A DISTRIBUTED STEERED RESPONSE POWER APPROACH TO SOURCE LOCALIZATION IN WIRELESS ACOUSTIC SENSOR NETWORKS

*Bilgesu Çakmak, *Thomas Dietzen, *Randall Ali, †Patrick Naylor, *Toon van Waterschoot

*Dept. of Electrical Engineering (ESAT-STADIUS), KU Leuven, Belgium
†Dept. of Electrical and Electronic Engineering, Imperial College London, UK

ABSTRACT

In wireless acoustic sensor networks (WASNs), the conventional steered response power (SRP) approach to source localization requires each node to transmit its microphone signal to a fusion center. As an alternative, this paper proposes two different fusion strategies for local, single-node SRP maps computed using only the microphone pairs within a node. In the first fusion strategy, we sum all single-node SRP maps in a fusion center, requiring less communication than the conventional SRP approach because the single-node SRP maps typically have less parameters than the raw microphone signals. In the second fusion strategy, the single-node SRP maps are distributively averaged without using a fusion center, requiring communication amongst connected nodes only. Simulations show that we achieve a good trade-off between communicational load and localization performance.

***Index Terms*—** Source localization, wireless acoustic sensor network, generalized cross correlation, steered response power, distributed microphone processing, distributed averaging

1. INTRODUCTION

In many applications of acoustic signal processing, source localization in noisy and reverberant environments is a crucial task [1, 2]. Over many years, microphone arrays have been used for this challenging task [3, 4, 5]. Lately, the production of small arrays for implementation in wireless acoustic sensor networks (WASN) has increased and source localization using a WASN has become a trending topic. In [6], the authors used a compact circular array, equipped with additional vertically placed microphones to estimate the direction of arrival (DOA). In [7], the authors proposed a method for the integration of spatial likelihood functions (SLFs) produced by microphone arrays. Deep learning based methods are also exploited in source localization using distributed microphones [8, 9].

For devices equipped with multiple microphones such as hearing aids and teleconferencing systems, acoustic signals from a sound source will arrive at the different microphones at different times. Hence a conventional approach to source localization problem consists of estimating the time difference of arrivals (TDOAs) between the microphone signals and then mapping these to a corresponding source position.

One particularly popular source localization approach is the steered response power (SRP) approach [10, 11, 12]. In SRP, locations are estimated using the generalized cross-correlation (GCC) often with a phase-transform (PHAT) weighting [1, 13]. A power map of the acoustic scene is built from the output power of a delay-and-sum beamformer (DSB) steered towards a set of candidate locations, as informed by the estimated TDOAs, and the source location is estimated to be at the position of maximum power over this spatial grid [1, 14]. Although SRP is a robust algorithm, it is often unsuitable for real-time applications due to its high computational complexity, which is in part due to the fact that cross-correlations between all microphone pairs have to be computed. In a WASN context, conventional SRP would require that the processing takes place in a centralized manner where it is assumed that all microphone signals are available in a fusion center. In the literature, the SRP algorithm has mostly been employed using a single microphone array [15, 16, 17]. As opposed to the conventional SRP strategy, in this paper, we propose two distributed SRP strategies which can be used in a WASN. A typical WASN consists of several nodes distributed over the acoustic environment that can communicate with one another or with a fusion center with a wireless communication protocol [18].

We propose two approaches to build the SRP maps in a distributed manner, by fusing local, single-node SRP maps computed using only the microphone pairs within a node, avoiding the need to share raw microphone signals in a fusion center. In the first fusion strategy, we sum all single-node SRP maps in a fusion center, requiring less communication than the conventional SRP approach because the single-node SRP maps typically have less parameters than the raw microphone signals. In the second fusion strategy, the single-node SRP maps are distributively averaged without using a fusion center, requiring communication amongst connected nodes only. Simulations show that we achieve a good trade-off between communicational load and localization performance.

The remainder of the paper is organized as follows. Section 2 describes the conventional SRP algorithm. In Section 3, we introduce the proposed distributed approaches. In Section 4, we present the results of simulations comparing the proposed approaches to the conventional SRP method and finally Section 5 concludes the paper.

2. STEERED RESPONSE POWER

SRP is a beamforming-based approach where a microphone array is steered towards a set of candidate source locations searching for the true source location. The location with maximum power of the output of the DSB is used as the estimate of the source position[19].

