Introduction to the Special Section on Blind Signal Processing for Speech and Audio Applications

PEOPLE can engage in comprehensible conversations at a noisy cocktail party. This is the well-known "cocktail-party effect," whereby our ears can extract what a person is saying under such conditions. The aim of blind signal processing for speech and audio applications is to provide computers with this ability, thus enabling them to determine individual speech waveforms from mixtures.

Blind signal processing has previously been applied to a variety of problems including wireless communication and biomedical applications. However, as speech and audio signal mixtures in a natural (i.e., reverberant) environment are generally convolutive mixtures, they involve a structurally much more challenging task than instantaneous mixtures, which are prevalent in many other applications. Here, we mainly focus on the blind source separation (BSS) and blind dereverberation of speech and audio signals under reverberant conditions. BSS is an approach for estimating source signals using only information about their mixtures observed in each input channel. The estimation is performed without possessing information on each source, such as its frequency characteristics and location, or on how the sources are mixed.

For overdetermined or critically determined BSS where the number of sources is smaller than or equal to the number of microphones, BSS using independent component analysis (ICA) has gained a firm position. This approach utilizes spatial diversity to discriminate between desired and undesired components, i.e., it reduces the undesired component by forming a spatial null towards them.

Underdetermined BSS, where there are fewer microphones than source signals, is a challenging problem because of the inherently adverse conditions. ICA cannot be used for this case, but the sparseness of audio sources is very useful; we can utilize time–frequency diversity, where sources are active in different regions of the time–frequency plane. Inspiration for solving this difficult problem may also come from an understanding of the mechanisms involved in human auditory scene analysis. Even with a single microphone, we can separate a mixture by using the harmonicity and temporal structure of the sources. We can build a probabilistic framework by assuming a source model and separate a mixture by maximizing the *a posteriori* probability of the sources given the observations.

The dereverberation of speech signals is also a challenging problem. There are two approaches to blind dereverberation. The first relies on the blind channel identification of acoustic systems and calculates inverse filters to obtain the dereberverated signal. On the other hand, the second approach does not explicitly try to estimate the channel or channel-inverse.

The aim of this special section is to offer an opportunity to link these techniques in different areas and to find effective ways of solving this problem. After the review process, we selected the following seven papers for publication that span the approaches identified above. Overdetermined or critically determined BSS is represented by the spatio–temporal extensions of the FastICA algorithm proposed by Douglas *et al.*, a nonparametric independent vector analysis algorithm proposed by Lee and Lee, and the blind adaptive beamforming study of Warsitz and Haeb-Umbach. Underdetermined BSS is tackled by exploiting spectral sparseness in the papers of Aissa-El-Bey *et al.* and He *et al.*, and a model-based approach with a single channel is taken by Ozerov *et al.* Robust dereverberation of speech is addressed in the paper by Furuya *et al.*

We hope that this special issue will stimulate interest in the challenging area of BSS for speech and audio applications, and we look forward to seeing an increasing body of high-quality research aligned to this idea.

We would like to express our gratitude to the authors of the papers in this special section and also to the more than 70 reviewers who helped us evaluate the submissions. We also thank Isabel Trancoso, the Editor-in-Chief who initiated this project, and Mari Ostendorf, the Editor-in-Chief under whom it was concluded. Finally, thanks to Kathy Jackson for her administrative assistance throughout the process.

SHOJI MAKINO, *Guest Editor*
NTT Communication Science Laboratories
Kyoto 619-0237, Japan
maki@cslab.kecl.ntt.co.jp

TE-WON LEE, *Guest Editor*
University of California at San Diego
La Jolla, CA 92093-0523 USA
tewon@ucsd.edu

GUY J. BROWN, *Guest Editor*
University of Sheffield
Sheffield S1 4DP, U.K.
g.brown@dcs.shef.ac.uk

Digital Object Identifier 10.1109/TASL.2007.899167

Shoji Makino (A'89–M'90–SM'99–F'04) received the B.E., M.E., and Ph.D. degrees from Tohoku University, Sendai, Japan, in 1979, 1981, and 1993, respectively.

He is now an Executive Manager at the NTT Communication Science Laboratories, Kyoto, Japan. He is also a Guest Professor at the Hokkaido University, Sapporo, Japan. His research interests include adaptive filtering technologies and realization of acoustic echo cancellation and blind source separation of convolutive mixtures of speech.

Dr. Makino received the ICA Unsupervised Learning Pioneer Award in 2006. He is the author or coauthor of more than 200 articles in journals and conference proceedings and is responsible for more than 150 patents. He is a Tutorial speaker at ICASSP2007. He is a member of both the Awards Board and the Conference Board of the IEEE Signal Processing Society. He was an Associate Editor of the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING. He is a member of the Technical Committee on Audio and Electroacoustics of the IEEE Signal Processing Society and the Chair-Elect of the Technical Committee on Blind Signal Processing of the IEEE Circuits and Systems Society. He is a member of the International IWAENC Standing committee and a member of the International ICA Steering Committee. He is the General Chair of the WASPAA2007 at Mohonk, New Paltz, NY, the General Chair of the IWAENC2003 in Kyoto, and the Organizing Chair of the ICA2003, Nara, Japan.

Te-Won Lee (S'96–M'03–SM'06) received the diploma degree and the Ph.D. degree (summa cum laude) in electrical engineering from the University of Technology Berlin, Berlin, Germany, in 1995 and 1997, respectively.

He is an Associate Research Professor at the Institute for Neural Computation, University of California at San Diego, La Jolla, and a Collaborating Professor in the Biosystems Department, Korea Advanced Institute of Science and Technology (KAIST), Daejeon. He is also Co-Founder, President, and Chief Technical Officer at SoftMax, Inc., a technology-driven company in San Diego that focuses on commercializing applications of its signal separation technology He was a Max-Planck Institute Fellow (1995–1997) and a Research Associate at the Salk Institute for Biological Studies (1997–1999).

Dr. Lee received the Erwin-Stephan Prize for excellent studies (1994) from the University of Technology Berlin, the Carl-Ramhauser prize (1998) for excellent dissertations from the Daimler–Chrysler Corporation and the ICA Unsupervised Learning Pioneer Award (2007).

Guy J. Brown received the B.Sc. degree in applied science from Sheffield Hallam University, Sheffield, U.K., in 1988, and the Ph.D. degree in computer science and the M.Ed. degree from the University of Sheffield, Sheffield, U.K., in 1992 and 1997, respectively.

He has been a Visiting Research Scientist at LIMSI-CNRS, Paris, France, ATR, Kyoto, Japan, The Ohio State University, Columbus, and the Helsinki University of Technology, Espoo, Finland. He is currently a Senior Lecturer in computer science with the University of Sheffield. He has a long-established interest in computational auditory modeling and also has research interests in automatic speech recognition and music technology. He has authored and coauthored more than 80 papers in books, journals, and conference proceedings.