

## Review Article

# A Survey of Sound Source Localization Methods in Wireless Acoustic Sensor Networks

Maximo Cobos,<sup>1</sup> Fabio Antonacci,<sup>2</sup> Anastasios Alexandridis,<sup>3,4</sup>  
Athanasios Mouchtaris,<sup>3,4</sup> and Bowon Lee<sup>5</sup>

<sup>1</sup>Department of Computer Science, Universitat de València, 46100 Burjassot, Spain

<sup>2</sup>Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milano, Italy

<sup>3</sup>Institute of Computer Science (ICS), Foundation for Research & Technology-Hellas (FORTH), Heraklion, 70013 Crete, Greece

<sup>4</sup>Department of Computer Science, University of Crete, Heraklion, 70013 Crete, Greece

<sup>5</sup>Department of Electronic Engineering, Inha University, Incheon 22212, Republic of Korea

Correspondence should be addressed to Maximo Cobos; maximo.cobos@uv.es

Received 18 May 2017; Accepted 28 June 2017; Published 17 August 2017

Academic Editor: Álvaro Marco

Copyright © 2017 Maximo Cobos et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Wireless acoustic sensor networks (WASNs) are formed by a distributed group of acoustic-sensing devices featuring audio playing and recording capabilities. Current mobile computing platforms offer great possibilities for the design of audio-related applications involving acoustic-sensing nodes. In this context, acoustic source localization is one of the application domains that have attracted the most attention of the research community along the last decades. In general terms, the localization of acoustic sources can be achieved by studying energy and temporal and/or directional features from the incoming sound at different microphones and using a suitable model that relates those features with the spatial location of the source (or sources) of interest. This paper reviews common approaches for source localization in WASNs that are focused on different types of acoustic features, namely, the energy of the incoming signals, their time of arrival (TOA) or time difference of arrival (TDOA), the direction of arrival (DOA), and the steered response power (SRP) resulting from combining multiple microphone signals. Additionally, we discuss methods not only aimed at localizing acoustic sources but also designed to locate the nodes themselves in the network. Finally, we discuss current challenges and frontiers in this field.

## 1. Introduction

With the rapid development in fields like circuit design and manufacturing, wireless nodes incorporating a variety of sensors, communication interfaces and compact microprocessors have become economical resources for the design of innovative monitoring systems. Networks of such type of devices, referred to as wireless sensor networks (WSNs) [1, 2], have been widely spread and used in many fields, with applications ranging from surveillance and military deployments to industrial and health-care systems [3]. When the nodes incorporate acoustic transducers and the processing involves the manipulation of audio signals, the resulting network is usually referred to as a wireless acoustic sensor network (WASN). A WASN consists of a set of sensor nodes

interconnected via a wireless medium [4]. Each node has one or several sensors (microphones), a processing unit, a wireless communication module, and, sometimes, also one or several actuators (loudspeakers) [5].

During the last decade, the use of location information and its potentiality in the development of ambient intelligence applications has promoted the design of local positioning systems with WSNs [6]. Using WSNs to perform localization tasks has always been a desirable property since, besides being considerably cheap, they are easily deployable. Localization and ranging in WSNs have been typically addressed by measuring the received signal strength (RSS) or time of arrival (TOA) of radio signals [7]. However, the RSS approach, while being significantly inexpensive, incurs significant errors due to channel fading, long distances,

and multipath. In the context of acoustic signal processing, WASNs also provide advantages with respect to traditional (wired) microphone arrays [8]. For example, they enable increased spatial coverage by distributing microphone nodes over a larger volume, a scalable structure, and possibly better signal-to-noise ratio (SNR) properties. In fact, since the ranging accuracy depends on both the signal propagation speed and the precision of the TOA measurement, acoustic signals may be preferred with respect to radio signals [9].

There are two typical localization tasks in WASNs: the localization of one or more sound sources of interest and the localization of the nodes that make up the network. The first case may involve locating the position of unknown sound sources, for example, talkers inside a room or unexpected acoustic events or sometimes other devices emitting known beacon signals. The second case is usually related to the self-calibration or automatic ranging of the nodes themselves.

To estimate the locations of the sound sources that are active in an acoustic environment monitored by a WASN, different methods exist in the literature. Usually, a centralized scheme is adopted where a dedicated node, known as the fusion center, is responsible for performing the localization task based on information it receives from the rest of nodes. The sensor network itself poses many limitations and challenges that must be considered when designing a localization approach in order to facilitate its use in practical scenarios. Such challenges include the bandwidth usage which limits the amount of information that can be transmitted in the network and the limited processing power of the nodes which prohibits them from carrying out very complex and computationally expensive operations. Moreover, each node has its own clock for sampling the signals and since the nodes operate individually, the resulting audio in the network will not be synchronized.

A taxonomy of sound source localization methods can be built up upon the nature of information from the sensors that it is utilized in order to estimate the locations. Hence, the WASN can estimate the locations of the acoustic sources based on (i) energy readings, (ii) time-of-arrival (TOA) measurements, (iii) time-difference-of-arrival (TDOA) measurements, and (iv) direction-of-arrival estimates or (v) by utilizing the steered response power (SRP) function.

In DOA-based approaches, each node estimates the DOA of the sources it can detect and transmits the DOA estimates to the fusion center. Although such approaches require increased computational power and multiple microphones in each node, they can attain very low-bandwidth usage, as only the DOA estimates need to be transmitted. Also, since the DOA estimation is carried out in each node individually, the audio signals at different nodes need not be synchronized: DOA-based approaches can tolerate unsynchronized input as long as the sources move at a rather slow rate relative to the analysis frame. The location estimators are generally based on estimating the location of a sound source as the intersection of lines emanating from the nodes at the direction of the estimated DOA. However, for multiple sources several challenges arise: the number of detected sources (and thus the number of DOA estimates) in each time instant can vary across the nodes due to missed detections (i.e., a source is not

detected by a node) or false-alarms (i.e., overestimation of the number of detected sources) and an association procedure is needed to find the DOA combinations that correspond to the same source. This is known as the data-association problem and is crucial for the localization task.

The TDOA is related to the difference in the time of flight (TOF) of the wavefront produced by the source at a pair of microphones in the same node. TDOAs can be estimated at a moderate computational cost through the generalized cross correlation (GCC) [10] of the signals acquired by microphones in the pair. The source location estimate is accomplished by combining TDOA measurements coming from multiple sensors. Notice that, as for the DOA, only the TDOA measurements must be transmitted over the wireless network, with clear advantages in terms of transmission power and required bandwidth. Though suitable for WASNs, in practical scenarios (reverberant environments, presence of noise, and interferers), TDOA measurements are prone to errors, which in their turn lead to wrong localization. In order to mitigate the impact of these adverse phenomena, several techniques have been presented with the aim of identifying and removing outliers in the TDOA set [11–13].

A TDOA measurement bounds the source to lie on a branch of hyperbola whose vertices are in microphone positions and whose aperture is determined by the TDOA value. When two (three in 3D) measurements from different pairs are available, the source can be localized through intersection of hyperbolas. The resulting cost function, however, is strongly nonlinear, and therefore its minimization is difficult and prone to errors. Linearized cost functions have been proposed to overcome this difficulty [14–16]. It is important to notice, however, that the linearized cost functions require the presence of a reference microphone, with respect to which the remaining microphones in all the sensors must be synchronized. This poses technological constraints that, in some cases, hinder the use of such techniques. More recently, methodologies that include the synchronization offset in the optimization of the cost function have been proposed to overcome this problem [12, 17].

TOA measurements are obtained by detecting the time instant at which the source signal arrives at the microphones present in the network. Since in passive source localization the source emission time is unknown, the TOA is not equivalent to the TOF of the signal, preventing a direct mapping from TOAs to source-to-node distances. While some applications involve the use of sound-emitting nodes that allows performing localization by using trilateration techniques, TDOA localization methods are usually chosen. In this case, although the source emission time does not need to be known, the registered TOAs need to be referenced to a common clock, requiring precise timing hardware and synchronization mechanisms.

Energy-based localization relies on the averaged energy readings computed over windows of signal samples acquired by the microphones incorporated by the nodes [18]. Compared to TDOA and DOA methods, energy-based approaches are attractive because they do not require the use of multiple microphones at the nodes and are free of synchronization issues unlike those based on TOA.

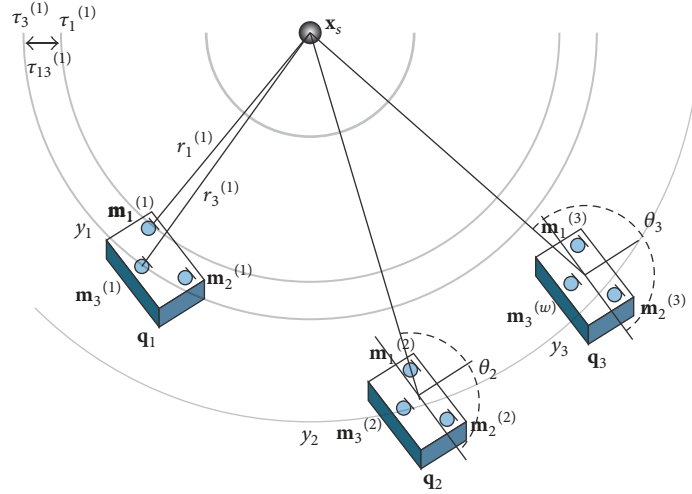


FIGURE 1: WASN with  $M = 3$  nodes and  $N = 3$  microphones per node.

However, TDOA- and DOA-based methods, considered as signal-based approaches, offer generally better performance than energy-based methods. This is due to the fact that the information conveyed by all the samples of the signal is directly exploited instead of their average, at the expense of more sophisticated capturing devices and transmission resources [18, 19].

SRP approaches are beamforming-based techniques that compute the output power of a filter-and-sum beamformer steered to a set of candidate source locations defined by a predefined spatial grid. Since the computation of the SRP involves the accumulation of GCCs from multiple microphone pairs, the synchronization requirements are usually the same as those of TDOA-based methods. The set of SRPs obtained at the different points of the grid make up the SRP power map, where the point accumulating the highest value corresponds to the estimated source location. When using unsynchronized nodes with multiple microphones, the SRP power maps computed at each node can be used to obtain their corresponding DOAs. Alternatively, the SRP power maps from the different sensors can be accumulated at a central node to obtain a combined SRP power map, identifying the true source location by its maximum.

Besides the localization of acoustic sources, approaches for localizing the nodes in the network are also of high interest within a WASN context. Based on the estimated TOAs and TDOAs, algorithms for self-localization of the sensor nodes usually assume known source positions playing known probe signals. In practical scenarios, each sensor node does not have any information regarding other nodes or synchronization between the sensor and the source. These assumptions allow all processing to take place on the node itself. Several issues, for example, reverberation, asynchrony between the sound source and the sensor, poor estimation of the speed of sound, and noise, need to be considered for robust self-localization methods. In addition, the processing needs to be computationally inexpensive in order to be run on the sensor node itself.

Some state-of-the-art solutions for acoustic sensor localization detail the challenges facing these algorithms and methods to tackle such problems in a unified context [20, 21]. Furthermore, recent methodologies have been proposed for probe signal design aimed at improving TOF estimation [22], the joint localization of sensors and sources in an ad hoc array by using low-rank approximation methods [23], and an iterative peak matching algorithm for fast node autocalibration [24].

The paper is structured as follows. Section 2 discusses some general considerations regarding a general WASN structure and the notation used throughout this paper. Section 3 presents the fundamentals of energy-based source localization methods. Section 4 discusses TOA-based localization approaches. Methods for TDOA-based localization are presented in Section 5. Section 6 discusses the use of DOA measurements to perform localization of one or several sound sources. In Section 7, the fundamentals of the conventional and modified SRP methods are explained. Section 8 reviews some recent methods for the self-localization of the nodes in the network. Some future directions in the field are discussed in Section 9. Finally, the conclusions of this review are summarized in Section 10.

## 2. General Considerations

In order to clarify the notation used throughout this paper and the type of location cues used to perform the localization task, Figure 1 shows a general WASN with a set of wireless nodes and an emitting sound source. It is assumed that the network consists of  $M$  nodes and that each node incorporates  $N$  microphones. In the example shown in Figure 1,  $M = 3$  and  $N = 3$ . The nodes are assumed to be located at positions  $\mathbf{q}_m = [q_{x,m}, q_{y,m}, q_{z,m}]^T$ ,  $m = 1, \dots, M$ , while the microphone locations are denoted as  $\mathbf{m}_i^{(m)} = [x_i^{(m)}, y_i^{(m)}, z_i^{(m)}]^T$ ,  $i = 1, \dots, N$ , where the superscript  $(m)$  identifies the node at which the microphone is located. The source position is denoted

as  $\mathbf{x}_s = [x_s, y_s, z_s]^T$ , while a general point in space is  $\mathbf{x} = [x, y, z]^T$ . Note that all these location vectors are referenced to the same absolute coordinate system. The distance from any microphone to the sound source is denoted as  $r_i^{(m)}$ , while the time it takes the sound wave to travel from the source to a microphone, that is, the time of flight (TOF), is denoted by  $\eta_i^{(m)}$ . The time instant at which the source signal arrives to a given microphone, that is, the TOA, is denoted as  $\tau_i^{(m)}$ . TDOAs are denoted by  $\tau_{ij}^{(m)}$  and correspond to the observed TOA differences between pairs of microphones ( $ij$ ). It is a common practice to identify different pairs of microphones by using an index  $p = 1, \dots, P$ , where  $P$  is the total number of microphone pairs involved in the localization task. The DOA corresponds to the angle that identifies the direction relative to the node microphone array that points to the sound source and is denoted by  $\theta_m$ . Finally, the energy values of the source signal measured at the nodes are denoted as  $y_m$ . These are negligible for microphones located at the same node (especially if the nodes are relatively far from the source), so it is usually assumed that only one energy reading is obtained for each node.

### 3. Energy-Based Source Localization

Traditionally, most localization methods for WASNs have been focused on sound energy measurements. This is motivated by the fact that the acoustic power emitted by targets usually varies slowly in time. Thus, the acoustic energy time series does not require as high a sampling rate as the raw signal time series, avoiding the need for high transmission rates and accurate synchronization. The energy-based model was first presented in [25]. In this model, the acoustic energy decreases approximately as the inverse of the square of the source to sensor distance. Without loss of generality, it will be assumed in this section that the node locations determine the microphone positions and that nodes only incorporate one microphone ( $M = N$ ). Note that, as opposed to time delay measurements, the differences in energy measurements obtained from different microphones at the same node are negligible.

**3.1. Energy-Decay Model.** Assuming that there is only one source active, the acoustic intensity signal received by the  $i$ th sensor at a time interval  $l$  is modeled as [25]

$$y_i(l) = g_i \frac{I(l - \eta_i)}{r_i^2} + \epsilon_i(l) = s_i(l) + \epsilon_i(l), \quad (1)$$

where  $s_i(l)$  is the source intensity at the sensor location,  $g_i$  is a sensor gain factor,  $I(l)$  denotes the intensity of the source signal at a distance of one meter from the source,  $\eta_i$  is the propagation delay from the source to the sensor,  $r_i$  is the distance between the sensor and the source, and  $\epsilon_i(l)$  is an additive noise component modeled as Gaussian noise. In

practice, for each time interval  $l$ , a set of  $T$  samples is used to obtain an energy reading  $y_i(l)$  at the sensor:

$$y_i(l) \approx \frac{1}{T} \sum_{l-T/2}^{l+T/2} x_i^2(t), \quad (2)$$

where  $x_i(t)$  are the samples obtained from the microphone of the  $i$ th node. In the case when several microphones are available at each node, the final energy reading is obtained by averaging the energies computed from each of the microphones.