Let $M \in \mathbb{N}$ be the number of microphones and let $y_m(t)$ with $m = 1, \ldots, M$ denote the m^{th} microphone signal in the time domain. The number of microphone pair combinations is given by $P = M(M-1)/2$.

This research work was carried out at the ESAT Laboratory of KU Leuven, and has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369 and from the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

978-1-6654-6867-1/22/$31.00 ©2022 IEEE

In the frequency domain, with ω the radial frequency, the GCC matrix $\mathbf{\Psi}(\omega) \in \mathbb{C}^{M \times M}$ is defined as

$$[\mathbf{\Psi}(\omega)]_{m,m'} = \psi_{m,m'}(\omega) = \frac{Y_m(\omega) Y^*_{m'}(\omega)}{|Y_m(\omega) Y^*_{m'}(\omega)|} \tag{1}$$

where $Y_m(\omega)$ and $Y_{m'}(\omega)$ are the Short Time Fourier Transform (STFT) of a pair of microphone signals and $*$ is the complex conjugation operator. Here, a PHAT weighting function $1/|Y_m(\omega)Y^*_{m'}(\omega)|$ is used, so that only the phase information remains in order to improve spatial resolution [1, 20].

Then, we can define the frequency-dependent SRP map as [15]

$$\begin{aligned} \text{SRP}(\omega, i) &= \mathbf{h}^H(\omega, i)\mathbf{\Psi}(\omega)\mathbf{h}(\omega, i) - \text{tr}[\mathbf{\Psi}(\omega)] \\ &= 2 \sum_{(m,m'):m>m'} \Re[\psi_{m,m'}(\omega)e^{j\omega\Delta t_{m,m'}(i)}] \end{aligned} \tag{2}$$

Here, $()^H$, $\Re()$, tr[ ], $\psi_{m,m'}(\omega)$ denote the Hermitian, the real part, the trace and the frequency-domain GCC respectively and $\mathbf{h}(\omega, i)$ is the DSB steering vector towards the candidate location i with $i = 1, \ldots, J$ and relative to the first microphone it is defined as [19]

$$\mathbf{h}(\omega, i) = [1 \; e^{j\omega\Delta t_{2,1}(i)} \ldots \; e^{j\omega\Delta t_{M,1}(i)}] \tag{3}$$

where $\Delta t_{m,m'}(i)$ is the TDOA observed at the microphone pair m, m' from a sound source located at $\mathbf{q}_i$ and is given by

$$\Delta t_{m,m'}(i) = \frac{(\|\mathbf{q}_i - \mathbf{p}_m\| - \|\mathbf{q}_i - \mathbf{p}_{m'}\|)}{c} \tag{4}$$

where $\mathbf{q}_i$ is the Cartesian coordinate vector of the i^{th} candidate location and $\mathbf{p}_m$ is the Cartesian coordinate vector of the m^{th} microphone position and c is the speed of sound. Commonly, tr$[\mathbf{\Psi}(\omega)]$ is subtracted here, as it creates an offset and does not contribute to the purpose of spatial mapping [1, 15]. The summation in the SRP functional is taken over all combinations of P microphone pairs (m, m') with $m > m'$.

By integrating SRP(ω, i) over frequency, one can obtain the broadband SRP value

$$\text{SRP}(i) = \int_{-\omega_0}^{\omega_0} \text{SRP}(\omega, i)\, d\omega \tag{5}$$

After calculating the broadband SRP value for each candidate location, the location of the source can be estimated by finding the location index with the highest SRP. So that

$$\hat{i} = \underset{i}{\text{argmax}} \; \text{SRP}(i). \tag{6}$$

This method will be referred to as the conventional SRP method from now on.

Given the frequency-domain GCC, $\psi_{m,m'}$ at K frequency bins, the computation of the conventional SRP requires JPK complex multiplications [15].

3. DISTRIBUTED SRP METHODS

In this work, our primary aim is to employ the SRP approach in the context of WASNs. In some applications, the nodes of a WASN can transmit their recorded microphone signals to a dedicated device which is also known as the fusion center. All signals are processed in the fusion center which results in a network with a centralized topology. In some other applications however a distributed approach is preferred where the nodes can make local processing and share their data only with their neighbouring nodes, not with the fusion center. This approach is very useful if no fusion center is available, if it is too far or if the compounded data of all microphone signals is too large to process in a single device [18].

