Learn the technology behind hearing aids, Siri, and Echo

Audio source separation and speech enhancement aim to extract one or more source signals of interest from an audio recording involving several sound sources. These technologies are among the most studied in audio signal processing today and bear a critical role in the success of hearing aids, hands-free phones, voice command and other noise-robust audio analysis systems, and music post-production software.

Research on this topic has followed three convergent paths, starting with sensor array processing, computational auditory scene analysis, and machine learning based approaches such as independent component analysis, respectively. This book is the first one to provide a comprehensive overview by presenting the common foundations and the differences between these techniques in a unified setting.

Key features:

• Consolidated perspective on audio source separation and speech enhancement.
• Both historical perspective and latest advances in the field, e.g. deep neural networks.
• Diverse disciplines: array processing, machine learning, and statistical signal processing.
• Covers the most important techniques for both single-channel and multichannel processing.

This book provides both introductory and advanced material suitable for people with basic knowledge of signal processing and machine learning. Thanks to its comprehensiveness, it will help students select a promising research track, researchers leverage the acquired cross-domain knowledge to design improved techniques, and engineers and developers choose the right technology for their target application scenario. It will also be useful for practitioners from other fields (e.g., acoustics, multimedia, phonetics, and musicology) willing to exploit audio source separation or speech enhancement as pre-processing tools for their own needs.

About the Author

Emmanuel Vincent is a Senior Research Scientist with Inria, Nancy, France. His research focuses on machine learning for speech and audio signal processing. He has been working on audio source separation for 15 years and co-authored over 180 publications in this field. His contributions include harmonic nonnegative matrix factorization, full-rank spatial covariance modeling, joint spatial/spectral estimation, deep learning based multichannel source separation, and objective performance metrics. He has given several keynotes, tutorials and summer school lectures, including at Interspeech 2012 and 2016, WASPAA 2015 and LVA/ICA 2015. He is a founding chair of the series of Signal Separation Evaluation Campaigns (SiSEC) and CHiME Speech Separation and Recognition Challenges and the chair of ISCA’s special interest group on Robust Speech Processing.

Tuomas Virtanen is a Professor with the Laboratory of Signal Processing, Tampere University of Technology, Finland, where he is leading the Audio Research Group. He is known for his pioneering work on single-channel sound source separation using nonnegative matrix factorization, and its application to noise-robust speech recognition, music content analysis, and sound event detection. His research interests also include content analysis and processing of audio signals in general. He has authored more than 170 publications and received four best paper awards. He is an IEEE Senior Member, a member of the Audio and Acoustic Signal Processing Technical Committee of IEEE Signal Processing Society, Associate Editor of IEEE/ACM Transaction on Audio, Speech, and Language Processing, and recipient of the ERC 2014 Starting Grant.

Sharon Gannot is a Full Professor at the Faculty of Engineering, Bar-Ilan University, Israel, where he is heading the Speech and Signal Processing laboratory and the Signal Processing Track. His research interests include multi-microphone speech processing; distributed algorithms for noise reduction and speaker separation; array processing on manifold; dereverberation; single-microphone speech enhancement; and speaker localization and tracking. He received the Bar-Ilan University's Outstanding Lecturer Award for 2010 and 2014 and the Bar-Ilan Rector Innovation in Research Award in 2018. He has co-authored over 200 publications and lectured tutorials at ICASSP 2012, EUSIPCO 2012, ICASSP 2013, and EUSIPCO 2013 and a keynote address at IWAENC 2012. He was a co-editor of the book Speech Processing in Modern Communication: Challenges and Perspectives
(Springer, 2012). He also served as an Associate Editor and a Senior Area Chair of the IEEE Transactions on Speech, Audio and Language Processing. He currently serves as the Chair of the IEEE Audio and Acoustic Signal Processing (AASP) Technical Committee.