By assuming that the maximum propagation delay between any pair of sensors is small compared to  $T$  and taking into account the averaging effect,  $\eta_i$  can be neglected for the energy-decay function, so that

$$y_i(l) \approx g_i \frac{I(l)}{r_i^2} + \epsilon_i(l). \quad (3)$$

**3.2. Energy Ratios.** Considering the energy measurements obtained by a group of  $N$  sensor nodes, the energy ratio  $\kappa_{ij}$  between the  $i$ th and the  $j$ th sensors is defined as

$$\kappa_{ij} \triangleq \left( \frac{y_i/y_j}{g_i/g_j} \right)^{-1/2} = \frac{\|\mathbf{x}_s - \mathbf{m}_i\|}{\|\mathbf{x}_s - \mathbf{m}_j\|} = \frac{r_i}{r_j}, \quad (4)$$

where  $\mathbf{x}_s$  is the location of the source and  $\mathbf{m}_i$  and  $\mathbf{m}_j$  are the locations of the two microphones. For  $0 < \kappa_{ij} \neq 1$ , all the possible source coordinates  $\mathbf{x}$  that satisfy (4) must reside on a hypersphere (sphere if  $\mathbf{x} \in \mathbb{R}^3$  or circle if  $\mathbf{x} \in \mathbb{R}^2$ ) described by

$$\|\mathbf{x} - \mathbf{c}_{ij}\|^2 = \rho_{ij}^2, \quad (5)$$

where the center  $\mathbf{c}_{ij}$  and the radius  $\rho_{ij}$  of this hypersphere are

$$\begin{aligned} \mathbf{c}_{ij} &\triangleq \frac{\mathbf{m}_i - \kappa_{ij}^2 \mathbf{m}_j}{1 - \kappa_{ij}^2}, \\ \rho_{ij} &\triangleq \frac{\kappa_{ij} \|\mathbf{m}_i - \mathbf{m}_j\|}{1 - \kappa_{ij}^2}. \end{aligned} \quad (6)$$

In the limiting case, when  $\kappa_{ij} \rightarrow 1$ , the solution of (4) forms a hyperplane between  $\mathbf{m}_i$  and  $\mathbf{m}_j$ :

$$\mathbf{w}_{ij}^T \mathbf{x} = \psi_{ij}, \quad (7)$$

where  $\mathbf{w}_{ij} \triangleq \mathbf{m}_i - \mathbf{m}_j$  and  $\psi_{ij} \triangleq (1/2)(\|\mathbf{m}_i\|^2 - \|\mathbf{m}_j\|^2)$ .

By using the energy ratios registered at a pair of sensors, the potential target location can be restricted to a hypersphere with center and radius that are functions of the energy ratio and the two sensor locations. If more sensors are used, more hyperspheres can be determined. If all the sensors that receive the signal from the same target are used, the corresponding target location hyperspheres must intersect at a particular point that corresponds to the source location. This is the basic idea underlying energy-based source localization. In the

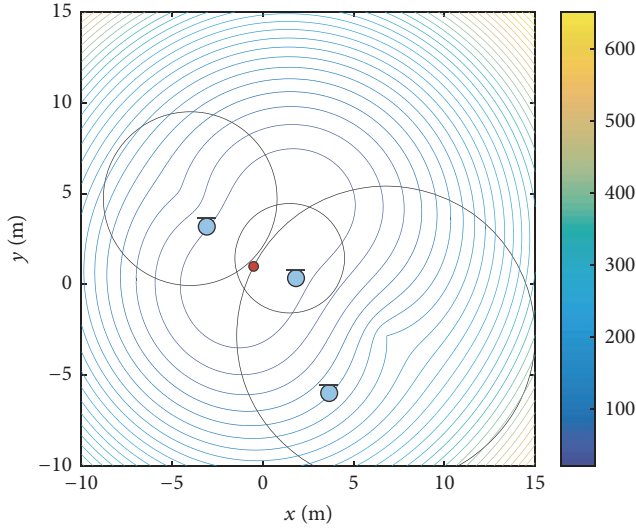


FIGURE 2: Example of energy-based localization with three nodes. The red circle indicates the true source location. Note that due to measurement noise circles do not intersect at the expected source location. The center of the contour plots indicates estimated target location, according to the nonlinear cost function of (9).

absence of noise, it can be shown that for  $N$  measurements only  $N - 1$  of the total  $N(N - 1)/2$  ratios are independent, and all the corresponding hyperspheres intersect at a single point for four or more sensor readings. For noisy measurements, however, more than  $N - 1$  ratios may be used for robustness, and the unknown source location  $\mathbf{x}_s$  is estimated by solving a nonlinear least squares problem, as explained in the next subsection. As an example, Figure 2 shows a 2D setup with three sensors and the circles resulting from noisy energy ratios.

**3.3. Localization through Energy Ratios.** Given  $N$  sensor nodes providing  $P = N(N - 1)/2$  energy ratios, the following least squares optimization problem can be formulated:

$$J_{\text{NLS}}^{(\text{ER})}(\mathbf{x}) = \sum_{p_1=1}^{P_1} \left| \|\mathbf{x} - \mathbf{c}_{p_1}\| - \varrho_{p_1} \right|^2 + \sum_{p_2=1}^{P_2} \left| \mathbf{w}_{p_2}^T \mathbf{x} - \psi_{p_2} \right|^2, \quad (8)$$

$$\hat{\mathbf{x}}_s^{(\text{ER-NLS})} = \arg \min_{\mathbf{x}} J_{\text{NLS}}^{(\text{ER})}(\mathbf{x}), \quad (9)$$

where  $P_1$  is the number of hyperspheres and  $P_2$  is the number of hyperplanes ( $\kappa_{ij}$  close to 1), with corresponding indices  $p_1$  and  $p_2$  indicating the associated sensor pairs ( $ij$ ) ( $P = P_1 + P_2$ ). Note that the above cost function is nonlinear with respect to  $\mathbf{x}$ , resulting in the energy-ratio nonlinear least squares (ER-NLS) problem. It can be shown that minimizing this cost function leads to an approximate solution for the maximum likelihood (ML) estimate. A set of strategies were proposed in [26] to minimize  $J_{\text{NLS}}^{(\text{ER})}(\mathbf{x})$  by using the complete set of measured energy ratios. A popular approach to solve the problem is the unconstrained least squares method. Since every pair of hyperspheres (with double indices  $ij$  replaced by a single pair index for the sake of brevity)  $\|\mathbf{x} - \mathbf{c}_m\|^2 = \varrho_m^2$  and

$\|\mathbf{x} - \mathbf{c}_n\|^2 = \varrho_n^2$ , a hyperplane can be determined by eliminating the common terms:

$$(\mathbf{c}_m - \mathbf{c}_n)^T \mathbf{x} = \frac{(\|\mathbf{c}_m\|^2 - \|\mathbf{c}_n\|^2) - (\varrho_m^2 - \varrho_n^2)}{2}. \quad (10)$$

The combination of (7) with (10) leads to a least squares optimization problem without inconvenient nonlinear terms, known as the energy-ratio least squares (ER-LS) method, with cost function:

$$J_{\text{LS}}^{(\text{ER})}(\mathbf{x}) = \sum_{p_1=1}^{P_1} \left| \mathbf{u}_{p_1}^T \mathbf{x} - \zeta_{p_1} \right|^2 + \sum_{p_2=1}^{P_2} \left| \mathbf{w}_{p_2}^T \mathbf{x} - \psi_{p_2} \right|^2, \quad (11)$$

where

$$\mathbf{u}_{ij} \triangleq \frac{2\mathbf{m}_i}{1 - \kappa_i^2} - \frac{2\mathbf{m}_j}{1 - \kappa_j^2}, \quad (12)$$

$$\zeta_{ij} \triangleq \frac{\|\mathbf{m}_i\|^2}{1 - \kappa_i^2} - \frac{\|\mathbf{m}_j\|^2}{1 - \kappa_j^2}.$$

The closed-form solution of the above unconstrained least squares formulation makes this method computationally attractive; however, it does not reach the Cramer-Rao bound. In [27], energy-based localization is formulated as a constrained least squares problem, and some well-known least squares methods for closed-form TDOA-based location estimation are applied, namely, linear intersection [28], spherical intersection [29], sphere interpolation (SI) [30], and subspace minimization [31]. In [32], an algebraic closed-form solution is presented that reaches the Cramer-Rao bound for Gaussian measurement noise as the SNR tends to infinite. The authors in [33] formulated the localization problem as a convex feasibility problem and proposed a distributed version of the projection onto convex sets method. A discussion of least squares approaches is provided in [19], presenting a low-complexity weighted direct least squares formulation. A recent review of energy-based acoustic source localization methods can be found in [18].

## 4. TOA-Based Localization

Typically, a WASN sound source location setup assumes that there is a sound-emitting source and a collection of fixed microphone anchor nodes placed at known positions. When the sound source emits a given signal, the different microphone nodes will estimate the time of arrival (TOA) or time of flight (TOF) of the sound. These two terms may not be equivalent under some situations. The TOF measures the time that it takes for the emitted signal to travel the distance between the source and the receiving node; that is,

$$\eta_i \triangleq \frac{1}{c} \|\mathbf{m}_i - \mathbf{x}_s\|. \quad (13)$$

In fact, when utilizing TOA measurements for source localization, it is often assumed that the source and sensor nodes cooperate such that the signal propagation time can

be detected at the sensor nodes. However, such collaboration between sources and sensor nodes is not always available. Thus, without knowing the initial signal transmission time at the source, from TOA alone, the sensor is unable to determine the signal propagation time.

In the more general situation when unknown acoustic signals such as speech or unexpected sound events are to be localized, the relation between distances and TOAs can be modeled as follows:

$$\tau_i = \eta_i + t_s + \epsilon_i, \quad (14)$$

where  $t_s$  is an unknown transmission time instant and  $\epsilon_i$  is the TOA measurement noise. Note that the time  $t_s$  appears due to the fact that typical sound sources do not encode a time stamp in their transmitted signal to indicate the starting transmission time to the sensor nodes and, moreover, there is not any underlying synchronization mechanism. Hence, the sensor nodes can only measure the signal arrival time instead of the propagation time or TOF. One way to tackle this problem is to exploit the difference of pairwise TOA measurements, that is, time difference of arrival (TDOA), for source localization (see Section 5). Although the dependence on the initial transmission time is eliminated by TDOA, the measurement subtraction strengthens the noise. To overcome such problems, some works propose methods to estimate both the source location and initial transmission time jointly [34].

When the TOA and the TOF are equivalent (i.e.,  $\tau_i = \eta_i$ ), for example, because there are synchronized sound-emitting nodes, the source-to-node distances can be calculated using the propagation speed of sound [35]. This may require an initial calibration process to determine factors that have a strong influence on the speed of sound. The computation of the source location can be carried out in a central node by using the estimated distances and solving the trilateration problem [36]. Trilateration is based on the formulation of one equation for each anchor that represents the surface of the sphere (or circle) centered at its position, with a radius equal to the estimated distance to the sound source. The solution to this series of equations finds the point where all the spheres intersect. For 2D localization, at least three sensors are needed, while one more sensor is necessary to obtain a 3D location estimate.

**4.1. Trilateration.** Let us consider a set of  $N$  sensor TOA measurements  $\tau_i$  that are transformed to distances by assuming a known propagation speed:

$$r_i = c \cdot \tau_i, \quad (15)$$

where  $c$  is the speed of sound (343 m/s). Then, the following system of equations can be formulated:

$$\|\mathbf{m}_i - \mathbf{x}_s\|^2 = r_i^2 \quad i = 1, \dots, N. \quad (16)$$

Each equation in (16) represents a circle in  $\mathbb{R}^2$  or a sphere in  $\mathbb{R}^3$ , centered at  $\mathbf{m}_i$  with a radius  $r_i$ . Note that the problem is the same as the one given by (5). Thus, solving the above system is equivalent to finding the intersection point/points

of a set of circles/spheres. Again, the trilateration problem is not straightforward to solve due to the nonlinearity of (16) and the errors in  $\mathbf{m}_i$  and  $r_i$  [37]. A number of algorithms have been proposed in the literature to solve the trilateration problem, including both closed form [37, 38] and numerical solutions. Closed-form solutions have low computational complexity when the solution of (16) actually exists. However, most closed-form solutions only solve for the intersection points of  $n$  spheres in  $\mathbb{R}^n$ . They do not attempt to determine the intersection point when  $N > n$ , where small errors can easily cause the involved spheres not to intersect at one point [39]. It is then necessary to find a good approximation that minimizes the errors contained in (16) considering the nonlinear least squares cost function:

$$J_{\text{NLS}}^{(\text{TOA})}(\mathbf{x}) = \sum_{i=1}^N (\|\mathbf{m}_i - \mathbf{x}\| - r_i)^2, \quad (17)$$

$$\hat{\mathbf{x}}_s^{(\text{TOA-NLS})} = \arg \min_{\mathbf{x}} J_{\text{NLS}}^{(\text{TOA})}(\mathbf{x}).$$

Numerical methods are in general necessary to estimate  $\hat{\mathbf{x}}_s^{(\text{TOA-NLS})}$ . However, compared with closed-form solutions, numerical methods have higher computational complexity. Some numerical methods are based on a linearization of the trilateration problem [40–42], introducing additional errors into position estimation. Common numerical techniques such as Newton's method or steepest descent have also been proposed [38, 40, 41]. However, most of these search algorithms are very sensitive to the choice of the initial guess, and a global convergence towards the desirable solution is not guaranteed [39].