We propose two SRP approaches that correspond to these two different parameter fusion strategies for WASNs.

3.1. All-node SRP

Let $N \in \mathbb{N}$ be the number of nodes and let $n = 1, \ldots, N$ denote the n^{th} node. In this method, we compute the SRP maps as in (2), but instead of taking the sum over all microphone pair combinations, we compute the SRP maps using all possible microphone pairs within that node. We compute this for every candidate location in the grid, and we obtain individual power maps for each node, which will be denoted as a "single-node SRP map" throughout this paper.

Let A_n be the set that contains all microphone pairs within the node n. Then

$$\text{SRP}_{\text{singlenode}}(\omega, i, n) = 2 \sum_{(m,m') \in A_n} \Re[\psi_{m,m'}(\omega)e^{j\omega\Delta t_{m,m'}(i)}] \tag{7}$$

By integrating SRP(ω, i) over frequency as in (5), one can obtain the broadband single-node SRP value:

$$\text{SRP}_{\text{singlenode}}(i, n) = \int_{-\omega_0}^{\omega_0} \text{SRP}_{\text{singlenode}}(\omega, i, n)\, d\omega \tag{8}$$

Next, we define a centralized SRP map computed from all nodes. For this, we take the sum of single-node SRP maps over all the nodes and name it "all-node SRP map",

$$\text{SRP}_{\text{allnode}}(i) = \sum_{n=1}^{N} \text{SRP}_{\text{singlenode}}(i, n). \tag{9}$$

The location of the source can be estimated by finding the location index with the highest SRP. So that

$$\hat{i} = \underset{i}{\text{argmax}} \; \text{SRP}_{\text{allnode}}(i). \tag{10}$$

This approach corresponds to a scenario in which the WASN has a fusion center where the summation of the single-node SRP maps can be carried out. Also the conventional SRP method can be used in a WASN with a fusion center, requiring each node to transmit its microphone signals to the fusion center. The proposed approach is then an alternative to the conventional SRP method requiring less communication from the nodes to the fusion center as the single-node SRP maps typically have a lower dimension than the microphone signals.

Given the frequency-domain GCC, $\psi_{m,m'}$ at K frequency bins, and P_n microphone pairs in node n, the computation of the all-node SRP requires JP_nK complex multiplications per node.

3.2. Multi-node SRP

The second approach is proposed for a scenario without a fusion center in which only the neighbouring nodes will communicate and share the SRP information with each other according to some topology. Instead of computing (9) directly, we approximate the all-node SRP map by means of a linear iteration that computes the average of the single-node SRP maps in a distributed manner. A fast distributed linear averaging (FDLA) algorithm has been proposed in [21], in which firstly an optimal (symmetric) weighting matrix

$\mathbf{W}_{\text{FDLA}}$ is calculated by solving the following convex optimization problem [21, 22]:

$$\mathbf{W}_{\text{FDLA}} = \underset{\mathbf{W}}{\text{argmin}} \, \|\mathbf{W} - \mathbf{1}\mathbf{1}^T/N\| \\ \text{s.t. } \mathbf{W} \in S(\mathbf{C}), \mathbf{1}^T\mathbf{W} = \mathbf{1}^T, \mathbf{W}\mathbf{1} = \mathbf{1} \quad (11)$$

where $\mathbf{1}$ denotes a length-N column vector with all elements equal to one and $S(\mathbf{C})$ is the class of matrices having the same sparsity pattern as the sensor connectivity matrix $\mathbf{C}$ in which the nodes that are connected correspond to a value 1 and the nodes that are not connected correspond to a value of 0. The resulting matrix $\mathbf{W}_{\text{FDLA}}$ will contain non-zero weights only for those nodes that are connected in the WASN.

