Editorial

Microphone Array Speech Processing

Sven Nordholm (EURASIP Member),[1] Thushara Abhayapala (EURASIP Member),[2] Simon Doclo (EURASIP Member),[3] Sharon Gannot (EURASIP Member),[4] Patrick Naylor (EURASIP Member),[5] and Ivan Tashev[6]

[1] *Department of Electrical and Computer Engineering, Curtin University of Technology, Perth, WA 6845, Australia*
[2] *College of Engineering & Computer Science, The Australian National University, Canberra, ACT 0200, Australia*
[3] *Institute of Physics, Signal Processing Group, University of Oldenburg, 26111 Oldenburg, Germany*
[4] *School of Engineering, Bar-Ilan University, 52900 Tel Aviv, Israel*
[5] *Department of Electrical and Electronic Engineering, Imperial College, London SW7 2AZ, UK*
[6] *Microsoft Research, USA*

Correspondence should be addressed to Sven Nordholm, s.nordholm@curtin.edu.au

Significant knowledge about microphone arrays has been gained from years of intense research and product development. There have been numerous applications suggested, for example, from large arrays (in the order of >100 elements) for use in auditoriums to small arrays with only 2 or 3 elements for hearing aids and mobile telephones. Apart from that, microphone array technology has been widely applied in speech recognition, surveillance, and warfare. Traditional techniques that have been used for microphone arrays include fixed spatial filters, such as, frequency invariant beamformers, optimal and adaptive beamformers. These array techniques assume either model knowledge or calibration signal knowledge as well as localization information for their design. Thus they usually combine some form of localisation and tracking with the beamforming. Today contemporary techniques using blind signal separation (BSS) and time frequency masking technique have attracted significant attention. Those techniques are less reliant on array model and localization, but more on the statistical properties of speech signals such as sparseness, non-Gaussianity, and non-stationarity. The main advantage that multiple microphones add from a theoretical perspective is the spatial diversity, which is an effective tool to combat interference, reverberation, and noise. The underpinning physical feature used is a difference in coherence in the target field (speech signal) versus the noise field. Viewing the processing in this way one can understand also the difficulty in enhancing highly reverberant speech given that we only can observe the received microphone signals.

This special issue contains contributions to traditional areas of research such as frequency invariant beamforming [1], hand-free operation of microphone arrays in cars [2], and source localisation [3]. The contributions show new ways to study these traditional problems and give new insights into those problems. Small size arrays have always a lot of applications and interest for mobile terminals, hearing aids, and close up microphones [4]. The novel way to represent small size arrays leads to a capability to suppress multiple interferers. Abnormalities in noise and speech stemming from processing are largely unavoidable, and using nonlinear processing results often in significant character change particularly in noise character. It is thus important to provide new insights into those phenomena particularly the so called musical noise [5]. Finally, new and unusual use of microphone arrays is always interesting to see. Distributed microphone arrays in a sensor network [6] provide a novel approach to find snipers. This type of processing has good opportunities to grow in interest for new and improved applications.

The contributions found in this special issue can be categorized to three main aspects of microphone array processing: (i) microphone array design based on eigenmode decomposition [1, 4]; (ii) multichannel processing methods [2, 5]; and (iii) source localisation [3, 6].

The paper by Zhang et al., "*Selective frequency invariant uniform circular broadband beamformer*" [1], describes a design method for Frequency-Invariant (FI) beamforming. This problem is a well-known array signal processing technique used in many applications such as, speech acquisition, acoustic imaging and communications purposes. However, many existing FI beamformers are designed to have a frequency invariant gain over all angles. This might not be necessary and if a gain constraint is confined to a specific angle, then the FI performance over that selected region (in frequency and angle) can be expected to improve. Inspired by this idea, the proposed algorithm attempts to optimize the frequency invariant beampattern solely for the mainlobe and relax the FI requirement on the sidelobes. This sacrifice on performance in the undesired region is traded off for better performance in the desired region as well as reduced number of microphones employed. The objective function is designed to minimize the overall spatial response of the beamformer with a constraint on the gain being smaller than a predefined threshold value across a specific frequency range and at a specific angle. This problem is formulated as a convex optimization problem and the solution is obtained by using the Second-Order Cone Programming (SOCP) technique. An analysis of the computational complexity of the proposed algorithm is presented as well as its performance. The performance is evaluated via computer simulation for different number of sensors and different threshold values. Simulation results show that the proposed algorithm is able to achieve a smaller mean square error of the spatial response gain for the specific FI region compared to existing algorithms.

