Introduction to the Special Issue on Processing Reverberant Speech: Methodologies and Applications

THIS special issue focuses on techniques designed to handle reverberant speech directed at both humans and computers. Typically, these techniques include speech dereverberation, as part of the general speech acquisition process, and automatic speech recognition (ASR) in reverberant environments. A speech signal captured by distant microphones in any enclosure (e.g., conference rooms, offices, living rooms) will inevitably contain noise and reverberation components. Both components are detrimental to the quality of the observed speech and cause serious degradation in such speech applications as voice recording, teleconferencing, hearing aids, human–machine dialogue systems, and meeting diarization. These speech applications are becoming almost ubiquitously desirable, requiring acoustic transduction to take place in a much wider range of environments than conceived for the handheld telephone. Although many successful techniques have been developed for dealing with noise, particularly uncorrelated noise with simple spectral characteristics (e.g., white noise), until recently the problem of sound reverberation has remained essentially unsolved. This problem hampers wider use of naturally untethered, effortless acoustic interfaces for many applications

Recently, researchers have recognized the importance of distant-microphone speech interfaces and tackled the reverberation problem. Substantial progress has been reported in a variety of areas. A fundamental problem is the reduction/compensation of the reverberation effect in observed signals without prior knowledge of the room characteristics. Furthermore, the solution must often consider application-specific requirements, e.g., incremental/real-time processing, and robustness to deviations in recording conditions, such as the speaker and microphone locations.

This special issue highlights some major aspects of the recent progress in the field. ASR in reverberant environments is a challenging sub-topic for which this special issue offers new and advanced findings. Sehr *et al.* propose a very generic approach (REMOS) that determines all clean-speech and reverberation estimates during decoding for ASR based on a combined acoustic model consisting of a hidden Markov model (HMM) network and a reverberation model (RM). This paper's algorithm operates in the log mel spectral domain and significantly outperforms the earlier version of REMOS that operated in the mel spectral domain. Krueger and Haeb-Umbach present an approach where dereverberation is performed in the log mel spectral domain based on a stochastic reverberant observation model and connect ASR with dereverberation preprocessing in an optimal way based on a Bayesian inference framework known as uncertainty decoding. Gomez and Kawahara's paper proposes a method to optimize the dereverberation preprocessing that performs in the power spectral domain by controlling multi-band scale factors so as to maximize the likelihood of the acoustic model used for back-end ASR.

Speech dereverberation for human listening (and for more general speech acquisition purposes) is another important sub-topic in this special issue. Nakatani *et al.* contribute to the blind dereverberation of nonstationary reverberant speech signals based on linear inverse filtering. They theoretically and experimentally show that the use of variance normalized delayed linear prediction effectively reduces reverberation by waveform while robustly preserving the statistical characteristics of direct signals, even in the presence of certain additive noise and model mismatch. Jeub *et al.* propose two-stage processing that combines late reverberation suppression and dual-channel Wiener filtering. An advantage of this proposal is that dereverberation processing hardly affects such binaural cues as inter-aural intensity differences so that human listeners can perceive the naturally improved localization of the sound from the processed signals. Erkelens and Heusdens also propose a new study for single channel late-reverberation suppression under conditions where the room impulse responses may change quickly due to speaker movements, for example.

Important sub-topics in this field also include quality measure of reverberant and dereverberated speech signals and intelligible sound reproduction in reverberant environments. Falk *et al.* propose a new non-intrusive quality and intelligibility measure obtained by auditory system-inspired filter bank analysis. An adaptive measure termed speech to reverberation modulation energy ratio was developed and shown to outperform conventional standard measures. Arai *et al.* provide evaluation results for steady-state suppression-based intelligibility improvement for speech reproduction in reverberant environments, particularly focusing on the effectiveness for elderly listeners.