**4.2. Estimating TOAs of Acoustic Events.** As already discussed, localizing acoustic sources from TOA measurements only is not possible due to the unknown source emission time of acoustic events. If the sensors are synchronized, the differences of TOA measurements can be used to cancel out the common term  $t_s$ , so that a set of TDOAs are obtained and used as discussed in Section 5. A low-complexity method to estimate the TOAs from acoustic events in WASNs was proposed in [43], where the cumulative-sum (CUSUM) change detection algorithm is used to estimate the source onset times at the nodes. The CUSUM method is a low-complexity algorithm that allows estimating change detection instants by maximizing the following log-likelihood ratio:

$$\hat{\tau}_i = \arg \max_{1 \leq \tau \leq k} \sum_{t=\tau}^k \ln \left( \frac{p(x_i(t), \theta_1)}{p(x_i(t), \theta_0)} \right). \quad (18)$$

The probability density function of each sample is given by  $p(x_i(t), \theta)$ , where  $\theta$  is a deterministic parameter (not to be confused with the DOA of the source). The occurrence of an event is modeled by an instantaneous change in  $\theta$ , so that  $\theta = \theta_0$  before the event at  $t = \tau_i$  and  $\theta = \theta_1$  when  $t \geq \tau_i$ . To simplify calculations at the nodes, the samples before the acoustic event are assumed to belong exclusively to a Gaussian noise component of variance  $\sigma_0^2$ , while the samples after the event are also normally distributed with variance

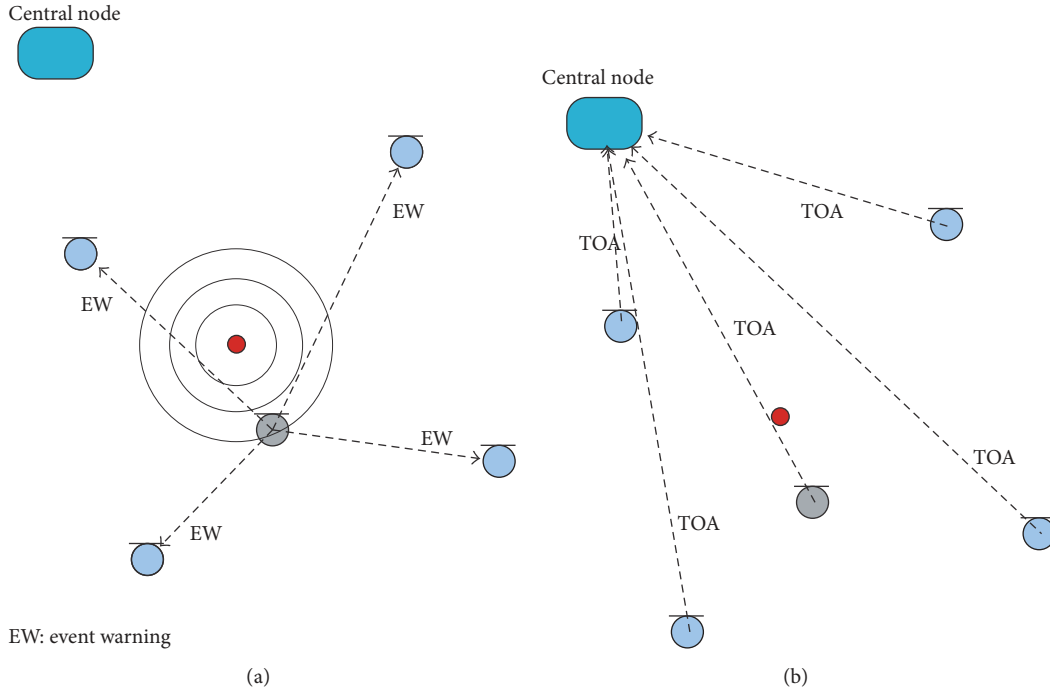


FIGURE 3: Node communication steps in CUSUM-based TOA estimation. (a) One of the nodes detects the sound event and sends an event warning alert to the other nodes. (b) The nodes estimate their TOAs and send it to a central node.

$\sigma_1^2 > \sigma_0^2$ . These variances are estimated from the beginning and tail of a window of samples where the nodes have strong evidence that an acoustic event has happened. The advantage of this approach is twofold. On the one hand, the estimation of the distribution parameters is more accurate. On the other hand, the CUSUM change detection algorithm needs only to be run when an acoustic event has actually occurred, allowing significant battery savings in the nodes. The detection of acoustic events is performed by assuming that, at least, one of the nodes has the sufficient SNR to detect the event by a simple amplitude threshold. The node (or nodes) detecting the presence of the event will notify the rest by sending an event warning alert in order to let them know that they must run the CUSUM algorithm over a window of samples (see Figure 3). The amplitude threshold selection is carried out by setting either the probability of detection or the probability of false alarm given an initial estimate of the ambient noise variance  $\sigma_0^2$ . Note that synchronization issues still persist (all  $\tau_i$  must have a common time reference), so that the nodes exchange synchronization information by using MAC layer time stamping in the deployment discussed in [43].

## 5. TDOA-Based Localization

When each sensor consists of multiple microphones, localization can be accomplished in an efficient way demanding as much as possible of the processing to each node and then combining measurements to yield the localization at the central node. When nodes are connected through low bitrate channels and no synchronization of the internal clocks

is guaranteed, this strategy becomes a must. Among all the possible measurements, a possible solution can be found in the time difference of arrival (TDOA).

**5.1. TDOA and Generalized Cross Correlation.** Consider the presence of  $M$  nodes in the network. For reasons of simplicity in the notation, all nodes are equipped with  $N$  microphones. The TDOA refers to the difference of propagation time from the source location to pairs of microphones. If the source is located at  $\mathbf{x}_s$ , and the  $i$ th microphone in the  $m$ th sensor is at  $\mathbf{m}_i^{(m)}$   $i = 1, \dots, N$ , the TDOA is related to the difference of the ranges from the source to the microphones  $i$  and  $j$  through

$$\tau_{ij}^{(m)} = \frac{r_{ij}^{(m)}}{c} = \frac{\|\mathbf{x}_s - \mathbf{m}_i^{(m)}\| - \|\mathbf{x}_s - \mathbf{m}_j^{(m)}\|}{c}, \quad (19)$$

$$i = 1, \dots, N, \quad j = 1, \dots, N, \quad i \neq j, \quad m = 1, \dots, M.$$

Throughout the rest of this subsection we will consider pairs of microphones within the same node, so we will omit the superscript  $(m)$  of the sensor. The estimate  $\hat{\tau}_{ij}$  of the TDOA  $\tau_{ij}$  can be accomplished performing the generalized cross correlation (GCC) between the signals acquired by microphones at  $\mathbf{m}_i$  and  $\mathbf{m}_j$ , as detailed in the following. Under the assumption of working in an anechoic scenario and in a single source context, the discrete-time signal acquired by the  $i$ th microphone is

$$x_i(t) = \alpha_i s(t - \eta_i) + \epsilon_i(t), \quad i = 1, \dots, N, \quad (20)$$

where  $\alpha_i$  is a microphone-dependent attenuation term that accounts for the propagation losses and air absorption,  $s(t)$

is the source signal,  $\eta_i$  is the propagation delay between the source and the  $i$ th microphone, and  $\epsilon_i(t)$  is an additive noise signal. In the discrete-time Fourier transform (DTFT) domain, the microphone signals can be written as

$$X_i(\omega) = \alpha_i S(\omega) e^{-j\omega\eta_i} + E_i(\omega), \quad i = 1, \dots, N, \quad (21)$$

where  $S(\omega)$  and  $E_i(\omega)$  are the DTFTs of  $s(t)$  and  $\epsilon_i(t)$ , respectively, and  $\omega \in \mathbb{R}$  denotes normalized angular frequency.

Given the pair of microphones  $i$  and  $j$ , with  $i \neq j$ , the GCC between  $x_i(t)$  and  $x_j(t)$  can be written as [10]

$$R_{ij}(\tau) \triangleq \frac{1}{2\pi} \int_{-\pi}^{\pi} X_i(\omega) X_j^*(\omega) \Psi_{ij}(\omega) e^{j\omega\tau} d\omega, \quad (22)$$

where  $X_i(\omega)$  is the DTFT of  $x_i(t)$ ,  $*$  is the conjugate operator, and  $\Psi_{ij}(\omega)$  is a suitable weighting function.

The TDOA from the pair ( $ij$ ) is estimated as

$$\hat{\tau}_{ij} = \frac{\arg \max_{\tau} R_{ij}(\tau)}{F_s}, \quad (23)$$

where  $F_s$  is the sampling frequency. The goal of  $\Psi_{ij}(\omega)$  is to make  $R_{ij}(\tau)$  sharper so that the estimate in (23) becomes more accurate. One of the most common choices is to use the PHASE Transform (PHAT) weighting function; that is,

$$\Psi_{ij}(\omega) = \frac{1}{|X_i(\omega) X_j^*(\omega)|}. \quad (24)$$

In an array of  $N$  microphones, the complete set of TDOAs counts  $N(N-1)/2$  measures. If these are not affected by any sort of measurement error, it can be easily demonstrated that only  $N-1$  of them are independent, the others being a linear combination of them. It is common practice, therefore, to adopt a reference microphone in the array and to measure  $N-1$  TDOAs with respect to it. We refer to the set of  $N-1$  measures as the reduced TDOA set. Without any loss of generality, for the reduced TDOA set, we can assume the reference microphone be with index 1, and the TDOAs  $\hat{\tau}_{1j}$  in the reduced set, for reasons of compactness in the notation, are denoted with  $\hat{\tau}_j$ ,  $j = 2, \dots, N$ .

**5.2. TDOA Measurement in Adverse Environments.** It is important to stress the fact that TDOA measurements are very sensitive to reverberation, noise, and the presence of possible interferers: in a reverberant environment, for some locations and orientation of the source, the peak of the GCC relative to a reflective path could overcome that of the direct path. Moreover, in a noisy scenario, for some time instants the noise level could exceed the signal, thus making the TDOAs unreliable. As a result, some TDOAs must be considered outliers and must be discarded from the measurement set before localization (as in the example shown in Figure 4).

Several solutions have been developed in order to alleviate the impact of outliers in TDOAs. It has been observed that GCCs affected by reverberation and noise do not exhibit a single sharp peak. In order to identify outliers, therefore, some works analyze the shape of the GCC from which

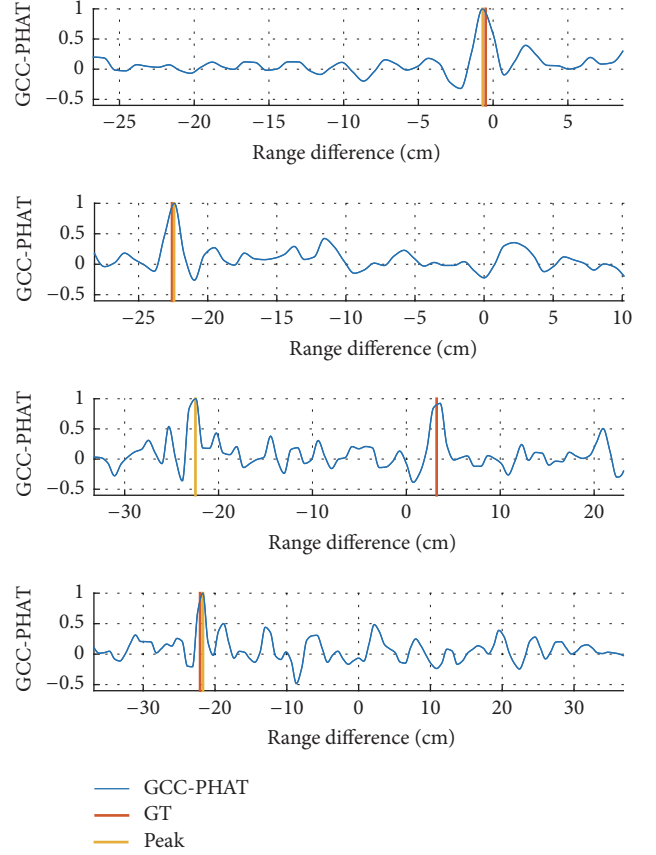


FIGURE 4: Examples of GCCs measured for pairs of microphones within the same node. It is possible to observe that for most of the above GCCs the measured TDOA (Peak) is very close to the ground truth (GT). This does not happen for one measurement, due to reverberation. Note also that TDOA values have been mapped to range differences.

they were extracted. In [12], authors propose the use of the function

$$\rho_{ij} \triangleq \frac{\sum_{\tau \in \mathcal{D}_{ij}} R_{ij}(\tau)^2}{\sum_{\tau \notin \mathcal{D}_{ij}} R_{ij}(\tau)^2}, \quad (25)$$

to detect GCCs affected by outliers. More specifically, the numerator sums the “power” of the GCC samples within the interval  $\mathcal{D}_{ij}$  centered around the candidate TDOA and compares it with the energy of the remaining samples. When  $\rho_{ij}$  overcomes a prescribed threshold, the TDOA is considered reliable. Two metrics of the GCC shape have been proposed in [13, 44]. The first one considers the value of the GCC at the maximum peak location, while the second compares the highest peak with the second highest one. When both metrics overcome prescribed thresholds, the GCC is considered reliable.

Another possible route to follow is described in [11] and is based on the observation that TDOAs along closed paths of microphones must sum to zero (zero-sum condition) and that there is a relationship between the local maxima of the



autocorrelation and cross correlation of the microphone signals (raster condition). The zero-sum condition on minimum length paths of three microphones with indexes  $i$ ,  $j$ , and  $k$ , in particular, states that

$$\tau_{ij} + \tau_{jk} + \tau_{ki} = 0. \quad (26)$$

By imposing zero-sum and raster conditions, authors demonstrate that they are able to disambiguate TDOAs in the case of multiple sources in reverberant environments.

In [17] authors combine different approaches. A redundant set of candidate TDOAs is selected by identifying local maxima of the GCCs. A first selection is operated by discarding TDOAs that do not honor the zero-sum condition. A second selection step is based on three quality metrics related to the shape of the GCC. The third final step is based on the inspection of the residuals of the source localization cost function: all the measurements related to residuals overcoming a prescribed threshold are discarded.

It is important to notice that all the referenced techniques for TDOA outlier removal do not involve the cooperation of multiple nodes, with clear advantages in terms of data to be transmitted.

**5.3. Localization through TDOAs.** In this section we will consider the problem of source localization by combining measurements coming from different nodes. In order to identify the sensor from which measurements have been extracted, we will use the superscript  $(m)$ . From a geometric standpoint, given a TDOA estimate  $\hat{\tau}_{ij}^{(m)}$ , the source is bound to lie on a branch of hyperbola (hyperboloid in 3D), whose foci are in  $\mathbf{m}_i^{(m)}$  and  $\mathbf{m}_j^{(m)}$ , and whose vertices are  $c\hat{\tau}_{ij}^{(m)}$  far apart. If source and microphones are coplanar, the location of the source can be ideally obtained by intersecting two or more hyperbolas [14], as in Figure 5 and some primitive solutions for source localization rely on this idea. It is important to notice that when the source is sufficiently far from the node, the branch of hyperbola can be confused with its asymptote: in this case the TDOA is informative only with respect to the direction towards which the source is located and not its distance from the array. In this context it is more convenient to work with DOAs (see Section 6).

