Introduction to the Issue on Acoustic Source Localization and Tracking in Dynamic Real-Life Scenes

I. Introduction

ACOUSTIC source localization and tracking is a well-studied topic in signal processing, but most traditional methods incorporate simplifying assumptions such as a point source, free-field propagation of the sound wave, static acoustic sources, time-invariant sensor constellations, and simple noise fields.

However, these assumptions may be seriously violated in a range of emerging real-life applications. In these applications, the environment is extremely challenging, with spatially distributed sources, reverberation and reflections, complex noise fields, multiple concurrent speakers, interference signals, and time-varying positions of sources and/or sensors.

Aside from classic source localization scenarios, present-day application areas include audio recording with mobile devices (e.g. cell phones, action cameras, and robots), video conferencing on the go, and recording for 3D reproduction and virtual reality. Underwater acoustic localization is another application, which is characterized by intricate propagation patterns.

This special issue presents recent advances in the development of signal processing methods for localization and tracking of acoustic sources and the associated theory and applications. Addressing the challenges raised by real-life environments, novel methods are introduced representing advances in array processing, speech processing, and the use of data inference tools.

II. Contributions

Fourteen manuscripts were accepted for publication in this special issue, addressing localization and tracking of acoustic sources in underwater environments or in air.

The papers can be roughly categorized in four groups: Papers proposing algorithms using either deep learning, statistical inference or array processing, as well as papers addressing sound propagation modelling. These four groups are summarized below.

A. Algorithms Using Deep Learning

In the paper "Multi-speaker DOA estimation using deep convolutional networks trained with noise signals" a convolutional neural network (CNN)-based supervised learning method for estimating the directions of arrival (DOAs) of multiple speakers is proposed. Multi-speaker DOA estimation is formulated as a multi-class multi-label classification problem, where the assignment of each DOA label to the input feature is treated as a separate binary classification problem. In a comprehensive evaluation, it was empirically shown that for a microphone array with M microphones, $M-1$ convolution layers are required for the best localization performance. The proposed method outperforms the multiple signal classification (MUSIC) and the steered response power with phase transform (SRP-PHAT) in terms of mean absolute error (MAE).

The paper "CRNN-based multiple DOA estimation using ambisonics acoustic intensity features" proposes a neural network to improve DOA estimation in reverberant environments using features derived from acoustic intensities as input to a convolutional recurrent neural network (CRNN). Analysis of the workings of the proposed method using layer-wise relevance propagation is also presented, together with a detailed description of experimental results. It is demonstrated that the technique can identify the sound "attack" times as being most relevant, in agreement with psychoacoustic studies.

In the paper "Sound event localization and detection of overlapping sources using convolutional recurrent neural networks", a CRNN for joint sound event localization and detection (SELD) of multiple overlapping sound events in 3D-space is proposed. The proposed network takes a sequence of consecutive spectrogram time frames as input and maps it to two outputs in parallel. At the first output, the sound event detection is carried out as a multi-label classification task on each time frame producing temporal activity for all the sound event classes. At the second output, localization is carried out by estimating the 3D Cartesian coordinates of the DOA for each sound event class using multi-output regression. The proposed method is able to associate multiple DOAs with respective sound event labels and further track this association over time. The paper also provides a survey of many neural network-based localizers.

B. Algorithms Using Statistical Inference

The paper "Robust ocean acoustic localization with sparse Bayesian learning in non-stationary noise" notes that the matched-field processing (MFP) problem can be thought of as an underdetermined set of linear equations, which suggests a compressed sensing (CS) approach. There is some literature showing that CS implemented using the non-greedy basis

Digital Object Identifier 10.1109/JSTSP.2019.2906923

pursuit (BP) method has significant shortcomings. The insight provided by this paper is to use sparse Bayesian learning (SBL) to learn the sparsity profile (common across snapshots and frequencies) corresponding to the location of present sources. The approach is evaluated using real data: the SWellEx-96 experiment.