The paper by Derkx, "*First-order azimuthal null-steering for the suppression of two directional interferers*" [4] shows that an azimuth steerable first-order super directional microphone response can be constructed by a linear combination of three eigenbeams: a monopole and two orthogonal dipoles. Although the response of a (rotation symmetric) first-order response can only exhibit a single null, the paper studies a slice through this beampattern lying in the azimuthal plane. In this way, a maximum of two nulls in the azimuthal plane can be defined. These nulls are symmetric with respect to the main-lobe axis. By placing these two nulls on maximally two-directional sources to be rejected and compensating for the drop in level for the desired direction, these directional sources can be effectively rejected without attenuating the desired source. An adaptive null-steering scheme for adjusting the beampattern, which enables automatic source suppression, is presented. Closed-form expressions for this optimal null-steering are derived, enabling the computation of the azimuthal angles of the interferers. It is shown that the proposed technique has a good directivity index when the angular difference between the desired source and each directional interferer is at least 90 degrees.

In the paper by Takahashi et al. "*Musical noise analysis in methods of integrating microphone array and spectral subtraction based on higher-order statistics*" [5], an objective analysis on musical noise is conducted. The musical noise is generated by two methods of integrating microphone array signal processing and spectral subtraction. To obtain better noise reduction, methods of integrating microphone array signal processing and nonlinear signal processing have been researched. However, nonlinear signal processing often generates musical noise. Since such musical noise causes discomfort to users, it is desirable that musical noise is mitigated. Moreover, it has been recently reported that higher-order statistics are strongly related to the amount of musical noise generated. This implies that it is possible to optimize the integration method from the viewpoint of not only noise reduction performance but also the amount of musical noise generated. Thus, the simplest methods of integration, that is, the delay-and-sum beamformer and spectral subtraction, are analysed and the features of musical noise generated by each method are clarified. As a result, it is clarified that a specific structure of integration is preferable from the viewpoint of the amount of generated musical noise. The validity of the analysis is shown via a computer simulation and a subjective evaluation.

The paper by Freudenberger et al., "*Microphone diversity combining for in-car applications*" [2], proposes a frequency domain diversity approach for two or more microphone signals, for example, for in-car applications. The microphones should be positioned separately to ensure diverse signal conditions and incoherent recording of noise. This enables a better compromise for the microphone position with respect to different speaker sizes and noise sources. This work proposes a two-stage approach: In the first stage, the microphone signals are weighted with respect to their signal-to-noise ratio and then summed similar to maximum-ratio-combining. The combined signal is then used as a reference for a frequency domain least-mean-squares (LMS) filter for each input signal. The output SNR is significantly improved compared to coherence-based noise reduction systems, even if one microphone is heavily corrupted by noise.

The paper by Ichikawa et al., "*DOA estimation with local-peak-weighted CSP*" [3], proposes a novel weighting algorithm for Cross-power Spectrum Phase (CSP) analysis to improve the accuracy of direction of arrival (DOA) estimation for beamforming in a noisy environment. As a sound source, a human speaker is used, and as a noise source broadband automobile noise is used. The harmonic structures in the human speech spectrum can be used for weighting the CSP analysis, because harmonic bins must contain more speech power than the others and thus give us more reliable information. However, most conventional methods leveraging harmonic structures require pitch estimation with voiced-unvoiced classification, which is not sufficiently accurate in noisy environments. The suggested approach employs the observed power spectrum, which is directly converted into weights for the CSP analysis by retaining only the local peaks considered to be coming from a harmonic structure. The presented results show that the proposed approach significantly reduces the errors in localization, and it also shows further improvement when used with other weighting algorithms.

The paper by Lindgren et al., "*Shooter localization in wireless microphone networks*" [6], is an interesting combination of microphone array technology with distributed

communications. By detecting the muzzle blast as well as the ballistic shock wave, the microphone array algorithm is able to locate the shooter in the case when the sensors are synchronized. However, in the distributed sensor case, synchronization is either not achievable or very expensive to achieve and therefore the accuracy of localization comes into question. Field trials are described to support the algorithmic development.

Sven Nordholm
Thushara Abhayapala
Simon Doclo
Sharon Gannot
Patrick Naylor
Ivan Tashev

References

[1] X. Zhang, W. Ser, Z. Zhang, and A. K. Krishna, "Selective frequency invariant uniform circular broadband beamformer," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 678306, 11 pages, 2010.

[2] J. Freudenberger, S. Stenzel, and B. Venditti, "Microphone diversity combining for In-car applications," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 509541, 13 pages, 2010.

[3] O. Ichikawa, T. Fukuda, and M. Nishimura, "DOA estimation with local-peak-weighted CSP," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 358729, 9 pages, 2010.

[4] R. M. M. Derkx, "First-order adaptive azimuthal null-steering for the suppression of two directional interferers," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 230864, 16 pages, 2010.

[5] Yu. Takahashi, H. Saruwatari, K. Shikano, and K. Kondo, "Musical-noise analysis in methods of integrating microphone array and spectral subtraction based on higher-order statistics," *EURASIP Journal on Advances in Signal Processing*, vol. 2010, Article ID 431347, 25 pages, 2010.

[6] D. Lindgren, O. Wilsson, F. Gustafsson, and H. Habberstad, "Shooter localization in wireless sensor networks," in *Proceedings of the 12th International Conference on Information Fusion (FUSION '09)*, pp. 404–411, July 2009.