In general, intersecting hyperbolas is a strongly nonlinear problem, with obvious negative consequences on the computational cost and the sensitivity to noise in the measurements. In [12] a solution is proposed to overcome this issue, which relies on a projective representation. The hyperbola derived from the TDOA  $\hat{\tau}_{ij}^{(m)}$  at microphones  $\mathbf{m}_i^{(m)}$  and  $\mathbf{m}_j^{(m)}$  is written as

$$a_{ij}^{(m)} x^2 + b_{ij}^{(m)} xy + c_{ij}^{(m)} y^2 + d_{ij}^{(m)} x + e_{ij}^{(m)} y + f_{ij}^{(m)} = 0, \quad (27)$$

where the coefficients  $a_{ij}^{(m)}$ ,  $b_{ij}^{(m)}$ ,  $c_{ij}^{(m)}$ ,  $d_{ij}^{(m)}$ ,  $e_{ij}^{(m)}$ ,  $f_{ij}^{(m)}$  are determined in closed form by  $\mathbf{m}_i^{(m)}$ ,  $\mathbf{m}_j^{(m)}$ , and  $\hat{\tau}_{ij}^{(m)}$ . Equation (27) represents a constraint on the source location. In the

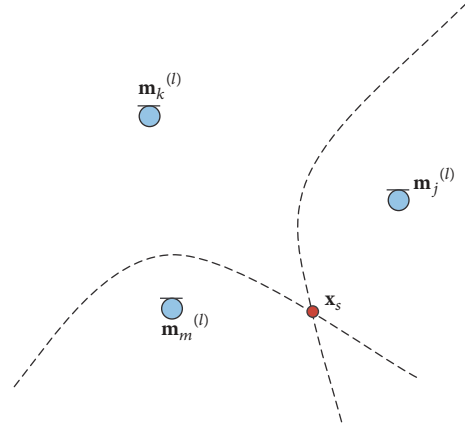


FIGURE 5: The source lies at the intersection of hyperbolas obtained from TDOAs.

presence of noise, the constraint is not honored and a residual can be defined as

$$\begin{aligned} \varepsilon_{ij}^{(m)}(\mathbf{x}) = & v_{ij}^{(m)} \left( a_{ij}^{(m)} x^2 + b_{ij}^{(m)} xy + c_{ij}^{(m)} y^2 + d_{ij}^{(m)} x \right. \\ & \left. + e_{ij}^{(m)} y + f_{ij}^{(m)} \right), \end{aligned} \quad (28)$$

where  $v_{ij}^{(m)} = 1$  for all the TDOAs that have been found reliable and 0 otherwise. The residuals are stacked in the vector  $\varepsilon(\mathbf{x})$ , and the source is localized by minimizing the cost function

$$J_{\text{HYP}}^{(\text{TDOA})}(\mathbf{x}) = \varepsilon(\mathbf{x})^T \varepsilon(\mathbf{x}). \quad (29)$$

If TDOA measurements are affected by additive white Gaussian noise, it is easy to demonstrate that (29) is proportional to the ML cost function.

It has been demonstrated that a simplification of the localization cost function can be brought if a reference microphone is adopted, at the cost of the nodes sharing a common clock. Without loss of generality, we assume the reference to be the first microphone in the first sensor (i.e.,  $i = 1$  and  $m = 1$ ), and we also set  $\mathbf{m}_1^{(1)} = \mathbf{0}$ ; that is, the origin of the reference frame coincides with the reference microphone. Moreover, we can drop the array index  $m$  upon assigning a global index to the microphones in different sensors, ranging from  $j = 1$  to  $j = NM$ . In this context, it is possible to linearize the localization cost function, as shown in the next paragraphs. By rearranging the terms in (19) and setting  $i = 1$ , it is easy to derive

$$\|\mathbf{x}_s\| - \hat{r}_j = \|\mathbf{x}_s - \mathbf{m}_j\|, \quad (30)$$

where  $\hat{r}_j = c\hat{\tau}_j$ . In [45] it has been proposed to represent the source localization problem in the space-range reference frame: a point  $\mathbf{x} = [x, y]^T$  in space is mapped onto the 3D space-range as  $[\mathbf{x}^T, r]^T$  where

$$r = \|\mathbf{x}_s - \mathbf{x}\| - \|\mathbf{x}_s\|. \quad (31)$$

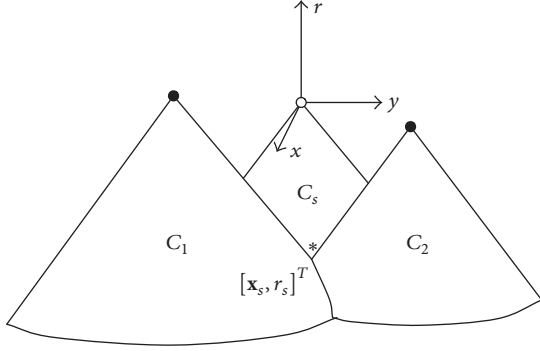


FIGURE 6: In the space-range reference frame the source is a point on the surface of the cone  $C_s$  and closest to the cones  $C_j$ .

We easily recognize that, in absence of noise and measurement errors,  $r$  is the range difference relative to the source in  $\mathbf{x}_s$  between the reference microphone and a microphone in  $\mathbf{x}$ . If we replace

$$r_s = -\|\mathbf{x}_s\|, \quad (32)$$

we can easily interpret (30) as the equation of a negative half-cone  $C_j$  in the space-range reference frame, whose apex is  $[\mathbf{m}_j^T, \hat{r}_j]^T$ , and with aperture  $\pi/4$ , and the source point  $[\mathbf{x}_s^T, r_s^T]$  lies on it. Equation (30) can be iterated for  $j = 2, \dots, N \times M$ , and the source is bound to lie on the surface of all the cones. Moreover, the range  $r_s$  of the source from the reference microphone must honor the constraint

$$r_s = -\|\mathbf{x}_s\|, \quad (33)$$

which is the equation of a cone  $C_s$  whose apex is in  $[\mathbf{m}_1^T, 0]^T$  and with aperture  $\pi/4$ . The ML source estimate, therefore, is the point on the surface of the cone in (33) closest to all the cones defined by (30), as represented in Figure 6.

By squaring both members of (30) and recognizing that

$$x_s^2 + y_s^2 - r_s^2 = 0, \quad (34)$$

(30) can be rewritten as

$$x_j x_s + y_j y_s - \hat{r}_j r_s - \frac{1}{2}(x_j^2 + y_j^2 - \hat{r}_j^2) = 0, \quad (35)$$

which is the equation of a plane in the space-range reference frame, on which the source is bound to lie. Under error in the range difference estimate  $\hat{r}_j$ , (35) is not satisfied exactly, and therefore a residual can be defined as

$$e_j = x_j x_s + y_j y_s - \hat{r}_j r_s - \frac{1}{2}(x_j^2 + y_j^2 - \hat{r}_j^2). \quad (36)$$

Based upon the definition of  $e_j$ , the LS spherical cost function is given by [46]

$$J_{\text{SP}}^{(\text{TDOA})}(\mathbf{x}_s, r_s) = \sum_{j=2}^{NM} e_j^2. \quad (37)$$

Most LS estimation techniques adopt this cost function and they differ only for the additional constraints. The Unconstrained Least Squares (ULS) estimator [16, 30, 31, 47, 48] localizes the source as

$$\hat{\mathbf{x}}_s^{(\text{ULS})} = \arg \min_{\mathbf{x}_s, r_s} J_{\text{SP}}^{(\text{TDOA})}(\mathbf{x}_s, r_s). \quad (38)$$

It is important to notice that the absence of any explicit constraint that relates  $\mathbf{x}_s$  and  $r_s$  provides in many applications a poor localization accuracy. Constrained Least Squares techniques, therefore, reintroduce this constraint as

$$\begin{aligned} \hat{\mathbf{x}}_s^{(\text{CLS})} = \arg \min_{\mathbf{x}_s, r_s} J_{\text{SP}}^{(\text{TDOA})}(\mathbf{x}_s, r_s) \\ \text{subject to } r_s = -\|\mathbf{x}_s\|. \end{aligned} \quad (39)$$

Based on (39), Spherical Intersection [29], Least Squares with Linear Correction [49], and Squared Range Difference Least Squares [15] estimators have been proposed, which differ for the minimization procedure, ranging from iterative to closed-form solutions. It is important to notice, however, that all these techniques assume the presence of a global reference microphone and synchronization valid for all the nodes. Alternative solutions that overcome this technological constraint have been proposed in [17, 50]. Here the concept of cone propagation in the space-range reference has been put at advantage. In particular, in [50], the propagation cone is defined slightly different from the one defined in (30): the apex is in the source and  $r_s = 0$ . As a consequence, in absence of synchronization errors, all the points  $[\mathbf{x}_j^{(m)T}, \hat{r}_j^{(m)T}]^T$ ,  $j = 1, \dots, N$ ,  $m = 1, \dots, M$  must lie on the surface of the propagation cone. If a sensor exhibits a clock offset, its measurements will be shifted along the range axis. The shift can be expressed as a function of the source location, and therefore it can be included in the localization cost function at a cost of some nonlinearity. The extension to the 3D localization cost function was then proposed in [17].

## 6. DOA-Based Localization

When each node in the WASN incorporates multiple microphones, the location of an acoustic source can be estimated based on direction of arrival (DOA), also known as *bearing*, measurements. Although, such approaches require increased computational complexity in the nodes—in order to perform the DOA estimation—they can attain very low-bandwidth usage as only DOA measurements need to be transmitted in the network. Moreover, they can tolerate unsynchronized input given that the sources are static or that they move at a rather slow rate relative to the analysis frame. DOA measurements describe the direction from which sound is propagating with respect to a sensor in each time instant and are an attractive approach to location estimation also due to the ease in which such estimates can be obtained: a variety of broadband DOA estimation methods for acoustic sources are available in the literature, such as the broadband MUSIC algorithm, [51] the ESPRIT algorithm [52], Independent Component Analysis (ICA) methods [53], or Sparse Component Analysis (SCA) methods [54]. When the microphones

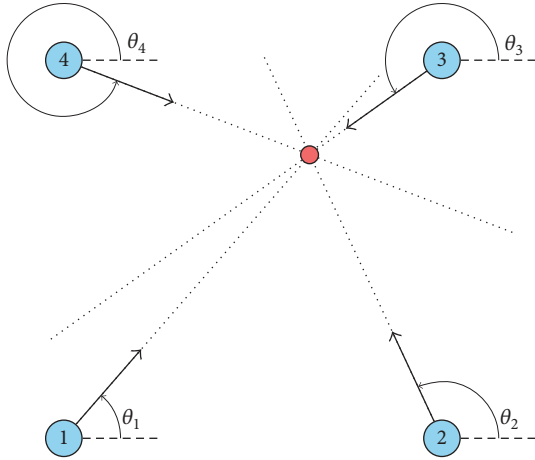


FIGURE 7: Triangulation using DOA estimates ( $\theta_1$ – $\theta_4$ ) in a WASN of 4 nodes (blue circles, numbered 1 to 4) and one active sound source (red circle).

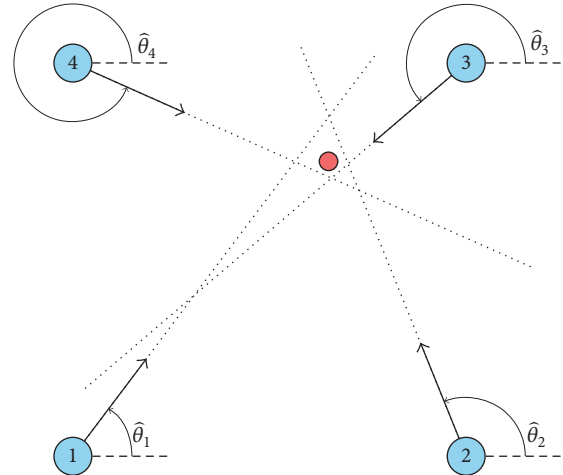


FIGURE 8: Triangulation using DOA estimates contaminated by noise ( $\hat{\theta}_1$ – $\hat{\theta}_4$ ) in a WASN of 4 nodes (blue circles, numbered 1 to 4) and the estimated location of the sound source (red circle).

at the nodes follow a specific geometry, for example, circular, methods such as Circular Harmonics Beamforming (CHB) [55] can also be applied.

In the sequel, we will first review DOA-based localization approaches when a single source is active in the acoustic environment. Then, we will present approaches for localization of multiple simultaneously active sound sources. Finally, we will discuss methods to jointly estimate the locations as well as the number of sources, a problem which is known as source counting.

**6.1. Single Source Localization through DOAs.** In the single source case, the location can be estimated as the intersection of bearing lines (i.e., lines emanating from the locations of the sensors at the directions of the sensors' estimated DOAs), a method which is known as *triangulation*. An example of triangulation is illustrated in Figure 7. The problem closely relates to that of target motion analysis, where the goal is to estimate the position and velocity of a target from DOA measurements acquired by a single moving or multiple observers. Hence, many of the methods were proposed for the target motion analysis problem but outlined here in the context of sound source localization in WASNs.

Considering a WASN of  $M$  nodes at locations  $\mathbf{q}_m = [q_{x,m} \ q_{y,m}]^T$ , the function that relates a location  $\mathbf{x} = [x \ y]^T$  with its true azimuthal DOA estimate at node  $m$  is

$$\theta_m(\mathbf{x}) = \arctan\left(\frac{y - q_{y,m}}{x - q_{x,m}}\right), \quad (40)$$

where  $\arctan(\cdot)$  is the four-quadrant inverse tangent function. Note that we deal with the two-dimensional location estimation problem; that is, only the azimuthal angle is needed. When information about the elevation angle is also available, location estimation can be extended to the three-dimensional space.

In any practical case, however, the DOA estimates  $\hat{\theta}_m$ ,  $m = 1, \dots, M$ , will be contaminated by noise and triangulation will not be able to produce a unique solution, craving for the need of statistical estimators to optimally tackle the triangulation problem. This scenario is illustrated in Figure 8. When the DOA noise is assumed to be Gaussian, the ML location estimator can be derived by minimizing the nonlinear cost function [56, 57]:

$$J_{\text{ML}}^{(\text{DOA})}(\mathbf{x}) = \sum_{m=1}^M \frac{1}{\sigma_m^2} (\hat{\theta}_m - \theta_m(\mathbf{x}))^2, \quad (41)$$

where  $\sigma_m^2$  is the variance of DOA noise at the  $m$ th sensor.

As information about the DOA error variance at the sensors is rarely available in practice, (41) is usually modified to

$$J_{\text{NLS}}^{(\text{DOA})}(\mathbf{x}) = \sum_{m=1}^M (\hat{\theta}_m - \theta_m(\mathbf{x}))^2, \quad (42)$$

which is termed as *nonlinear least squares (NLS)* [58] cost function. Minimizing (42) results in the ML estimator, when the DOA noise variance is assumed to be the same at all sensors.

While asymptotically unbiased, the nonlinear nature of the above cost functions requires numerical search methods for minimization, which comes with increased computational complexity compared to closed-form solutions and can become vulnerable to convergence problems under bad initialization, poor geometry between sources and sensors, high noise, or insufficient number of measurements. To surpass some of these problems, some methods form geometrical constraints between the measured data and result in better convergence properties than the maximum likelihood estimator [59] or try to directly minimize the mean squared

location error [60] instead of minimizing the total bearing error in (41) and (42).