The paper "Distributed source localization in acoustic sensor networks using the coherent-to-diffuse power ratio" proposes an approach to acoustic source localization that "learns" the source-to-microphone distances in reverberant rooms. The feature used in the Bayesian learning model is based on the coherent-to-diffuse power ratio (CDR), the power of the coherent signal component to the power of the reverberation, assumed here to be a function only of the relative source-to-microphone distances and not of position. The trained model based on Gaussian processes produces estimates of source-to-microphone distances given estimates of CDRs. The source is then localized by weighted least-squares combining the estimated source-to-microphone distances. The algorithm has been designed to minimize data communication among microphone nodes and is proposed as a viable distributed-sensor processing technique.

The paper "Speaker tracking based on distributed particle filter and iterative covariance intersection in distributed microphone networks" presents a method for tracking an acoustic source using a network of spatially distributed microphone pairs. The method is based on distributed particle filtering, with a modified iterative covariance intersection (MICI) fusion strategy. First, the classical generalized cross-correlation phase transform (GCC-PHAT) function is used to estimate the TDOA at each node. Then, multiple TDOA measurements are sent to the local particle filter at each node to estimate the speakers local state posterior. Finally, the distributed calculation of the local posteriors is implemented via the MICI fusion strategy to obtain a global consistent state estimate at each node. The performance of the method is verified in both simulated and real-life noisy and reverberant environments. It is also shown to be robust to node failures and to unknown network size.

In the paper "Online localization and tracking of multiple moving speakers in reverberant environment" the authors use the direct-path relative transfer function (DP-RTF), an inter-channel feature that encodes acoustic information against reverberation in a robust fashion, and propose an online algorithm well-suited for estimating the DP-RTFs associated with moving audio sources. They also assign the DP-RTFs to audio-source directions. To this end, they adopt a maximum-likelihood formulation and use an exponentiated gradient to efficiently update the source-direction estimates starting from their currently available values. To track multiple speakers, a Bayesian framework is adopted. In this context, the authors propose a variational approximation of the posterior filtering distribution associated with multiple speaker tracking and an efficient variational expectation maximization solver. The proposed online localization and tracking method is evaluated using two datasets that contain recordings in real environments.

The paper "An active acoustic track-before-detect approach for finding underwater mobile targets" considers the challenge of tracking and estimating the size of a single submerged target in a highly reverberant underwater environment using a single active acoustic transceiver. This problem occurs in many applications, ranging from security to environmental research. Considering that the target can be either slow (e.g., a scuba diver) or fast moving (e.g., a shark), the authors avoid assumptions about the targets reflection patterns or motion type, and instead, perform probabilistic tracking using a constraint Viterbi algorithm, whereby detection is performed based on the maximum likelihood criterion, implemented using the expectation-maximization (EM) approach. Based on the tracked path, the authors then evaluate the target's size. To test their approach, they have performed extensive simulations as well as eight sea experiments in different environmental settings to track both a scuba diver and a sandbar shark. The simulation results show a tracking performance that is close to the Cramér-Rao lower bound.

C. Algorithms Based on Array Processing Methodologies

The authors of the paper "An underwater acoustic positioning algorithm for compact arrays with arbitrary configuration" use simple geometrical insight to find a fast approach for localization in a small underwater ultra-short baseline (USBL) array with an arbitrary configuration. In the proposed algorithm, the bearing is estimated by applying a group of linear equations. The method is evaluated by both simulations and real data. Both numerical and experimental results indicate that the proposed algorithm outperforms competing methods in terms of positioning accuracy when some of the array sensors malfunction.

The paper "Direction of arrival estimation for reverberant speech based on enhanced decomposition of the direct sound" presents an approach for speaker localization in reverberant conditions, and compares this to existing approaches. The proposed approach uses a single-term signal-subspace MUSIC to determine a threshold for time-frequency (TF) bins that contain coherent single-source information, and utilize these to extract the DOA of the direct sound of the sources. The proposed direct-path dominance (DPD) test is compared to previous DPD tests, and to other recently proposed reverberation robust methods, using computer simulations and an experimental study, demonstrating its potential advantage. The studies include multiple speakers in highly reverberant environments, therefore representing challenging real-life acoustics scenes.