We can now form our new map which will be denoted as a "multi-node SRP map" by multiplying the vector formed from the single-node SRP maps with the weight matrix iteratively,

$$\begin{bmatrix} \text{SRP}^{(k+1)}_{\text{multinode}}(i,1) \\ \text{SRP}^{(k+1)}_{\text{multinode}}(i,2) \\ \vdots \\ \text{SRP}^{(k+1)}_{\text{multinode}}(i,N) \end{bmatrix} = \mathbf{W}_{\text{FDLA}} \cdot \begin{bmatrix} \text{SRP}^{(k)}_{\text{multinode}}(i,1) \\ \text{SRP}^{(k)}_{\text{multinode}}(i,2) \\ \vdots \\ \text{SRP}^{(k)}_{\text{multinode}}(i,N) \end{bmatrix} \quad (12)$$

in which the initialization corresponds to the single-node SRP values,

$$\text{SRP}^{(0)}_{\text{multinode}}(i,n) = \text{SRP}_{\text{singlenode}}(i,n) \quad n = 1,\dots,N \quad (13)$$

where k is the iteration number; $k = 1,\dots,k_{\text{max}}$ and the iterative algorithm of (12) is run for each candidate location index i separately.

The location of the source can be estimated by finding the location index with the highest SRP for each node. So that

$$\hat{i} = \underset{i}{\text{argmax}} \, \text{SRP}_{\text{multinode}}(i,n) \quad n = 1,\dots,N \quad (14)$$

This approach, in which the single-node SRP maps are distributively averaged and a multi-node SRP map is formed, is a fusion strategy that does not require a fusion center and relies only on communication of single-node SRP maps amongst neighboring nodes.

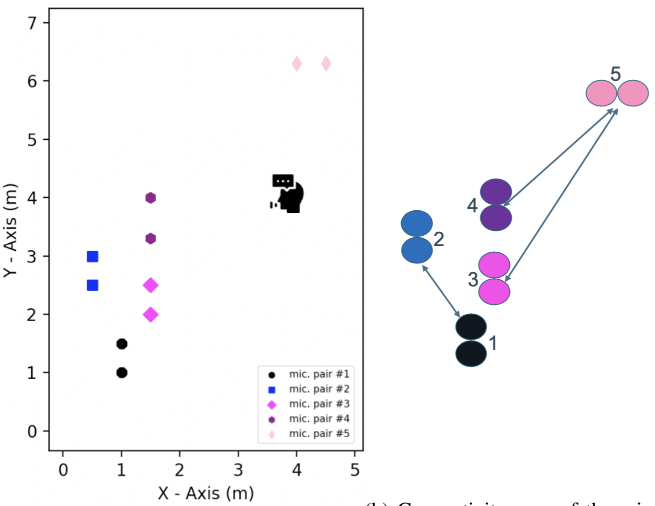
Given the frequency-domain GCC, $\psi_{m,m'}$ at K frequency bins, and P_n microphone pairs in node n, let C_n be the number of connected nodes in that node and N_{it} be the number of iterations. Then the computation of the multi-node SRP requires JP_nK complex multiplications and JC_nN_{it} real multiplications per node.

4. SIMULATIONS

In this section, we consider a scenario in which the WASN consists of N = 5 nodes, and each node consists of a microphone pair hence M = 2N = 10. We compare the localization errors of the all-node SRP, the multi-node SRP and the conventional SRP method. The acoustic scenario is modelled using the randomized image method [23] at a sampling rate of 16 kHz. We simulate a shoe-box room of 5 × 7 x 5 m with 600 ms reverberation time, wherein 5 microphone pairs, as well as a single source at coordinates (4,4,1) m, are placed. Female speech is used as the source signal [24]. Additive white Gaussian noise (AWGN) is added at signal-to-noise ratios (SNRs) of SNR $= [-3, 0, 3, 6, 12, 24, 48, 60]$ dB. Fig. 1a shows a top view of our acoustic setup.

Fig. 2a shows the conventional SRP map which is computed from the GCCs of all microphone pair combinations; that is $M(M-1)/2$ GCCs.

As explained in Section 3.1, single-node SRP maps are computed for each microphone pair and Fig. 2b shows an example with



(a) Top view of the position of the microphone pairs and the source

(b) Connectivity map of the microphone pairs considered in the simulation scenario

Fig. 1: Acoustic Setup

an SNR = 6 dB. The leftmost plot shows the map obtained from the first microphone pair only and the second from left shows the map obtained from the second microphone pair only etc. With these maps only, it is not possible to locate the source. However if we sum the single-node SRP values obtained from all 5 nodes, a fairly accurate localization may be possible as seen in Fig. 2c.