Other approaches are targeted at linearizing the above nonlinear cost functions. Stansfield [61] developed a weighted linear least squares estimator based on the cost function of (41) under the assumption that range information  $r_m$  is available and DOA errors are small. Under the small DOA errors assumption  $\hat{\theta}_m - \theta_m(\mathbf{x})$  can be approximated by  $\sin(\hat{\theta}_m - \theta_m(\mathbf{x}))$  and the ML cost function can be modified to

$$J_{\text{ST}}^{(\text{DOA})}(\mathbf{x}) = \frac{1}{2} (\mathbf{A}\mathbf{x} - \mathbf{b})^T \mathbf{W}^{-1} (\mathbf{A}\mathbf{x} - \mathbf{b}), \quad (43)$$

where

$$\mathbf{A} = \begin{bmatrix} \sin \hat{\theta}_1 & -\cos \hat{\theta}_1 \\ \vdots & \vdots \\ \sin \hat{\theta}_M & -\cos \hat{\theta}_M \end{bmatrix}, \quad (44)$$

$$\mathbf{b} = \begin{bmatrix} q_{x,1} \sin \hat{\theta}_1 - q_{y,1} \cos \hat{\theta}_1 \\ \vdots \\ q_{x,M} \sin \hat{\theta}_M - q_{y,M} \cos \hat{\theta}_M \end{bmatrix},$$

$$\mathbf{W} = \begin{bmatrix} r_1^2 \sigma_1^2 & & \mathbf{0} \\ & r_2^2 \sigma_2^2 & \\ & & \ddots \\ \mathbf{0} & & & r_M^2 \sigma_M^2 \end{bmatrix},$$

which is linear and has a closed-form solution:

$$\hat{\mathbf{x}}_s^{(\text{ST})} = (\mathbf{A}^T \mathbf{W}^{-1} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}^{-1} \mathbf{b}. \quad (45)$$

When range information is not available, the weight matrix  $\mathbf{W}$  can be replaced by the identity matrix. In this way, the Stansfield estimator is transformed to the *orthogonal vectors estimator*, also known as the *pseudolinear estimator* [62]:

$$\hat{\mathbf{x}}_s^{(\text{OV})} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{b}. \quad (46)$$

While simple in their implementation and computationally very efficient due to their closed-form solution, these linear estimators suffer from increased estimation bias [63]. A comparison between the Stansfield estimator and the ML estimator in [64] reveals that the Stansfield estimator provides biased estimates. Moreover, the bias does not vanish even for a large number of measurements. To reduce that bias various methods have been proposed based on instrumental variables [65–67] or total least squares [57, 68].

Motivated by the need for computational efficiency, the intersection point method [69] is based on finding the location of a source by taking the centroid of the intersections of *pairs* of bearing lines. The centroid is simply the mean of the set of intersection points and minimizes the sum of squared distances between itself and each point in the set.

To increase robustness in poor geometrical conditions, the method incorporates a scheme of identifying and excluding outliers that occur from the intersection of pairs of bearing lines that are almost parallel. Nonetheless, the performance of the method is very similar to that of the pseudolinear estimator.

To attain the accuracy of nonlinear least squares estimators and improve their computational complexity, the grid-based (GB) method [70, 71] is based on making the search space discrete by constructing a grid  $\mathcal{G}$  of  $N_g$  grid points over the localization area. Moreover, as the measurements are angles, the GB method proposes the use of the Angular Distance—taking values in the range of  $[0, \pi]$ —as a more proper measure of similarity than the absolute distance of (42). The GB method estimates the source location by finding the grid point whose DOAs most closely match the estimated DOAs from the sensors by solving

$$J_{\text{GB}}^{(\text{DOA})}(\mathbf{x}) = \sum_{m=1}^M [A(\hat{\theta}_m, \theta_m(\mathbf{x}))]^2, \quad (47)$$

$$\hat{\mathbf{x}}_s^{(\text{GB})} = \arg \min_{\mathbf{x} \in \mathcal{G}} J_{\text{GB}}^{(\text{DOA})}(\mathbf{x}), \quad (48)$$

where  $A(\cdot, \cdot)$  denotes the angular distance between the two arguments.

To eliminate the location error introduced by the discrete nature of this approach, a very dense grid (high  $N_g$ ) is required. The search for the best grid point is performed in an iterative manner: it starts with a coarse grid (low value of  $N_g$ ) and once the best grid point is found—according to (47)—a new grid centered on this point is generated, with a smaller spacing between grid points but also a smaller scope. Then, the best grid point in this new grid is found and the procedure is repeated until the desired accuracy is obtained, while keeping the complexity under control, as it does not require an exhaustive search over a large number of grid points. In [70] it is shown that the GB method is much more computationally efficient than the nonlinear least squares estimators and attains the same accuracy.

**6.2. Multiple Source Localization.** When considering multiple sources, a fundamental problem is that the correct association of DOAs from the nodes to the sources is unknown. Hence, in order to perform triangulation, one must first estimate the correct DOA combinations from the nodes that correspond to the same source. The use of DOAs that belong to different sources will result in “ghost” sources, that is, locations that do not correspond to real sources, thus severely affecting localization performance. This is known as the *data-association problem*. The data-association problem is illustrated with an example in Figure 9: in a WASN with two nodes (blue circles) and two active sound sources (red circles), let the solid lines show the DOAs to the first source and the dashed lines show the DOAs to the second source. Intersecting the bearing lines will result in 4 intersection points: the red circles, that correspond to the true sources’ locations and are estimated by using the correct DOA combinations (i.e., the DOAs from the node that correspond

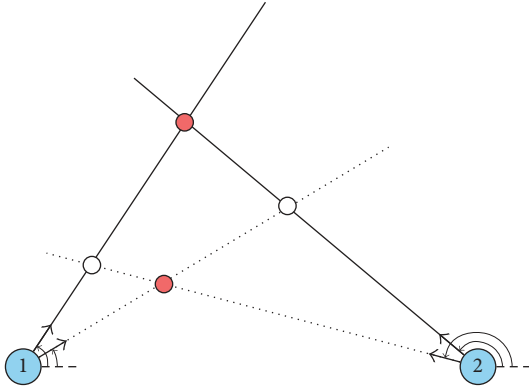


FIGURE 9: Illustration of the data-association problem. A two-node WSN with two active sound sources. The solid lines show the DOAs to the first source, while the dashed lines show the DOAs to the second source. Intersecting the bearing lines from the sensors will result in 4 possible intersection points: the true sources' locations (red circles) that are estimated by using the correct combination of DOAs and the "ghost" sources (white circles) that are the results of using the erroneous DOA combinations.

to the same source) and the white circles that are the result of using the erroneous DOA combinations ("ghost" sources).

Also, with multiple active sources, some arrays might underestimate their number, especially when some nodes are far from some sources or when the sources are close together in terms of their angular separation [54]. Thus, *missed detections* can occur, meaning that the DOAs of some sources from some arrays may be missing. As illustrated in [72], missed detections can occur very often in practice. In the sequel, we review approaches for the data-association and localization problem of multiple sources whose number is assumed to be known.

Some approaches tried to tackle the data-association problem by enumerating all possible DOA combinations from the sensors and deciding on the correct DOA combinations based on the resulting location estimates from all combinations. In general, if  $C_s$  denotes the number of sensors that detected  $s$  sources, the number of possible DOA combinations is

$$N_{\text{comb}} = \prod_{s=1}^S s^{C_s}. \quad (49)$$

The position nonlinear least squares (P-NLS) estimator developed in [73] incorporates the association procedure in the ML cost function, which takes the form

$$J_{\text{P-NLS}}^{(\text{DOA})}(\mathbf{x}) = \sum_{m=1}^M \min_i |\hat{\theta}_{m,i} - \theta_m(\mathbf{x})|^2, \quad (50)$$

where  $\hat{\theta}_{m,i}$  is the  $i$ th DOA estimate of sensor  $m$ . To minimize (50),  $N_{\text{comb}}$  initial locations are estimated (one for each DOA combination) using a linear least squares estimator, such as the pseudolinear transform of (46). Then, the cost function (50) is minimized—using numerical search methods— $N_{\text{comb}}$  times, each time using a different initial location estimate.

Each time, for each sensor the DOA closest to the DOA of the initial location estimate is used to take part in the minimization procedure. In that way, for all initial locations, the estimator is expected to converge to a location of a true source. However, as illustrated in [70], in the presence of missed detections and high noise the approach is not able to completely eliminate "ghost" sources.

The multiple source grid-based method [70] estimates an initial location for each possible DOA combination from the sensors by solving (47). It then decides which of the initial location estimates correspond to a true source, heuristically by selecting the estimated initial locations whose DOAs are closer to the DOAs from the combination used to estimate that location.

Other approaches focus on solving the data-association problem prior to the localization procedure. In this way, the correct association of DOAs from the sensors to the sources is estimated beforehand and the multiple source localization problem decomposes into multiple single source localization problems. In [74] the data-association problem is viewed as an assignment problem and is formulated as a statistical estimation problem which involves the maximization of the ratio of the likelihood that the measurements come from the same target to the likelihood that the measurements are false-alarms. Since the proposed solution becomes NP-hard for more than three sensors, suboptimal solutions tried to solve the same problem in pseudopolynomial time [75, 76].

An approach based on clustering of intersections of bearing lines in scenarios with no missed detections is discussed in [77]. It is based on the observation that intersections between pairs of bearing lines that correspond to the same source will be close to each other. Hence, intersections between bearing lines will cluster around the true sources, revealing the correct DOA associations, while intersections from erroneous DOA combinations will be randomly distributed in space.

Permitting the transmission of low-bandwidth additional information from the sensors can lead to more efficient approaches to the data-association problem. The idea is that the sensors can extract and transmit features associated with each source they detect. Appropriate features for the data-association problem must possess the property of being "similar" for the same source in the different sensors. Then, the correct association of DOAs to the sources can be found by comparing the corresponding features.

In [78] such features are extracted using Blind Source Separation. The features are binary masks [79] in the frequency domain for each detected source that when applied to the corresponding source signals they perform source separation. The extraction of such features relies on the W-disjoint orthogonality assumption [80], which states that in a given time-frequency point only one source is active, an assumption which has been showed to be valid especially for speech signals [81]. The association algorithm works by finding the binary masks from the different arrays that correlate the most. However, the method is designed for scenarios with no missed detections and, as illustrated in [72], performance significantly drops when missed detections

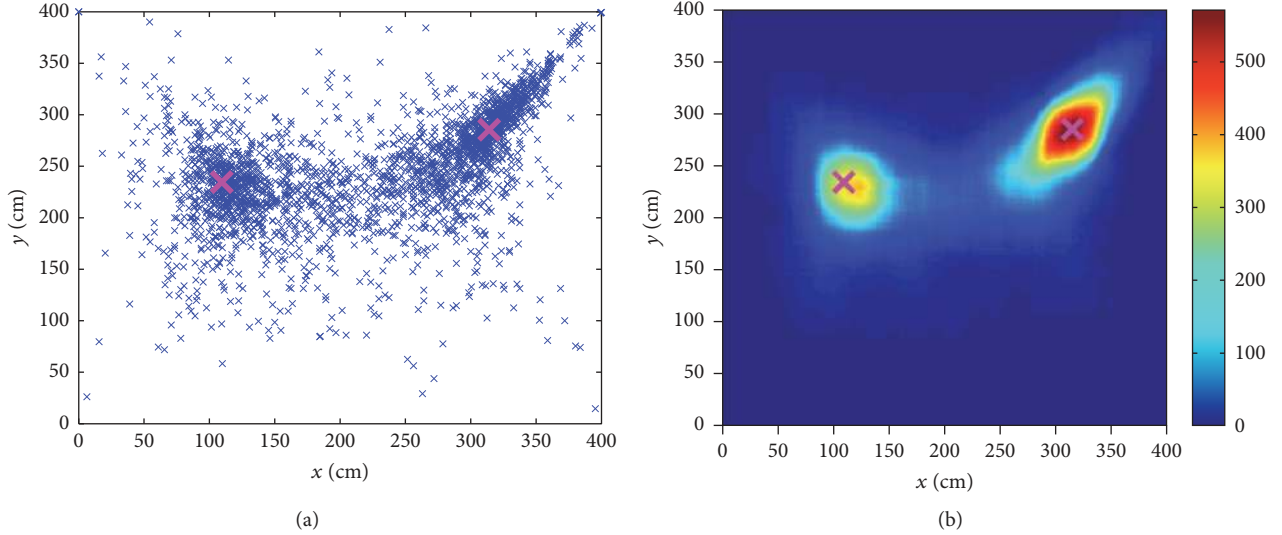


FIGURE 10: Example of (a) narrowband location estimates and (b) their corresponding histogram for a scenario of two active sound sources (the red X's). The narrowband location estimates form two clusters around the true sources' locations, while their corresponding histogram describes the plausibility that a source is present at a given location.

occur. Moreover, the association algorithm is designed for the case of two sensors.

The design of association features that are robust to missed detections is considered in [72] along with a greedy association algorithm that can work with an arbitrary number of sources and sensors. The association features describe how the frequencies of the captured signals are distributed to the sources. To do that, the method estimates a DOA  $\phi(\omega, k)$  in each time-frequency point, where  $\omega$  and  $k$  denote the frequency and time frame index, respectively. Then, a time-frequency point  $(\omega, k)$  is assigned to source  $s$  if the following conditions are met:

$$A(\phi(\omega, k), \hat{\theta}_s(k)) < A(\phi(\omega, k), \hat{\theta}_q(k)), \quad \forall q \neq s, \quad (51)$$

$$A(\phi(\omega, k), \hat{\theta}_s(k)) < \epsilon, \quad (52)$$

where  $\hat{\theta}_s(k)$  is the DOA estimate at time frame  $k$  for the  $s$ th source at the sensor of interest and  $\epsilon$  is a predefined threshold. Equations (51) and (52) imply that each frequency is assigned to the source whose DOA is closest to the estimated DOA in this frequency, as long as their distance does not exceed a certain threshold  $\epsilon$ . The second condition (see (52)) adds robustness to missed detections as it rejects the frequencies with DOA estimates whose distance from the detected sources' DOAs is significantly large.

**6.3. Source Counting.** Assuming that the number of sources is also unknown and can vary arbitrarily in time, other approaches were developed to jointly solve the source counting and location estimation problem. In these approaches, the central idea is to utilize narrowband DOA estimates—for each time-frequency point—from the nodes in order to

estimate narrowband location estimates. Appropriate processing of the narrowband location estimates can infer the number and locations of the sound sources. The location for each time-frequency point is estimated using triangulation based on the corresponding narrowband DOA estimates from the sensors at that time-frequency point. Figure 10 shows an example of such narrowband location estimates and their corresponding histogram, which also describes the plausibility that a source is at a given location. The processing of these narrowband location estimates is usually done by statistical modeling methods: in [82], the narrowband location estimates are modeled by a Gaussian Mixture Model (GMM), where the number of Gaussian components corresponds to the number of sources, while the means of the Gaussians determine the sources' locations. A variant of the Expectation-Maximization (EM) algorithm is proposed that incorporates empirical criteria for removing and merging Gaussian components in order to determine the number of sources as well. A Bayesian view of the Gaussian Mixture Modeling is adopted in [83, 84], where a variant of the  $K$ -means algorithm is utilized that is able to determine both the number of clusters (i.e., number of sources) and the cluster centroids (i.e., sources' locations) using split and merge operations on the Gaussian components.

## 7. SRP-Based Localization

Approaches based on the steered response power (SRP) have attracted the attention of many researchers due to their robustness in noisy and reverberant environments. Particularly, the SRP-PHAT algorithm is today one of the most popular approaches for acoustic source localization using microphone arrays [85–87]. Basically, the goal of SRP methods is to maximize the power of the received sound

source signal using a steered filter-and-sum beamformer. To this end, the method uses a grid-search procedure where candidate source locations are explored by computing a functional that relates spatial location to the TDOA information extracted from multiple microphone pairs. The power map resulting from the values computed at all candidate source locations (also known as Global Coherence Field [88]) will show a peak at the estimated source location.

Since SRP approaches are based on the exploitation of TDOA information, synchronization issues also arise when applying SRP in WASNs. As in the case of source localization using DOAs or TDOAs, SRP-based approaches for WASNs have been proposed considering that multiple microphones are available at each node [89–91]. In these cases, the SRP method can be used for acquiring DOA estimates at each node or collecting source location estimates that are merged by a central node. Next, we describe the fundamentals of SRP-PHAT localization.