The paper "Acoustic source localization based on geometric projection in reverberant and noisy environments" deals with the problem of acoustic source localization from the perspective of geometric projection with four types of narrowband power functions and three fusion methods for broadband sources. Interestingly, it is shown that widely-used conventional algorithms (e.g. MUSIC, SRP-PHAT) could be cast into the presented framework based on the projection of the observation signal vector onto a hypothesized steering vector.

The paper "Subspace-based algorithms for localization and tracking of multiple near-field sources" investigates the task of estimating and tracking the DOAs and ranges of multiple near-field (NF) narrowband sources impinging on a uniform linear array (ULA). It presents a simple subspace-based algorithm for the localization of NF sources (SALONS), where the computationally costly eigen-decomposition and spectrum peak search are avoided. In the proposed SALONS algorithm, the DOAs and the ranges are estimated separately using a one-dimensional subspace-based estimation technique. Furthermore, an online algorithm is developed for tracking multiple moving NF sources with possible crossover points of their trajectories. The effectiveness and the theoretical analysis of the presented algorithms are verified through numerical examples.

D. Sound-Field Modelling

The paper “Sparse representation of a spatial sound field in a reverberant environment” proposes a general model for sparse sound field decomposition in a reverberant environment. Numerical and experimental results indicate that this decomposition enables an accurate sound field estimation, inside a region including sources. This study demonstrates that, in a reverberant environment, the reverberating component of the field cannot be treated as a residual, but must be explicitly treated. It was shown that this reverberating component can be modeled as the sum of a small number of plane waves, and a low-rank part.

The paper “Passive source depth discrimination in deep-water” addresses the problem of passive source depth discrimination in ocean acoustics using a horizontal line array. The scope is restricted to low-frequency sources, broadband signals, a deep-water environment, and distant sources at the endfire position. In this context, the environment acts as a dispersive waveguide, and classical source localization methods based on plane waves or any other simplistic wave model should not be used. Instead, a method based on the modal behavior driving the propagation is proposed. It uses the concept of the waveguide invariant, a scalar that summarizes the waveguide dispersion. In deep water, the waveguide invariant largely depends on source depth and thus represents an important input for source depth discrimination. An algorithm is proposed to compute the energy ratio in groups of modal interferences. The input data for the algorithm is a range-frequency intensity, as measured on a horizontal line array. The modal interference groups are defined based on their respective waveguide invariant values that in turn depend on the source depth. This paper proposes a source depth discrimination method that is expressed as a binary classification problem. As long as the sound speed profile features a surface thermocline, the algorithm does not require detailed knowledge about the environment and it allows the classification of sources under two hypotheses, above or under a user-chosen threshold depth.

S. GANNOT, *Editor*
Bar-Ilan University
Ramat Gan 5290002, Israel

M. HAARDT, *Editor*
Technische Universität Ilmenau
Ilmenau 98693, Germany

W. KELLERMANN, *Editor*
Friedrich-Alexander Universität
Erlangen-Nürnberg
Erlangen 91054, Germany

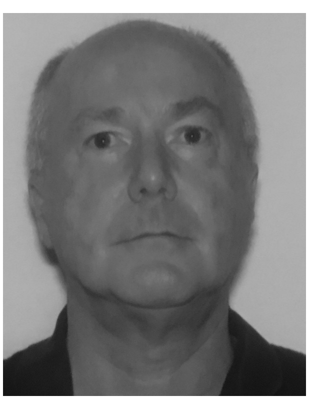
P. WILLETT, *Editor*
University of Connecticut
Storrs, CT 06269 USA

Sharon Gannot (S'92–M'01–SM'06) received the B.Sc. degree (*summa cum laude*) in electrical engineering from the Technion-Israel Institute of Technology, Haifa, Israel, in 1986, and the M.Sc. (*cum laude*) and Ph.D. degrees in electrical engineering from Tel-Aviv University, Tel Aviv, Israel, in 1995 and 2000, respectively. In 2001, he held a postdoctoral position with the Department of Electrical Engineering, KU Leuven, Leuven, Belgium. From 2002 to 2003, he held a research and teaching position with the Faculty of Electrical Engineering, Technion-Israel Institute of Technology. He is currently a Full Professor with the Faculty of Engineering, Bar-Ilan University, Ramat Gan, Israel, where he is heading the Acoustic Signal Processing Laboratory and the Signal Processing Track and serving as a Deputy Dean. Since April 2018, he has also been a part-time Professor with the Technical Faculty of IT and Design, Aalborg University, Aalborg, Denmark. He is currently a Moderator for arXiv in the field of audio and speech processing. He is also a Reviewer for many IEEE journals and conferences.

