APSIPA Transactions on Signal and Information Processing, 2023, 12, e24

Editorial

Editorial for the Special Issue on Advanced Acoustic, Sound and Audio Processing Techniques and Their Applications

Yu Tsao[1], Shoji Makino[2], Yoshinobu Kajikawa[3] and Nobutaka Ono[4]

[1] *Academia Sinica, 11529, Taiwan; yu.tsao@citi.sinica.edu.tw*
[2] *Waseda University, Kitakyushu, 808-0135, Japan; s.makino@waseda.jp*
[3] *Kansai University, Osaka, 564-8680, Japan; kaji@kansai-u.ac.jp*
[4] *Tokyo Metropolitan University, Tokyo, 192-0397, Japan; onono@tmu.ac.jp*

With recent advancements in sensing, computing, and communication capabilities, substantial quantities of acoustic, sound, and audio (ASA) data have become conveniently accessible. The extensive and diverse dataset presents an opportunity to develop systems for a wide range of applications using state-of-the-art artificial intelligence (AI) algorithms. Despite the proliferation of AI-driven systems utilizing ASA data and system frameworks, there is still untapped potential for improving performance and exploring novel directions. This special issue focuses on all aspects of pattern recognition, information retrieval, and front-end processing (enhancement, separation, and noise cancellation) of ASA signals. This special issue has collected nine excellent articles reviewed and highly recommended by the editors and reviewers.

The first paper is "Movable Virtual Sound Source Construction Based on Wave Field Synthesis using a Linear Parametric Loudspeaker Array," authored by Yuting Geng, Shiori Sayama, Masato Nakayama, and Takanobu Nishiura. This paper outlines a novel approach for constructing a movable virtual sound source (VSS) using a linear arrangement of parametric loudspeakers. Experimental findings indicate that, in comparison to a VSS generated using conventional electro-dynamic loudspeakers, the proposed method can

ISSN 2048-7703; DOI 10.1561/116.00001100

attain a higher precision when the VSS is moving parallel to the loudspeaker array.

The second paper is "BASPRO: A Balanced Script Producer for Speech Corpus Collection Based on the Genetic Algorithm," authored by Yu-Wen Chen, Hsin-Min Wang, and Yu Tsao. This paper introduces the BAlanced Script PROducer (BASPRO) system, which can automatically create phonetically balanced and diverse collections of Chinese sentences for collecting Mandarin speech data, with the aim of efficiently training speech processing models and testing their performance fairly. Experimental results show that speech enhancement and automatic speech recognition models trained using the created speech datasets outperform the models trained on randomly compiled speech datasets.

The third paper is "Missing Data Completion of Multi-channel Signals Using Autoencoder for Acoustic Scene Classification," authored by Yuki Shiroma, Yuma Kinoshita, Keisuke Imoto, Sayaka Shiota, Nobutaka Ono, and Hitoshi Kiya. This paper presents an autoencoder-based missing data completion method for multi-channel acoustic scene classification (ASC). Experimental results illustrate the effectiveness of the proposed autoencoder in accurately completing missing data, thereby improving the accuracy of ASC systems by utilizing the completed multi-channel signals.

The fourth paper is "A Review of Speech-centric Trustworthy Machine Learning: Privacy, Safety, and Fairness," authored by Tiantian Feng, Rajat Hebbar, Nicholas Mehlman, Xuan Shi, Aditya Kommineni, and Shrikanth Narayanan. This paper provides a comprehensive survey of speech-centric and trustworthy machine learning (ML) topics related to privacy, security, and fairness. In addition to providing a conclusive overview, this paper highlights several promising future research directions to motivate researchers interested in further exploration in this area.

The fifth paper is "Automatic Analyses of Dysarthric Speech based on Distinctive Features," authored by Ka Ho Wong and Helen Mei-Ling Meng. This paper presents an automatic analysis framework for dysarthric speech, using a linguistically motivated representation based on distinctive features (DFs). Experimental results show that there is little difference between the articulatory error rate profiles derived from manual and automatic speech transcriptions. The results also confirm the feasibility of the proposed framework as an automated method for processing dysarthric speech to achieve the pronunciation analysis described by DFs.

The sixth paper is "A Dual-branch Convolutional Network Architecture Processing on both Frequency and Time Domain for Single-channel Speech Enhancement," authored by Kanghao Zhang, Shulin He, Hao Li, and Xueliang Zhang. This paper proposes a novel real-time speech enhancement framework, called DBCN, which consists of a two-branch architecture: one branch takes the waveform as input, and the other branch takes the shifted real spectrum

as input. Experimental results show that the proposed system notably outperforms related algorithms in both causal and non-causal speech enhancement in very challenging environments.

The seventh paper is "Repeated Update of Demixing Vectors in Independent Low-rank Matrix Analysis for Better Separation," authored by Taishi Nakashima and Nobutaka Ono. This paper proposes an improved update algorithm for Independent Low-Rank Matrix Analysis (ILRMA). Experimental results on a music source separation task show that the proposed algorithm with the repeated update of demixing vectors outperforms conventional ILRMA in terms of separation performance and convergence speed.

The eighth paper is "Wavelength-Proportional Interpolation and Extrapolation of Virtual Microphone for Underdetermined Speech Enhancement," authored by Ryoga Jinzai, Kouei Yamaoka, Shoji Makino, Nobutaka Ono, Mitsuo Matsumoto and Takeshi Yamada. This paper proposes to applying the extrapolation of a virtual microphone as preprocessing of the maximum signal-to-noise ratio (SNR) beamformer and compare its speech enhancement performance with that of using the interpolation of a virtual microphone. Furthermore, the paper proposes to considering a trade-off relationship between performance at low and high frequencies to improve speech enhancement performance. Experimental results confirm that speech enhancement using virtual microphone extrapolation outperforms speech enhancement using virtual microphone interpolation.

The ninth paper is "EEG-based Auditory Attention Detection in Cocktail Party Environment," authored by Siqi Cai, Hongxu Zhu, Tanja Schultz, and Haizhou Li. This paper provides a comprehensive overview of state-of-the-art EEG-based auditory attention detection techniques and evaluation methods for assessing their performance. This paper reviews statistical and deep learning methods, indicates the gap between state-of-the-art techniques and practical needs in real-world applications, and outlines available resources for EEG-based auditory attention detection research.

The papers published in this special issue cover a wide range of ASA topics, highlight limitations of existing methods, offer novel solutions, and point to new research directions. This special issue is expected to provide readers with a comprehensive overview of the advances and potential applications in the field of ASA research. We also hope that this special issue will facilitate researchers to explore new directions and inspire new researchers to pursue ASA-related research areas. Finally, we would like to thank all reviewers for their devoted cooperation and constructive feedback.

Guest Editors:

Dr. Yu Tsao
Academia Sinica, Taiwan

Dr. Shoji Makino
Waseda University, Japan

Dr. Yoshinobu Kajikawa
Kansai University, Japan

Dr. Nobutaka Ono
Tokyo Metropolitan University, Japan