*7.1. Conventional SRP-PHAT (C-SRP).* Consider a set of  $N$  different microphones capturing the signal arriving from a sound source located at a spatial position  $\mathbf{x}_s \in \mathbb{R}^3$  in an anechoic scenario, following the model of (20). The SRP is defined as the output power of a filter-and-sum beamformer steered to a given spatial location. DiBiase [85] demonstrated that the SRP at a spatial location  $\mathbf{x} \in \mathbb{R}^3$  calculated over a time interval of  $T$  samples can be efficiently computed in terms of GCCs:

$$J_C^{(\text{SRP})}(\mathbf{x}) = \frac{2}{T} \sum_{i=1}^N \sum_{j=i+1}^N R_{ij}(\tau_{ij}(\mathbf{x})) + \sum_{i=1}^N R_{ii}(0), \quad (53)$$

where  $\tau_{ij}(\mathbf{x})$  is the time difference of arrival (TDOA) that would produce a sound source located at  $\mathbf{x}$ ; that is,

$$\tau_{ij}(\mathbf{x}) = \frac{\|\mathbf{x} - \mathbf{m}_i\| - \|\mathbf{x} - \mathbf{m}_j\|}{c}. \quad (54)$$

The last summation term in (53) is usually ignored, since it is a power offset independent of the steering location. When GCCs are computed with PHAT, the resulting SRP is known as SRP-PHAT.

In practice, the method is implemented by discretizing the location space region  $\mathbb{V}$  using a search grid  $\mathcal{G}$  consisting of candidate source locations in  $\mathbb{V}$  and computing the functional of (53) at each grid position. The estimated source location is the one providing the maximum functional value:

$$\hat{\mathbf{x}}_s^{(\text{C-SRP})} = \arg \max_{\mathbf{x} \in \mathcal{G}} J_C^{(\text{SRP})}(\mathbf{x}). \quad (55)$$

*7.2. Modified SRP-PHAT (M-SRP).* Reducing the computational cost of SRP is an important issue in WASNs, since power-related constraints in the nodes may render impractical its implementation in real-world applications. The vast amount of modified solutions based on SRP is aimed at reducing the computational cost of the grid-search step [92, 93]. A problem of these methods is that they are prone to discard part of the information available, leading

to some performance degradation. Other recent approaches are based on analyzing the volume surrounding the grid of candidate source locations [87, 94]. By taking this into account, the methods are able to accommodate the expected range of TDOAs at each volume in order to increase the robustness of the algorithm and relax its computational complexity. The modified SRP-PHAT collects and uses the TDOA information related to the volume surrounding each point of the search grid by considering a modified functional [87]:

$$J_M^{(\text{SRP})}(\mathbf{x}) = \sum_{i=1}^N \sum_{j=i+1}^N \sum_{\tau=L_{ij}^l(\mathbf{x})}^{L_{ij}^u(\mathbf{x})} R_{ij}(\tau), \quad (56)$$

where  $L_{ij}^l(\mathbf{x})$  and  $L_{ij}^u(\mathbf{x})$  are the lower and upper accumulation limits of GCC delays, which depend on the spatial location  $\mathbf{x}$ .

The accumulation limits can be calculated beforehand in an exact way by exploring the boundaries separating the regions corresponding to the points of the grid. Alternatively, they can be selected by considering the spatial gradient of the TDOA  $\nabla_{\tau_{ij}}(\mathbf{x}) = [\nabla_{x\tau_{ij}}(\mathbf{x}), \nabla_{y\tau_{ij}}(\mathbf{x}), \nabla_{z\tau_{ij}}(\mathbf{x})]^T$ , where each component  $\gamma \in \{x, y, z\}$  of the gradient is

$$\nabla_{\gamma\tau_{ij}}(\mathbf{x}) = \frac{1}{c} \left( \frac{\gamma - \gamma_i}{\|\mathbf{x} - \mathbf{m}_i\|} - \frac{\gamma - \gamma_j}{\|\mathbf{x} - \mathbf{m}_j\|} \right). \quad (57)$$

For a rectangular grid where neighboring points are separated a distance  $r_g$  and the lower and upper accumulation limits are given by

$$\begin{aligned} L_{ij}^l(\mathbf{x}) &= \tau_{ij}(\mathbf{x}) - \left\| \nabla_{\tau_{ij}}(\mathbf{x}) \right\| \cdot d_g, \\ L_{ij}^u(\mathbf{x}) &= \tau_{ij}(\mathbf{x}) + \left\| \nabla_{\tau_{ij}}(\mathbf{x}) \right\| \cdot d_g, \end{aligned} \quad (58)$$

where  $d_g = (r_g/2) \min(|1/\sin(\theta) \cos(\phi)|, |1/\sin(\theta) \sin(\phi)|, |1/\cos(\theta)|)$ , and the gradient direction angles are given by

$$\theta = \cos^{-1} \left( \frac{\nabla_{z\tau_{ij}}(\mathbf{x})}{\left\| \nabla_{\tau_{ij}}(\mathbf{x}) \right\|} \right), \quad (59)$$

$$\phi = \arctan_2 \left( \nabla_{y\tau_{ij}}(\mathbf{x}), \nabla_{x\tau_{ij}}(\mathbf{x}) \right).$$

The estimated source location is again obtained as the point in the search grid providing the maximum functional value:

$$\hat{\mathbf{x}}_s^{(\text{M-SRP})} = \arg \max_{\mathbf{x} \in \mathcal{G}} J_M^{(\text{SRP})}(\mathbf{x}). \quad (60)$$

Figure 11 shows the normalized SRP power maps obtained by C-SRP using two different grid resolutions and the one obtained by M-SRP using a coarse spatial grid. In (a), the fine search grid shows clearly the hyperbolas intersecting at the true source location. However, when the number of grid points is reduced in (b), the SRP power map does not provide a consistent maximum. As shown in (c), M-SRP is able to fix

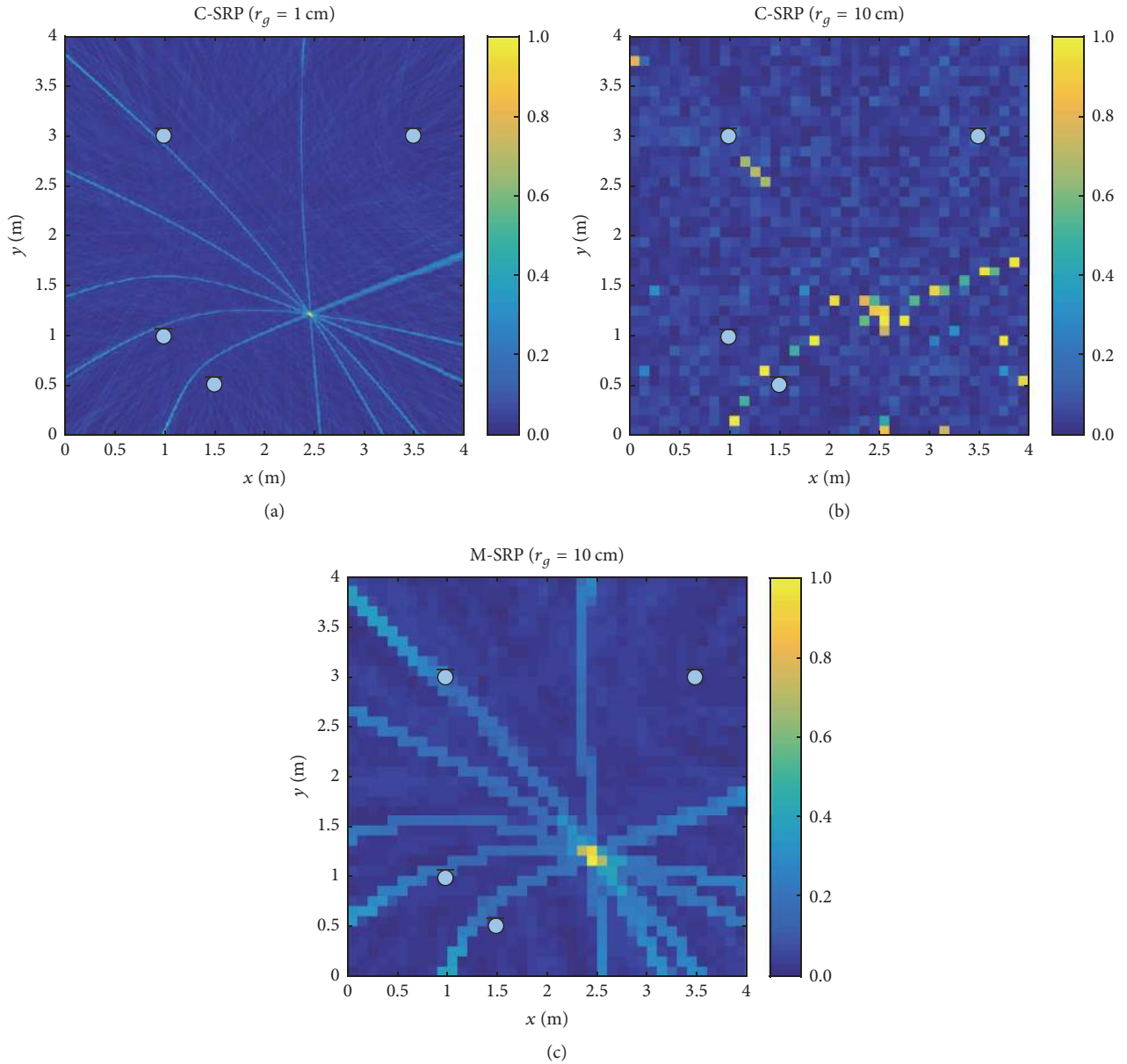


FIGURE 11: Example of SRP power maps obtained by C-SRP and M-SRP for a speech frame in anechoic acoustic conditions using different spatial grid resolutions. Yellow circles indicate the node locations.

this situation, showing a consistent maximum even when a coarse spatial grid is used.

An iterative approach of the M-SRP method was described in [95], where the M-SRP is initially evaluated using a coarse spatial grid. Then, the volume surrounding the point of highest value is iteratively decomposed by using a finer spatial grid. This approach allows obtaining almost the same accuracy as the fine-grid search with a substantial reduction of functional evaluations.

Finally, recent works are also focusing on hardware aspects in the nodes with the aim of efficiently computing the SRP. In this context, the use of graphics processing units (GPUs) for implementing SRP-based approaches is

specially promising [96, 97]. In [98] the performance of SRP-PHAT is analyzed over a massive multichannel processing framework in a multi-GPU system, analyzing its performance as a function of the number of microphones and available computational resources in the system. Note, however, that the performance of SRP approaches is also related to the properties of the sound sources, such as their bandwidth or their low-pass/pass-band nature [99, 100].

## 8. Self-Localization of Acoustic Sensor Nodes

Methods for sound source localization discussed in previous sections assume that  $\mathbf{q}_m$  for  $m = 1, \dots, M$ , the locations of



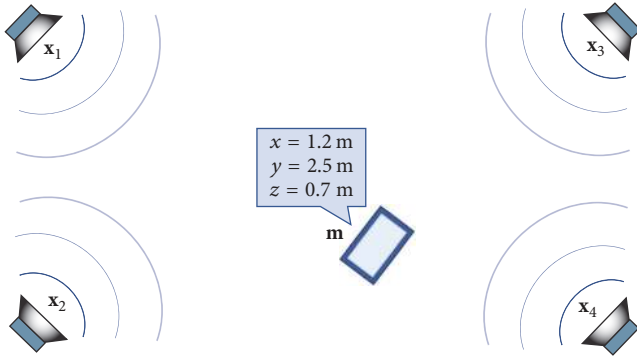


FIGURE 12: Illustration of a WASN comprised with  $M = 1$  node and  $S = 4$  loudspeakers for self-localization scenario.

the acoustic sensor nodes, or  $\mathbf{m}_i$  for  $i = 1, \dots, N$ , those of microphones, are known to the system and fixed in time. In practical situations, however, the precise locations of the sensor nodes or the microphones may not be known (e.g., for the deployment of ad hoc WASNs). Furthermore, in some WASN applications, the node locations may change over time. Due to these reasons, self-calibration for adjusting the known node/microphone locations or self-localization of unknown nodes of the WASNs becomes necessary.

The methods for self-localization of WASNs can be divided into three categories [101]. The first one uses non-acoustic sensors such as accelerometer and magnetometers, the second one uses the signal strength, and the third one uses the TOA or TDOA of acoustic signals. In this section, we focus on the last category since it has been shown to enable localization with fine granularity, centimeter-level localization [21], and requires only the acoustic sensors.

The TOA/TDOA-based algorithms can be further divided into two types depending upon whether the source positions are known for self-localization. Most of the early works addressed this problem as microphone array calibration [102–105] with known source locations. The general problem of joint source and sensor localization was addressed by [5] as a nonlinear optimization problem. The work in [106] presented a solution to explicitly tackle the synchronization problem. In [107], a solution considering multiple sources and sensors per device was described.

This section will focus on algorithms that assume known source positions generating known probe signals without the knowledge of the sensor positions as well as the synchronization between the sensors and the sources. This approach allows all processing to take place on the sensor node for self-localization. The system illustration for this problem is given in Figure 12. The remainder of this section describes the TOA/TDOA-based methods for acoustic sensor localization by modeling the inaccurate TOA/TDOA measurements for robust localization, followed by some recent approaches.

**8.1. Problem Formulation.** Consider a WASN comprised of  $S$  sources and  $M$  nodes. The microphone locations  $\mathbf{m}_i^{(m)}$ , for  $i = 1, 2, \dots, N$ , at node  $\mathbf{q}_m$  can be determined with respect to the node location, its orientation, and its microphone

configuration. So we consider the sensor localization of finding the microphone location  $\mathbf{m}_i^{(m)}$  as the same problem as finding the node location  $\mathbf{q}_m$  in this section. Without loss of generality, we can consider the case with only one node and one microphone ( $M = N = 1$ ) because each node determines its location independently from others. In addition, we consider that the sources are the loudspeakers of the system with fixed and known locations in this scenario.

Let  $\mathbf{m}$  and  $\mathbf{x}_s$  be the single microphone position and the position of the  $s$ th source for  $s \in \mathcal{S} \triangleq \{1, 2, \dots, S\}$ , respectively. The goal is to find  $\mathbf{m}$  by means of  $S$  received acoustic signals emitted by  $S$  sources where the location of each source  $\mathbf{x}_s$  is known.

The TOF  $\eta_s$  from the  $s$ th loudspeaker to the sensor is defined as

$$\eta_s \triangleq \frac{1}{c} \|\mathbf{m} - \mathbf{x}_s\|, \quad (61)$$

where  $c$  is the speed of sound. Note that this equation is equivalent to (13) except that we consider a single microphone ( $\mathbf{m}$ ), multisource ( $\{\mathbf{x}_s\}$ ) case for  $s \in \mathcal{S}$ ; thus the TOF is indexed with respect to  $s$  instead of  $i$ . From (61), it is evident that  $\mathbf{m}$  can be found if a sufficient number of TOFs are known. In practice, we need to rely on TOAs instead of TOFs due to measurement errors.