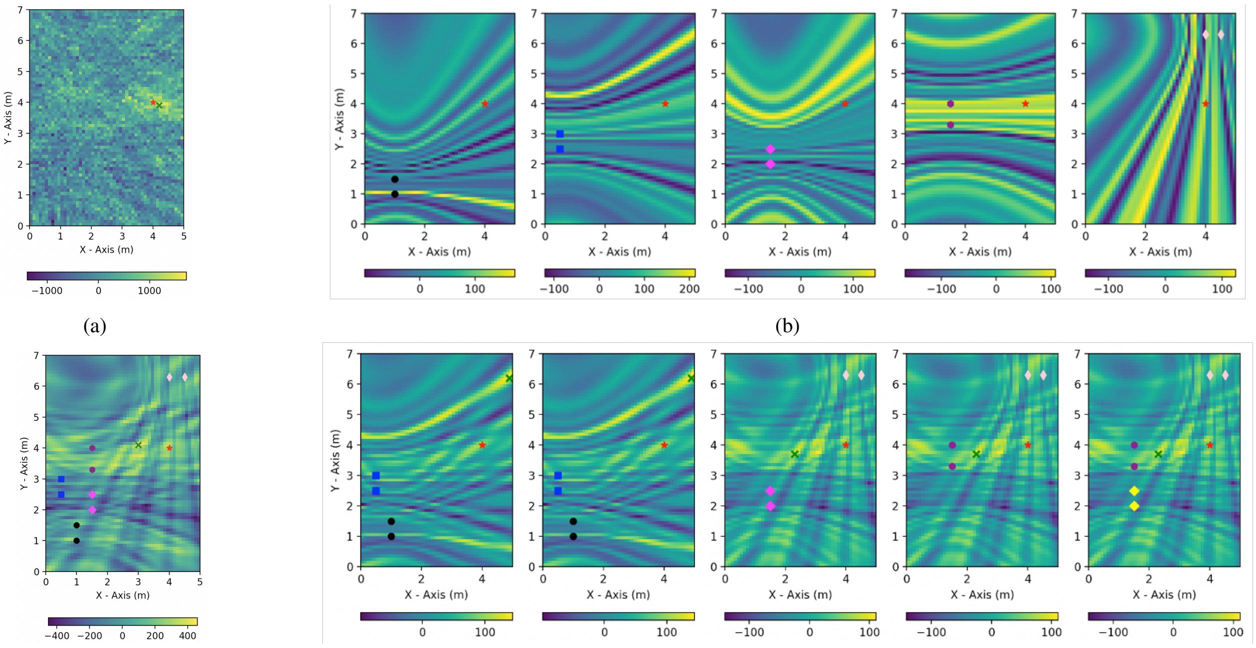
In section 3.2, a multi-node SRP approach has been proposed, for which we must first define the connectivity matrix, $\mathbf{C}$ of the microphone pairs. There are many possible configurations and Fig. 1b shows one possible configuration that we used in our simulations. In this configuration, the first and second pair are connected and the fifth pair is connected to both the fourth and the third pair.

Fig. 2d shows the resulting multi-node SRP maps for each node obtained after 50 iterations of (12). We observe that the nodes that are connected converge to the same SRP value. This way, we can achieve source localization based only on communication of single-node SRP maps amongst neighboring nodes.

Figure 3 shows the localization errors and their averages for 50 realizations of AWGN added to the female speech with SNR $= [-3, 0, 3, 6, 12, 24, 48, 60]$ dB. The root-mean-square-error (RMSE) is used as a performance metric. The conventional SRP approach gives the best localization performance. Compared with the conventional SRP approach the proposed approaches show acceptable errors and we achieve a good trade-off between communicational load and localization performance.