The other important sub-topic features reverberation-robust microphone array signal processing techniques, including source location/distance estimation and source separation. Recently, even though a large amount of work has been done to develop these techniques, reverberation remains a difficult problem that severely degrades the performances of the microphone array signal processing. The proposals of these techniques in this special issue are thus aimed at widening usability in real acoustical environments. For estimating source location and/or source distance, this issue has advanced results. In particular, an interesting new trend is discussed where reverberation is utilized as a clue to improve the estimation reliability. Ribeiro *et al.* show that early reflections can be used to provide more information about the source location, including resolution for range and elevation than are available in anechoic scenarios. In addition, Lu and Cooke utilize the direct-to-reverberation energy ratio for estimating the distance of static and moving sources. A new binaural equalization-cancel-

Digital Object Identifier 10.1109/TASL.2010.2062590

lation technique is proposed and examined with a probabilistic inference framework based on particle filtering. In contrast, Talantzis proposes another particle filtering-based inference for source localization and tracking based on an information theoretic approach. An integrated system that also incorporates voice activity detection is presented and tested in a real-world meeting scenario.

To improve the source separation performance in reverberant environments, this special issue mainly discusses new models of observed signals. Kowalski *et al.* address the problem of estimating source signals from an underdetermined convolutive mixture assuming known mixing filters. This paper's key contribution is the use of a wideband (time-domain) mixture fitting cost that circumvents the conventionally adopted narrowband approximation. Duong *et al.* model microphone signals in the context of under-determined convolutive blind source separation. In their proposed model, the covariance matrix of the desired signal is decomposed into a time-varying source variance and a time-invariant spatial covariance matrix and utilized for identifying the relationships between sources and observations. Masnadi-Shirazi *et al.* present a new framework for a convolutive BSS that extends independent vector analysis (IVA) to overcomplete and undercomplete cases. The paper by Woodruff and Wang describes a method to segregate and localize multiple speeches in reverberant environments by combining monaural and binaural cues. Hummersone *et al.* test a previously proposed binaural source separation that utilizes the precedence effect in real rooms with a range of reverberant conditions and shows that adaptation is necessary and can yield significant gains in the separation performance. On the other hand, Mandel and Ellis provide comparative investigations about speech source separation performance in reverberant conditions and show that the proposed metrics can be good predictors of ASR performance.

We believe that this is the first special issue on this topic published by any journal and that it will constitute a milestone on the road from research ideas to industrial deployment. We thank the authors for their contributions. We also express our appreciation to the reviewers for their help and offer our sincere gratitude to the Editor-in-Chief, Professor Helen Meng, the past Editor-in-Chief, Professor Mari Ostendorf, and the Publications Coordinator, Ms. Kathy Jackson, for their kind support and assistance on this special issue.

TOMOHIRO NAKATANI, *Lead Guest Editor*
NTT Communication Science Laboratories
NTT Corporation
Kyoto 619-0237, Japan

WALTER KELLERMANN, *Guest Editor*
Chair of Multimedia Communications and Signal Processing
University Erlangen-Nuremberg
91058 Erlangen, Germany

PATRICK NAYLOR, *Guest Editor*
Department of Electrical and Electronic Engineering
Imperial College London
London SW7 2AZ, U.K.

MASATO MIYOSHI, *Guest Editor*
Graduate School of Natural Science and Technology
Kanazawa University
Kanazawa 920-1192, Japan

BIING HWANG (FRED) JUANG, *Guest Editor*
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0250 USA

Tomohiro Nakatani (M'04–SM'06) received the B.E., M.E., and Ph.D. degrees from Kyoto University, Kyoto, Japan, in 1989, 1991, and 2002, respectively.

He is a Senior Research Scientist with NTT Communication Science Labs, NTT Corporation, Kyoto. His research interests include speech enhancement technologies for intelligent human–machine interfaces. In 2005, he visited the Georgia Institute of Technology, Atlanta, as a Visiting Scholar for a year. Since 2008, he has been a Visiting Assistant Professor at Nagoya University, Nagoya, Japan.

Dr. Nakatani has been a member of the IEEE Signal Processing Audio and Acoustics Technical Committee since 2009. He served as a Technical Program Chair of IEEE WASPAA-2007.

Walter Kellermann (M'89–SM'06–F'08) received the Dipl.-Ing. (Univ.) degree in electrical engineering from the University of Erlangen-Nuremberg, Erlangen, Germany, in 1983, and the Dr.-Ing. degree (with distinction) from the Technical University Darmstadt, Darmstadt, Germany, in 1988.