In order to remove the effect from such unknown factors, the TDOA can be used instead of TOA, which is given by

$$\tau_{s_1 s_2} \triangleq \eta_{s_1} - \eta_{s_2} = \frac{\|\mathbf{m} - \mathbf{x}_{s_1}\| - \|\mathbf{m} - \mathbf{x}_{s_2}\|}{c}, \quad (62)$$

regarding a pair of sources  $s_1, s_2 \in \mathcal{S}$ . Please note that the subscripts for the TDOA indicate the source indexes that are different from those defined for sensor indexes in (19).

Since the probe signals generated from the sources along with their locations are assumed to be known to the sensor nodes, the derivations of the self-localization methods hereafter rely on the GCC between the probe signal and the received signal at the sensor. Provided that the direct line-of-sight between the source and the sensor is guaranteed, then the time delay found by the GCC in (22) between the probe signal and the signal received at the sensor provides the TOA information.

**8.2. Modeling of Time Measurement Errors.** Two main factors—asynchrony and the sampling frequency mismatch between sources and sensors—can be considered for the modeling of time measurement errors. When there exists asynchrony between a source and a sensor, the TOA can be modeled as

$$\tau_s = \eta_s + \Delta\eta_s, \quad (63)$$

where  $\eta_s$  is the true TOF from  $s$ th source to the sensor and  $\Delta\eta_s$  is the bias caused by the asynchrony. If there exists sampling frequency mismatch, then the sampling frequency at the sensor can be modeled as  $F_s + \Delta F_s$ , where  $F_s$  is that of the source. Considering these and ignoring the rounding of the

discrete-time index, the relationship between the discrete-time TOA  $\hat{\tau}_s$  and the actual TOF  $\eta_s$  is given by

$$\hat{\tau}_s = \frac{\eta_s + \Delta\eta_s}{F_s + \Delta F_s} = \frac{\eta_s + \Delta\eta_s}{F_s} \left( \frac{1}{1 + \Delta F_s/F_s} \right), \quad (64)$$

which can be further simplified for  $|\Delta F_s| \ll F_s$  as

$$\hat{\tau}_s \approx \alpha \hat{\eta}_s + \beta_s, \quad (65)$$

where  $\alpha \triangleq (F_s - \Delta F_s)/F_s$ ,  $\hat{\eta}_s \approx \eta_s/F_s$ , and  $\beta_s \triangleq \Delta\eta_s(F_s - \Delta F_s)/F_s^2$ . If the sources are connected to the playback system with the same clock such that the sources share the common playback delay, then  $\beta_s = \beta$  for all  $s \in \mathcal{S}$ . Therefore

$$\|\mathbf{m} - \mathbf{x}_s\| = c\eta_s \approx a\hat{\tau}_s + b, \quad (66)$$

where  $a = c/\alpha$  and  $b = -c\beta/\alpha$ .

**8.3. Least Squares Method.** Given a sufficient number of TOA or TDOA estimates, the least squares (LS) method can be used to estimate the sensor position. The TOA-based and the TDOA-based LS methods proposed in [20, 21] are described in this subsection.

**8.3.1. TOA-Based Formulation.** Motivated by the relation in (66), the localization error  $e_s$  corresponding to the  $s$ th loudspeaker can be defined as

$$e_s \triangleq (a\hat{\tau}_s + b)^2 - \|\mathbf{m} - \mathbf{x}_s\|^2, \quad \forall s \in \mathcal{S}. \quad (67)$$

If we define the error vector as  $\mathbf{e}^{(\text{TOA})} \triangleq [e_1 \ e_2 \ \dots \ e_S]^T$  with unknown parameters  $a$ ,  $b$ , and  $\mathbf{m}$ , then the cost function can be defined as

$$J_{\text{LS}}^{(\text{TOA})}(\mathbf{m}, a, b) = \|\mathbf{e}^{(\text{TOA})}\|^2. \quad (68)$$

Then the localization problem is formulated as

$$\hat{\mathbf{m}}_{\text{LS}}^{(\text{TOA})} = \arg \min_{\mathbf{m}, a, b} J_{\text{LS}}^{(\text{TOA})}(\mathbf{m}, a, b), \quad (69)$$

which does not have a closed-form solution.

**8.3.2. TDOA-Based Formulation.** We can consider the first loudspeaker for  $s = 1$  as the reference loudspeaker, whose TOA and position can be set to  $\hat{\tau}_1$  and  $\mathbf{x}_1 = \mathbf{0}$ . Given a set of TDOAs in the LS framework, it can be used with (66) as

$$a\hat{\tau}_s + b = (a\hat{\tau}_s - a\hat{\tau}_1) + (a\hat{\tau}_1 + b) \approx a\hat{\tau}_{s1} + \|\mathbf{m}\|, \quad (70)$$

where  $\hat{\tau}_{s1}$  is the TDOA between the  $s$ th and the first loudspeakers. With the first as the reference, we can define length  $S - 1$  vectors  $\tilde{e}_s \triangleq (a\hat{\tau}_{s1} + \|\mathbf{m}\|)^2 - \|\mathbf{m} - \mathbf{x}_s\|^2$ , for  $s = 2, 3, \dots, S$  and  $\mathbf{e}_{\text{LS}}^{(\text{TDOA})} \triangleq [\tilde{e}_2, \dots, \tilde{e}_S]^T$ , the cost function can be defined as

$$J_{\text{LS}}^{(\text{TDOA})}(\mathbf{m}, a) = \|\mathbf{e}_{\text{LS}}^{(\text{TDOA})}\|^2, \quad (71)$$

and the LS problem can be written as

$$\hat{\mathbf{m}}_{\text{LS}}^{(\text{TDOA})} = \arg \min_{\mathbf{m}, a} J_{\text{LS}}^{(\text{TDOA})}(\mathbf{m}, a). \quad (72)$$

Note that the TDOA-based approach is not dependent upon the parameter  $b$  unlike the TOA-based approach.

**8.3.3. LS Solutions.** For both the TOA- and TDOA-based approaches, the error vector can be formulated as  $\mathbf{e} = \mathbf{H}\Theta - \mathbf{g}$ , where the elements of the vector  $\Theta$  are unknown and those of the matrix  $\mathbf{H}$  and the vector  $\mathbf{g}$  are both known to the system. The error vectors for both approaches in this formulation can be expressed as follows [21]:

$$\begin{aligned} \mathbf{e}^{(\text{TOA})} &= \begin{bmatrix} 1 & 2\mathbf{x}_1^T & \hat{\tau}_1^2 & 2\hat{\tau}_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 2\mathbf{x}_S^T & \hat{\tau}_S^2 & 2\hat{\tau}_S \end{bmatrix} \begin{bmatrix} b^2 - \|\mathbf{m}\|^2 \\ \mathbf{m} \\ a^2 \\ ab \end{bmatrix} \\ &\quad - \begin{bmatrix} \|\mathbf{x}_1\|^2 \\ \vdots \\ \|\mathbf{x}_S\|^2 \end{bmatrix}, \\ \mathbf{e}^{(\text{TDOA})} &= \begin{bmatrix} 2\hat{\tau}_{21} & 2\mathbf{x}_2^T & \hat{\tau}_{21}^2 \\ \vdots & \vdots & \vdots \\ 2\hat{\tau}_{S1} & 2\mathbf{x}_S^T & \hat{\tau}_{S1}^2 \end{bmatrix} \begin{bmatrix} \|\mathbf{m}\| a \\ \mathbf{m} \\ a^2 \end{bmatrix} - \begin{bmatrix} \|\mathbf{x}_2\|^2 \\ \vdots \\ \|\mathbf{x}_S\|^2 \end{bmatrix}. \end{aligned} \quad (73)$$

Although the elements of the vector  $\Theta$  are not independent from one another, the constraints can be removed for computational efficiency [15, 21], then the nonlinear problems in (69) and (72) can be considered as the ULS problem as follows:

$$\min_{\Theta} \|\mathbf{e}\|^2, \quad (74)$$

which has the closed-form solution given by

$$\hat{\Theta} \triangleq (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{g}. \quad (75)$$

It has been shown that it requires  $S \geq 6$  to find the closed-form solution for both approaches [21].

For the case when  $a = c$ , that is, no sampling frequency mismatch with known speed of sound, the TDOA-based approach can be further simplified as

$$\mathbf{e} = \begin{bmatrix} 2c\hat{\tau}_{21} & 2\mathbf{x}_2^T \\ \vdots & \vdots \\ 2c\hat{\tau}_{S1} & 2\mathbf{x}_S^T \end{bmatrix} \begin{bmatrix} \|\mathbf{m}\| \\ \mathbf{m} \end{bmatrix} - \begin{bmatrix} \|\mathbf{x}_2\|^2 - c^2\hat{\tau}_{21}^2 \\ \vdots \\ \|\mathbf{x}_S\|^2 - c^2\hat{\tau}_{S1}^2 \end{bmatrix}, \quad (76)$$

which is related to the methods developed for the sensor localization problem [15, 108].

The self-localization results of the LS approaches are highly sensitive to the estimated values of TOA/TDOA. If they are estimated poorly, then the localization accuracy may significantly suffer from those inaccurate estimates. In order to address this issue, a sliding window technique is proposed to improve the accuracy of TOA/TDOA estimates [21].

*8.4. Other Approaches.* More recently, several papers have tackled the problem of how to design good probe signals between source and sensor and how to improve TOF estimation; in [22], a probe signal design using pulse compression technique and hyperbolic frequency modulated signals is presented that is capable of localizing an acoustic source and estimates its velocity and direction in case it is moving. A matching pursuit-based algorithm for TOF estimation is described in [109] and refined in [101]. The joint localization of sensor and source in an ad hoc array by using low-rank approximation methods has been addressed in [23]. In [24] an iterative peak matching algorithm for the calibration of a wireless acoustic sensor network is described in an unsynchronized network by using a fast calibration process. The method is valid for nodes that incorporate a microphone and a loudspeaker and is based on the use of a set of orthogonal probe signals that are assigned to the nodes of the network. The correlation properties of pseudonoise sequences are exploited to simultaneously estimate the relative TOAs from multiple acoustic nodes, substantially reducing the total calibration time. In a final step, synchronization issues are removed by following a BeepBeep strategy [106, 110], providing range estimates that are converted to absolute node positions by means of multidimensional scaling [104].

## 9. Challenges and Future Directions

*9.1. Practical Challenges.* Some real-world challenges arise in the design of localization systems using WASNs. To build a robust and accurate localization system, it is necessary to ensure a tradeoff among aspects related to cost, energy effectiveness, ease of calibration, deployment difficulty, and precision. Achieving such tradeoff is not straightforward and encompasses many practical challenges as discussed below.

*9.1.1. Cost-Effectiveness.* The potentiality of WASNs to provide high-accuracy acoustic localization features is highly dependent on the underlying hardware technologies in the nodes. For example, localization techniques based on DOA, TDOA, or SRP need intensive in-node processing for computing GCCs as well as input resources permitting multichannel audio recording. With the advent of powerful single-board computers, high-performance in-node signal processing can be easily achieved. Nonetheless, important aspects should be considered regarding cost and energy dissipation, especially for battery-powered nodes that are massively deployed.

*9.1.2. Deployment Issues.* Localization methods usually require a predeployment configuration process. For example, TOA-based methods usually need to properly set up synchronization mechanisms before starting to localize targets. Similarly, energy-based methods need nodes with calibrated gains in order to obtain high-quality energy-ratio measurements. All these tasks are usually complex and time-consuming. Moreover, they tend to need human supervision during an offline profiling phase. Such predeployment phase can become even more complex in some application

environments where nodes can be accessed by unauthorized subjects. Moreover, WASNs are also applied outside of closed buildings. Thus they are subject to daily and seasonal temperature variations and corresponding variations of the speed of sound [35]. To cope with this shortcoming, calibration needs to be automated and made environment adaptive.

*9.1.3. System Resiliency.* Besides predeployment issues, a WASN should also implement self-configuration mechanisms dealing with network dynamics such as those related to node failures. In this context, system design must take into account the number of anchor nodes that are needed in the deployment and their placement strategy. The system should assure that, if some of the nodes get out of the network, the rest are still able to provide location estimates appropriately. To this end, it is important to maximize the coverage area while minimizing the number of required anchor nodes in the system.

*9.1.4. Scalability.* Depending on the specific application, the WASN that needs to be deployed can vary from a very small and simple network of a few nodes to very large WASNs with tens or hundreds of nodes and complex network topologies. For example, in wildlife monitoring applications a very large number of sensors are utilized to acoustically monitor very large environments, while the topology of the network can be constantly changing due to sensors being displaced by the wind or by passing animals. The challenge in such applications is to design localization and self-configuration methods that can easily scale to complex WASNs.

*9.1.5. Measurement Errors.* It is well known that RF-based localization in WSNs are prone to errors due to irregular propagation patterns induced by environmental conditions (pressure and temperature) and random multipath effects such as reflection, refraction, diffraction, and scattering. In the case of WASNs, acoustic signals are also subject to similar distortions caused by environmental changes and effects produced by noise and interfering sources, reflected echoes, object obstruction, or signal diffraction. Other errors are related to the aforementioned predeployment process, resulting in synchronization errors or inaccuracies in the positions of anchor nodes. These errors must be analyzed in order to filter measurement noise out and improve the accuracy of location estimates.

*9.1.6. Benchmarking.* In terms of performance evaluation, so far there are no specific benchmarking methodologies and datasets for the location estimation problem in WASNs. Due to their heterogeneity—in terms of sensors, number of sensors and microphones, topology, and so on—works comparing different localization methods using a common sensor setup are difficult to find in the literature. The definition of formal methodologies in order to evaluate localization performance and the recording of evaluation datasets using real-life WASNs still remain a big challenge.

## 9.2. Future Directions

**9.2.1. Real-Life Application.** In our days, the need for real-life realizations of WASN with sound source localization capabilities is becoming more and more evident. An important direction in the future will thus be the application of the localization methodologies in real-life WASNs. In this direction, the integration of methodologies from a diverse range of scientific fields will be of paramount importance. Such fields include networks (e.g., to design the communication and synchronization protocols), network administration (e.g., to organize the nodes of the network and identify and handle potential failures), signal processing (e.g., to estimate the sources' locations with many potential applications), and hardware design (e.g., to design the acoustic nodes that can operate individually featuring communication and multi-channel audio processing capabilities in a power efficient way). While many of these fields have flourished individually, the practical issues that will arise from their integration in practical WASNs still remain unseen and the need for benchmarking and methodologies for their efficient integration is becoming more and more urgent.

**9.2.2. Machine Learning-Based Approaches.** One of the practical challenges for the deployment of real-life applications is the huge variability of acoustic signals received at the WASNs due to the acoustic signal propagation in the physical domain as well as the inaccuracies caused at the system level and uncertainties associated with measurements of TOAs and TDOAs. With the help of large datasets and vastly increased computational power of off-the-shelf processors, these variabilities can be learned by machines for designing more robust algorithms.