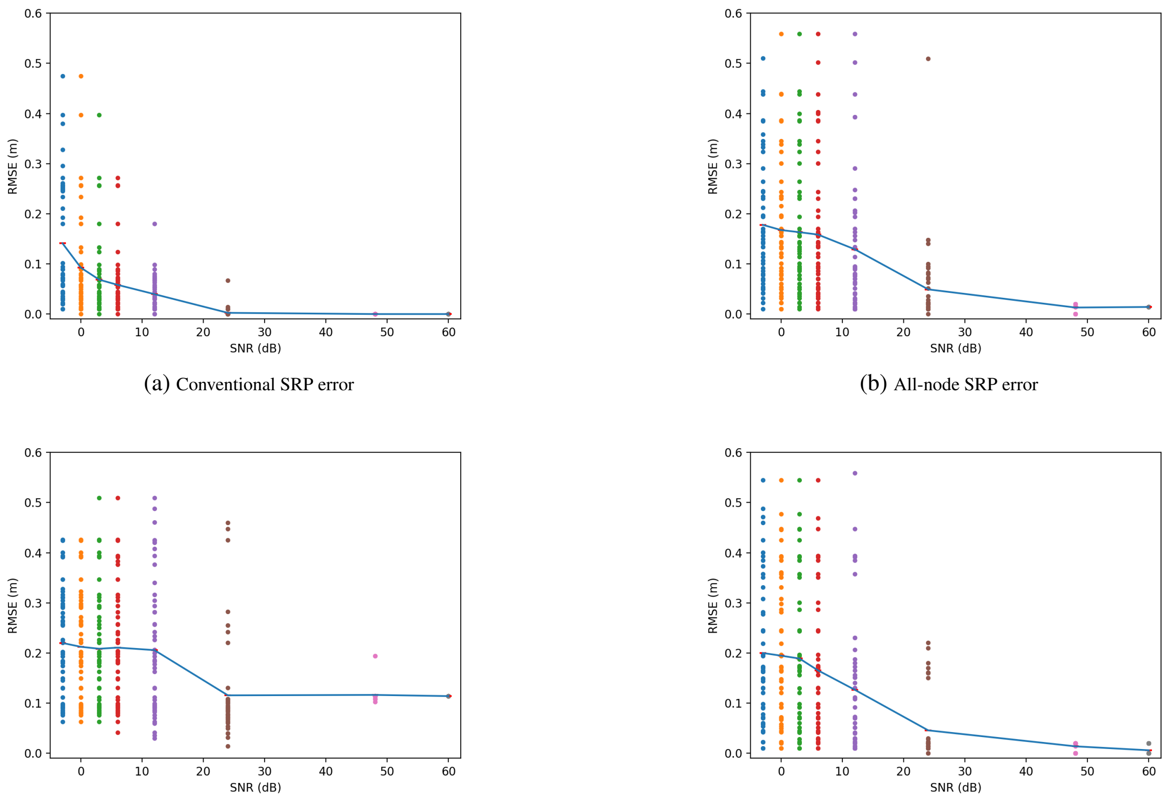
5. CONCLUSION

In this work, instead of using a microphone array in a centralized fashion, we have proposed two methods for source localization by using distributed microphones in a WASN. Both methods are based on the SRP algorithm. The first method applies to a WASN which has a fusion center and the second method can be used in a WASN that does not require a fusion center and relies only on the communication of single-node SRP maps amongst neighboring nodes. Simulations comparing the proposed methods and the conventional SRP show low errors in the localization and we achieve a good trade-off between communicational load and localization performance. A future work would involve investigating the communicational load in more detail.



(a) (b) (c) (d)

Fig. 2: SRP maps (an example with SNR = 6 dB) using (a) Conventional SRP, (b) Single-node SRPs, (c) All-node SRP, (d) Multi-node SRPs of each node $n = 1, \ldots, N$ from left to right. Red star: source position. Green cross: estimated source position. Legends for the microphone pairs are the same as Fig. 1a



(a) Conventional SRP error

(b) All-node SRP error

(c) Multi-node SRP error (nodes 1, 2)

(d) Multi-node SRP error (nodes 3, 4, 5)

Fig. 3: SRP error for 50 AWGN realizations (dots) and RMSE (solid line)

6. REFERENCES

[1] Joseph Hector DiBiase, *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*, Ph.D. thesis, Brown University, Providence, Rhode Island, USA, 2000.

[2] Maximo Cobos, Fabio Antonacci, Anastasios Alexandridis, Athanasios Mouchtaris, and Bowon Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," *Wireless Communications and Mobile Computing*, vol. 2017, 2017.

[3] J.-M. Valin, F. Michaud, J. Rouat, and D. Letourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003) (Cat. No.03CH37453)*, 2003, vol. 2, pp. 1228–1233 vol.2.