His research interests include statistical signal processing and machine learning algorithms (including manifold learning and deep learning) with applications to single- and multi-microphone speech processing. Specifically, distributed algorithms for ad hoc microphone arrays, speech enhancement, noise reduction and speaker separation, dereverberation, single microphone speech enhancement, speaker localization, and tracking. He was an Associate Editor for the *EURASIP Journal of Advances in Signal Processing* during 2003–2012, and an Editor for several special issues on multi-microphone speech processing of the same journal. He was a Guest Editor for the Elsevier *Speech Communication* and *Signal Processing* journals. He is currently the Lead Guest Editor of a special issue on speaker localization for the IEEE Journal of Selected Topics in Signal processing. He was an Associate Editor for the IEEE Transactions on Audio, Speech, and Language Processing during 2009–2013, and the Area Chair for the same journal during 2013–2017. Since January 2010, he has been a member of the Audio and Acoustic Signal Processing technical committee of the IEEE and served as the Committee Chair during 2017–2018. Since 2005, he has also been a member of the technical and steering committee of the International Workshop on Acoustic Signal Enhancement (IWAENC) and was the General Co-Chair of the IWAENC held in Tel-Aviv, Israel, in August 2010. He was the General Co-Chair of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics in October 2013. He was selected (with colleagues) to present tutorial sessions at ICASSP 2012, EUSIPCO 2012, ICASSP 2013, EUSIPCO 2013, and EUSIPCO 2019 and was a keynote speaker for IWAENC 2012, LVA/ICA 2017, and ITG 2018. He is the recipient of the Bar-Ilan University Outstanding Lecturer Award in 2010 and 2014 and the Rector Innovation in Research Award in 2018. He is also a co-recipient of 11 Best Paper Awards.

Martin Haardt (S'90–M'98–SM'99–F'18) received the degrees in electrical engineering from Ruhr-University Bochum, Bochum, Germany, and from Purdue University, Lafayette, IN, USA, the Diplom-Ingenieur (M.S.) degree from Ruhr-University Bochum in 1991, and the Doktor-Ingenieur (Ph.D.) degree from the Munich University of Technology, Munich, Germany, in 1996. He has been a Full Professor with the Department of Electrical Engineering and Information Technology and the Head of the Communications Research Laboratory, Ilmenau University of Technology, Ilmenau, Germany, since 2001. From 2012 to 2017, he was an Honorary Visiting Professor with the Department of Electronics, University of York, U.K. In 1997, he joined Siemens Mobile Networks, Munich, Germany, where he was responsible for strategic research for third generation mobile radio systems. From 1998 to 2001, he was the Director for International Projects and University Cooperations in the mobile infrastructure business of Siemens in Munich, where his work focused on mobile communications beyond the third generation. During his time at Siemens, he also taught in the International Master of Science in Communications Engineering Program at the Munich University of Technology. In the fall of 2006 and the fall of 2007, he was a Visiting Professor with the University of Nice, Sophia-Antipolis, France, and with the University of York, U.K, respectively. His research interests include wireless communications, array signal processing, high-resolution parameter estimation, as well as numerical linear and multi-linear algebra. He was the recipient of the 2009 Best Paper Award from the IEEE Signal Processing Society, the Vodafone (formerly Mannesmann Mobilfunk) Innovations-Award for outstanding research in mobile communications, the ITG Best Paper Award from the Association of Electrical Engineering, Electronics, and Information Technology (VDE), and the Rohde & Schwarz Outstanding Dissertation Award. He was an Associate Editor for the IEEE Transactions on Signal Processing (2002–2006 and 2011–2015), the IEEE Signal Processing Letters (2006–2010), the *Research Letters in Signal Processing* (2007–2009), the Hindawi *Journal of Electrical and Computer Engineering* (since 2009), the *EURASIP Signal Processing Journal* (2011–2014), and as a Guest Editor for the *EURASIP Journal on Wireless Communications and Networking*. Since 2011, he has been an elected member of the Sensor Array and Multichannel (SAM) technical committee of the IEEE Signal Processing Society, where he was the Vice-Chair (2015–2016), Chair (2017–2018), and Past Chair (2019). Moreover, he was the Technical Co-Chair of PIMRC 2005 in Berlin, Germany, ISWCS 2010 in York, UK, the European Wireless 2014 in Barcelona, as well as the Asilomar Conference on Signals, Systems, and Computers 2018 and as the General Co-Chair of ISWCS 2013 in Ilmenau, Germany, CAMSAP 2013 in Saint Martin, French Antilles, WSA 2015 in Ilmenau, SAM 2016 in Rio de Janeiro, Brazil, CAMSAP 2017 in Curacao, Dutch Antilles, and SAM 2020 in Hangzhou, China.