He is a Professor for communications at the Chair of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, Erlangen. From 1989 to 1990, he was a Postdoctoral Member of Technical Staff at AT&T Bell Laboratories, Murray Hill, NJ. In 1990, he joined Philips Kommunikations Industrie, Nuremberg, Germany. From 1993 to 1999, he was a Professor at the Fachhochschule Regensburg before he joined the University Erlangen-Nuremberg as a Professor and Head of the Audio Research Laboratory in 1999. In 1999, he cofounded the consulting firm DSP Solutions. He authored or coauthored 16 book chapters and more than 170 refereed papers in journals and conference proceedings.

Dr. Kellermann served as an Associate Editor and as Guest Editor to various journals, e.g., to the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING from 2000 to 2004, and to the *EURASIP Journal on Signal Processing*. Currently, he serves as an Associate Editor to the *EURASIP Journal on Advances in Signal Processing* and as a Member of the Overview Editorial Board for the IEEE Signal Processing Society. He was the general chair of the 5th International Workshop on Microphone Arrays in 2003 and the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics in 2005. He was the general co-chair of the 2nd International Workshop on Hands-Free Speech Communication and Microphone Arrays (HSCMA) in 2008. He was a Distinguished Lecturer of the IEEE Signal Processing Society for 2007 and 2008. He currently serves as a Chair of the Technical Committee for Audio and Acoustic Signal Processing of the IEEE Signal Processing Society.

Patrick A. Naylor (M'89–SM'07) received the B.Eng. degree from the University of Sheffield, Sheffield, U.K., in 1986 and the Ph.D. degree from Imperial College London, London, U.K., in 1990.

In 1990, he joined the academic staff in the Electrical and Electronics Engineering Department at Imperial College London where he is also Director of Postgraduate Studies. His research interests are in the areas of speech, audio, and acoustic signal processing. He also enjoys fruitful links with industry.

Dr. Naylor serves as an Associate Editor of the IEEE TRANSACTIONS ON AUDIO SPEECH AND LANGUAGE PROCESSING and a member of the IEEE Signal Processing Society Technical Committee on Audio and Acoustic Signal Processing.

Masato Miyoshi (M'87–SM'04) received M.E. and D.E. degrees from Doshisha University, Kyoto, Japan, in 1983 and 1991, respectively.

From 1983 to 2009, he engaged, as a Research Staff Member with Nippon Telegraph and Telephone Corporation, Kyoto, in research on audio signal processing. Since 2009, he has been serving as a Professor of the Graduate School of Natural Science and Technology, Kanazawa University, Kanazawa, Japan.

Prof. Miyoshi was honored to receive the 1988 IEEE Senior Award, the 1989 ASJ Awaya Prize Young Researcher Award, the 1990 and 2006 ASJ Sato Prize Paper Awards, the 2005 IEICE Best Paper Award, and the 2009 ASJ Technical Development Award.

Biing Hwang (Fred) Juang (M'80–SM'87–F'92) received the Ph.D. degree from the University of California, Los Angeles.

His research career, before joining the Georgia Institute of Technology (Georgia Tech), Atlanta, in 2002, spans over Speech Communication Research Lab. (1978–1979), Signal Technology, Inc. (1979–1982), Bell Labs (1982–2001), and Avaya Labs (2001–2002). At Georgia Tech, he holds the Motorola Foundation Chair Professorship, and is a Georgia Research Alliance Eminent Scholar. He has published extensively, including the book *Fundamentals of Speech Recognition* (Prentice-Hall, NJ), coauthored with L. R. Rabiner.

Prof. Juang has served a number of positions in the IEEE Signal Processing Society, including Editor-in-Chief of the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING and Chair of its Fellow Evaluation Committee. He received a number of technical awards, notable among which are several Best Paper awards in the area of speech communications and processing, the Technical Achievement Award from the Signal Processing Society of the IEEE, and the IEEE Third Millennium Medal. He is a Fellow of Bell Labs, a member of the U.S. National Academy of Engineering, and an Academician of Academia Sinica.