[4] Despoina Pavlidi, Anthony Griffin, Matthieu Puigt, and Athanasios Mouchtaris, "Real-time multiple sound source localization and counting using a circular microphone array," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 10, pp. 2193–2206, 2013.

[5] Despoina Pavlidi, Matthieu Puigt, Anthony Griffin, and Athanasios Mouchtaris, "Real-time multiple sound source localization using a circular microphone array based on single-source confidence measures," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 2625–2628.

[6] Sergei Astapov, Julia Berdnikova, and Jürgo-Sören Preden, "A two-stage approach to 2d doa estimation for a compact circular microphone array," in *2015 International Conference on Informatics, Electronics Vision (ICIEV)*, 2015, pp. 1–6.

[7] Parham Aarabi, "The fusion of distributed microphone arrays for sound localization," *EURASIP Journal on Advances in Signal Processing*, vol. 2003, no. 4, pp. 1–10, 2003.

[8] Luca Comanducci, Federico Borra, Paolo Bestagini, Fabio Antonacci, Stefano Tubaro, and Augusto Sarti, "Source localization using distributed microphones in reverberant environments based on deep learning and ray space transform," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2238–2251, 2020.

[9] Stijn Kindt, Alexander Bohlender, and Nilesh Madhu, "2d acoustic source localisation using decentralised deep neural networks on distributed microphone arrays," in *Speech Communication; 14th ITG Conference*. VDE, 2021, pp. 1–5.

[10] Dmitry N Zotkin and Ramani Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 499–508, 2004.

[11] Amparo Marti, Maximo Cobos, Jose J Lopez, and Jose Escolano, "A steered response power iterative method for high-accuracy acoustic source localization," *The Journal of the Acoustical society of America*, vol. 134, no. 4, pp. 2627–2630, 2013.

[12] Hoang Do, Harvey F Silverman, and Ying Yu, "A real-time srp-phat source location implementation using stochastic region contraction (src) on a large-aperture microphone array," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*. IEEE, 2007, vol. 1, pp. I–121.

[13] Maximo Cobos, Amparo Marti, and Jose J Lopez, "A modified srp-phat functional for robust real-time sound source localization with scalable spatial sampling," *IEEE Signal Processing Letters*, vol. 18, no. 1, pp. 71–74, 2010.

[14] Jacek P Dmochowski, Jacob Benesty, and Sofiene Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 8, pp. 2510–2526, 2007.

[15] Thomas Dietzen, Enzo De Sena, and Toon van Waterschoot, "Low-complexity steered response power mapping based on nyquist-shannon sampling," in *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2021, pp. 206–210.

[16] Markus VS Lima, Wallace A Martins, Leonardo O Nunes, Luiz WP Biscainho, Tadeu N Ferreira, Maurício VM Costa, and Bowon Lee, "Efficient steered-response power methods for sound source localization using microphone arrays," *arXiv preprint arXiv:1407.2351*, 2014.

[17] Zhao Xiaoyan, Tang Jie, Zhou Lin, and Wu Zhenyang, "A fast search method of steered response power with small-aperture microphone array for sound source localization [j]," *Journal of Electronics (China)*, vol. 5, pp. 483–490, 2013.

[18] Alexander Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," in *2011 18th IEEE symposium on communications and vehicular technology in the Benelux (SCVT)*. IEEE, 2011, pp. 1–6.

[19] Michael Brandstein and Darren Ward, *Microphone arrays: signal processing techniques and applications*, Springer Science & Business Media, 2001.

[20] Cha Zhang, Dinei Florêncio, and Zhengyou Zhang, "Why does phat work well in lownoise, reverberative environments?," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 2565–2568.

[21] Lin Xiao and Stephen Boyd, "Fast linear iterations for distributed averaging," *Systems & Control Letters*, vol. 53, no. 1, pp. 65–78, 2004.

[22] Toon van Waterschoot and Marc Moonen, "Distributed estimation and equalization of room acoustics in a wireless acoustic sensor network," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*. IEEE, 2012, pp. 2709–2713.

[23] Enzo De Sena, Niccolo Antonello, Marc Moonen, and Toon van Waterschoot, "On the modeling of rectangular geometries in room acoustic simulations," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 4, pp. 774–786, 2015.

[24] Bang and Olufsen, "Music for archimedes," Compact Disc B&O, 1992.