Walter Kellermann received the Dipl.-Ing. (Univ.) degree in electrical engineering from the University of Erlangen-Nuremberg, Erlangen, Germany, in 1983, and the Dr.-Ing. degree from the Technical University Darmstadt, Darmstadt, Germany, in 1988. From 1989 to 1990, he was a postdoctoral Member of Technical Staff with AT&T Bell Laboratories, Murray Hill, NJ. In 1990, he joined Philips Kommunikations Industrie, Nuremberg, Germany, to work on hands-free communication in cars. From 1993 to 1999, he was a Professor with the Fachhochschule Regensburg, where he also became the Director of the Institute of Applied Research in 1997. He has been a Professor for communications with the University of Erlangen-Nuremberg, Germany, since 1999. In 1999, he cofounded DSP Solutions, a consulting firm in digital signal processing, and joined the University Erlangen-Nuremberg as a Professor and the Head of the Audio Research Laboratory. In 2016, he was a Visiting Fellow with Australian National University, Canberra, Australia. He authored or coauthored 21 book chapters, more than 300 refereed papers in journals and conference proceedings, as well as more than 70 patents. His current research interests include speech signal processing, array signal processing, adaptive filtering, and its applications to acoustic human–machine interfaces. He is a co-recipient of ten Best Paper Awards. He was an Associate Editor and Guest Editor to various journals, including the IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING from 2000 to 2004, the IEEE SIGNAL PROCESSING MAGAZINE in 2015, and is currently an Associate Editor for the *EURASIP Journal on Applied Signal Processing*. He was the General Chair of seven mostly IEEE-sponsored workshops and conferences. He served as a Distinguished Lecturer of the IEEE Signal Processing Society (SPS) from 2007 to 2008. He was the Chair of the IEEE SPS Technical Committee for Audio and Acoustic Signal Processing from 2008 to 2010, a member of the IEEE James L. Flanagan Award Committee from 2011 to 2014, a member of the SPS Board of Governors (2013–2015), Vice-President Technical Directions of the IEEE Signal Processing Society (2016–2018) and is currently a member of the SPS Nominations Appointments Committee (2019–2020). He was awarded the Julius von Haast Fellowship by the Royal Society of New Zealand in 2012 and the Group Technical Achievement Award of the European Association for Signal Processing in 2015.



Peter Willett (F'03) has been a faculty member with the Electrical and Computer Engineering Department, University of Connecticut, Storrs, CT, USA, since 1986. Since 1998, he has been a Professor. His primary areas of research have been statistical signal processing, detection, machine learning, communications, data fusion, and tracking. He has been a Chief Editor for IEEE AEROSPACE AND ELECTRONIC SYSTEMS SOCIETY MAGAZINE (since 2018). He was an Editor-in-Chief for the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS from 2006 to 2011 and the IEEE SIGNAL PROCESSING LETTERS from 2014 to 2016. He was also an AESS Vice President for publications during 2012–2014. He is a member of IEEE Signal Processing Society's Technical Activities and Conference Boards. He is the Chief Editor for the IEEE AEROSPACE AND ELECTRONIC SYSTEMS MAGAZINE. He is a member of the IEEE AESS Board of Governors and was the Chair of IEEE Signal Processing Society's Sensor-Array and Multichannel technical committee.