## 10. Conclusion

Sound source localization through WASNs offers great potential for the development of location-aware applications. Although many methods for locating acoustic sources have been proposed during the last decades, most methods assume synchronized input signals acquired by a traditional microphone array. As a result, when designing WASN-oriented applications, many assumptions of traditional localization approaches have to be revisited. This paper has presented a review of sound source localization methods using commonly used measurements in WASNs, namely, energy, direction of arrival (DOA), time of arrival (TOA), time difference of arrival (TDOA), and steered response power (SRP). Moreover, since most algorithms assume perfect knowledge on the node locations, self-localization methods used to estimate the location of the nodes in the network are also of high interest within a WASN context. The practical challenges and future directions arising in the deployment of WASNs have also been discussed, emphasizing important aspects to be considered in the design of real-world applications relying on acoustic localization systems.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Authors' Contributions

All authors contributed equally to this work.

## References

- [1] A. Swami, Q. Zhao, Y. W. Hong, and L. Tong, Eds., *Wireless Sensor Networks: Signal Processing and Communications*, Wiley, West Sussex, UK, 2007.
- [2] J. Segura-Garcia, S. Felici-Castell, J. J. Perez-Solano, M. Cobos, and J. M. Navarro, "Low-cost alternatives for urban noise nuisance monitoring using wireless sensor networks," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 836–844, 2015.
- [3] M. Cobos, J. J. Perez-Solano, and L. T. Berger, "Acoustic-based technologies for ambient assisted living," in *Introduction to Smart eHealth and eCare Technologies*, pp. 159–177, CRC Press, Taylor & Francis, 2017.
- [4] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: a signal processing perspective," in *Proceedings of the 18th IEEE Symposium on Communications and Vehicular Technology in the Benelux (SCVT '11)*, pp. 1–6, IEEE, Ghent, Belgium, November 2011.
- [5] V. C. Raykar, I. V. Kozintsev, and R. Lienhart, "Position calibration of microphones and loudspeakers in distributed computing platforms," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 70–83, 2005.
- [6] L. Cheng, C. Wu, Y. Zhang, H. Wu, M. Li, and C. Maple, "A survey of localization in wireless sensor network," *International Journal of Distributed Sensor Networks*, vol. 2012, Article ID 962523, 12 pages, 2012.
- [7] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Transactions on Systems, Man and Cybernetics C*, vol. 37, no. 6, pp. 1067–1080, 2007.
- [8] H. Wang, *Wireless sensor networks for acoustic monitoring [Ph.D. dissertation]*, University of California, Los Angeles (UCLA), Los Angeles, Calif, USA, 2006.
- [9] N. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking (MobiCom '00)*, pp. 32–43, August 2000.
- [10] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.
- [11] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 8, pp. 1479–1489, 2008.
- [12] A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 2, pp. 439–443, 2013.
- [13] D. Bechler and K. Kroschel, "Reliability criteria evaluation for TDOA estimates in a variety of real environments," in *Proceedings of the IEEE International Conference on Acoustics,*

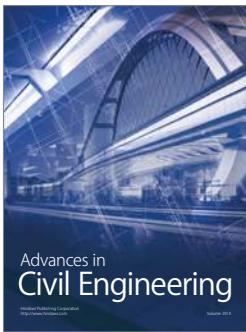
- Speech, and Signal Processing (ICASSP '05)*, vol. 4, pp. iv/985–iv/988, March 2005.
- [14] Y. A. Huang and J. Benesty, *Audio Signal Processing for Nextgeneration Multimedia Communication Systems*, Springer Science & Business Media, 2007.
- [15] A. Beck, P. Stoica, and J. Li, “Exact and approximate solutions of source localization problems,” *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1770–1778, 2008.
- [16] M. D. Gillette and H. F. Silverman, “A linear closed-form algorithm for source localization from time-differences of arrival,” *IEEE Signal Processing Letters*, vol. 15, pp. 1–4, 2008.
- [17] A. Canclini, P. Bestagini, F. Antonacci, M. Compagnoni, A. Sarti, and S. Tubaro, “A robust and low-complexity source localization algorithm for asynchronous distributed microphone networks,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 23, no. 10, pp. 1563–1575, 2015.
- [18] W. Meng and W. Xiao, “Energy-based acoustic source localization methods: a survey,” *Sensors*, vol. 17, no. 2, p. 376, 2017.
- [19] C. Meesookho, U. Mitra, and S. Narayanan, “On energy-based acoustic source localization for sensor networks,” *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 365–377, 2008.
- [20] D. B. Haddad, L. O. Nunes, W. A. Martins, L. W. P. Biscainho, and B. Lee, “Closed-form solutions for robust acoustic sensor localization,” in *Proceedings of the 14th IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '13)*, pp. 1–4, October 2013.
- [21] D. B. Haddad, W. A. Martins, M. D. V. M. Da Costa, L. W. P. Biscainho, L. O. Nunes, and B. Lee, “Robust acoustic self-localization of mobile devices,” *IEEE Transactions on Mobile Computing*, vol. 15, no. 4, pp. 982–995, 2016.
- [22] R. Pfeil, M. Pichler, S. Schuster, and F. Hammer, “Robust acoustic positioning for safety applications in underground mining,” *IEEE Transactions on Instrumentation and Measurement*, vol. 64, no. 11, pp. 2876–2888, 2015.
- [23] L. Wang, T.-K. Hon, J. D. Reiss, and A. Cavallaro, “Self-localization of ad-hoc arrays using time difference of arrivals,” *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 1018–1033, 2016.
- [24] M. Cobos, J. J. Perez-Solano, Ó. Belmonte, G. Ramos, and A. M. Torres, “Simultaneous ranging and self-positioning in unsynchronized wireless acoustic sensor networks,” *IEEE Transactions on Signal Processing*, vol. 64, no. 22, pp. 5993–6004, 2016.
- [25] D. Li and Y. H. Hu, “Energy-based collaborative source localization using acoustic microsensor array,” *Eurasip Journal on Applied Signal Processing*, vol. 2003, Article ID 985029, 2003.
- [26] X. Sheng and Y.-H. Hu, “Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. 53, no. 1, pp. 44–53, 2005.
- [27] D. Li and Y. H. Hu, “Least square solutions of energy based acoustic source localization problems,” in *Proceedings of the Workshops on Mobile and Wireless Networking/High Performance Scientific, Engineering Computing/Network Design and Architecture/Optical Networks Control and Management/Ad Hoc and Sensor Networks/Compil*, pp. 443–446, IEEE, Montreal, Canada, August 2004.
- [28] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, “A closed-form location estimator for use with room environment microphone arrays,” *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 1, pp. 45–50, 1997.
- [29] H. C. Schau and A. Z. Robinson, “Passive source localization employing intersecting spherical surfaces from time-of-arrival differences,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 8, pp. 1223–1225, 1987.
- [30] J. O. Smith and J. S. Abel, “The spherical interpolation method of source localization,” *IEEE Journal of Oceanic Engineering*, vol. 12, no. 1, pp. 246–252, 1987.
- [31] J. O. Smith and J. S. Abel, “Closed-form least-squares source location estimation from range-difference measurements,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 12, pp. 1661–1669, 1987.
- [32] K. C. Ho and M. Sun, “An accurate algebraic closed-form solution for energy-based source localization,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, no. 8, pp. 2542–2550, 2007.
- [33] D. Blatt and A. O. Hero III, “Energy-based sensor network source localization via projection onto convex sets,” *IEEE Transactions on Signal Processing*, vol. 54, no. 9, pp. 3614–3619, 2006.
- [34] H. Shen, Z. Ding, S. Dasgupta, and C. Zhao, “Multiple source localization in wireless sensor networks based on time of arrival measurement,” *IEEE Transactions on Signal Processing*, vol. 62, no. 8, pp. 1938–1949, 2014.
- [35] P. Annibale, J. Filos, P. A. Naylor, and R. Rabenstein, “TDOA-based speed of sound estimation for air temperature and room geometry inference,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 21, no. 2, pp. 234–246, 2013.
- [36] F. Thomas and L. Ros, “Revisiting trilateration for robot localization,” *IEEE Transactions on Robotics*, vol. 21, no. 1, pp. 93–101, 2005.
- [37] D. E. Manolakis, “Efficient solution and performance analysis of 3-D position estimation by trilateration,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 4, pp. 1239–1248, 1996.
- [38] I. Coope, “Reliable computation of the points of intersection of  $n$  spheres in  $R^n$ ,” *ANZIAM Journal*, vol. 42, pp. 461–477, 2000.
- [39] Y. Zhou, “An efficient least-squares trilateration algorithm for mobile robot localization,” in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3474–3479, IEEE, October 2009.
- [40] W. H. Foy, “Position-location solutions by Taylor’s series estimation,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 12, no. 2, pp. 187–194, 1976.
- [41] W. Navidi, W. S. Murphy Jr., and W. Hereman, “Statistical methods in surveying by trilateration,” *Computational Statistics and Data Analysis*, vol. 27, no. 2, pp. 209–227, 1998.
- [42] M. Pent, M. A. Spirito, and E. Turco, “Method for positioning GSM mobile stations using absolute time delay measurements,” *Electronics Letters*, vol. 33, no. 24, pp. 2019–2020, 1997.
- [43] M. Cobos, J. J. Perez-Solano, S. Felici-Castell, J. Segura, and J. M. Navarro, “Cumulative-sum-based localization of sound events in low-cost wireless acoustic sensor networks,” *IEEE/ACM Transactions on Speech and Language Processing*, vol. 22, no. 12, pp. 1792–1802, 2014.
- [44] D. Bechler and K. Kroschel, “Three different reliability criteria for time delay estimates,” in *Proceedings of the 12th European Signal Processing Conference*, pp. 1987–1990, September 2004.
- [45] P. Bestagini, M. Compagnoni, F. Antonacci, A. Sarti, and S. Tubaro, “TDOA-based acoustic source localization in the space-range reference frame,” *Multidimensional Systems and Signal Processing*, vol. 25, no. 2, pp. 337–359, 2014.

- [46] Y. Huang, "Source localization," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, chapter 8, pp. 229–253, Springer, 2004.
- [47] Y. Huang, J. Benesty, and G. W. Elko, "Passive acoustic source localization for video camera steering," in *Proceedings of the 25th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '00)*, vol. 2, pp. II909–II912, June 2000.
- [48] J. Abel and J. Smith, "The spherical interpolation method for closed-form passive source localization using range difference measurements," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '87)*, vol. 12, pp. 471–474.
- [49] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereau, "Real-time passive source localization: a practical linear-correction least-squares approach," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 943–956, 2001.
- [50] M. Compagnoni, P. Bestagini, F. Antonacci, A. Sarti, and S. Tubaro, "Localization of acoustic sources through the fitting of propagation cones using multiple independent arrays," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 7, pp. 1964–1975, 2012.
- [51] S. Argentieri and P. Danès, "Broadband variations of the MUSIC high-resolution method for sound source localization in robotics," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '07)*, pp. 2009–2014, October 2007.
- [52] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.
- [53] F. Nesta and M. Omologo, "Generalized state coherence transform for multidimensional TDOA estimation of multiple sources," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 1, pp. 246–260, 2012.
- [54] D. Pavlidi, A. Griffin, M. Puigt, and A. 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.
- [55] A. M. Torres, M. Cobos, B. Pueo, and J. J. Lopez, "Robust acoustic source localization based on modal beamforming and time-frequency processing using circular microphone arrays," *Journal of the Acoustical Society of America*, vol. 132, no. 3, pp. 1511–1520, 2012.
- [56] L. M. Kaplan, Q. Le, and P. Molnár, "Maximum likelihood methods for bearings-only target localization," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01)*, vol. 5, pp. 3001–3004, May 2001.
- [57] K. Doğançay, "Bearings-only target localization using total least squares," *Signal Processing*, vol. 85, no. 9, pp. 1695–1710, 2005.
- [58] L. M. Kaplan and Q. Le, "On exploiting propagation delays for passive target localization using bearings-only measurements," *Journal of the Franklin Institute*, vol. 342, no. 2, pp. 193–211, 2005.
- [59] A. N. Bishop, B. D. O. Anderson, B. Fidan, P. N. Pathirana, and G. Mao, "Bearing-only localization using geometrically constrained optimization," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 1, pp. 308–320, 2009.
- [60] Z. Wang, J.-A. Luo, and X.-P. Zhang, "A novel location-penalized maximum likelihood estimator for bearing-only target localization," *IEEE Transactions on Signal Processing*, vol. 60, no. 12, pp. 6166–6181, 2012.
- [61] R. G. Stansfield, "Statistical theory of D.F. fixing," *Journal of the Institute of Electrical Engineers—Part IIIA: Radiocommunication*, vol. 94, no. 15, pp. 762–770, 1947.
- [62] S. C. Nardone, A. G. Lindgren, and K. F. Gong, "Fundamental properties and performance of conventional bearings-only target motion analysis," *IEEE Transactions on Automatic Control*, vol. 29, no. 9, pp. 775–787, 1984.
- [63] K. Doğançay, "On the bias of linear least squares algorithms for passive target localization," *Signal Processing*, vol. 84, no. 3, pp. 475–486, 2004.
- [64] M. Gavish and A. J. Weiss, "Performance analysis of bearing-only target location algorithms," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 3, pp. 817–828, 1992.
- [65] K. Doğançay, "Passive emitter localization using weighted instrumental variables," *Signal Processing*, vol. 84, no. 3, pp. 487–497, 2004.
- [66] Y. T. Chan and S. W. Rudnicki, "Bearings-only and doppler-bearing tracking using instrumental variables," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 28, no. 4, pp. 1076–1083, 1992.
- [67] K. Doğançay, "Bias compensation for the bearings-only pseudolinear target track estimator," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 59–68, 2006.
- [68] K. Doğançay, "Reducing the bias of a bearings-only tls target location estimator through geometry translations," in *Proceedings of the European Signal Processing Conference (EUSIPCO '04)*, pp. 1123–1126, IEEE, Vienna, Austria, September 2004.
- [69] A. Griffin and A. Mouchtaris, "Localizing multiple audio sources from DOA estimates in a wireless acoustic sensor network," in *Proceedings of the 14th IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '13)*, pp. 1–4, IEEE, October 2013.
- [70] A. Griffin, A. Alexandridis, D. Pavlidi, Y. Mastorakis, and A. Mouchtaris, "Localizing multiple audio sources in a wireless acoustic sensor network," *Signal Processing*, vol. 107, pp. 54–67, 2015.
- [71] A. Griffin, A. Alexandridis, D. Pavlidi, and A. Mouchtaris, "Real-time localization of multiple audio sources in a wireless acoustic sensor network," in *Proceedings of the 22nd European Signal Processing Conference (EUSIPCO '14)*, pp. 306–310, September 2014.
- [72] A. Alexandridis, G. Borboudakis, and A. Mouchtaris, "Addressing the data-association problem for multiple sound source localization using DOA estimates," in *Proceedings of the 23rd European Signal Processing Conference (EUSIPCO '15)*, pp. 1551–1555, August 2015.
- [73] L. M. Kaplan, P. Molnár, and Q. Le, "Bearings-only target localization for an acoustical unattended ground sensor network," in *Unattended Ground Sensor Technologies and Applications III*, vol. 4393 of *Proceedings of SPIE*, pp. 40–51, April 2001.
- [74] K. R. Pattipati, S. Deb, Y. Bar-Shalom, and R. B. Washburn, "A new relaxation algorithm and passive sensor data association," *IEEE Transactions on Automatic Control*, vol. 37, no. 2, pp. 198–213, 1992.
- [75] S. Deb, M. Yeddanapudi, K. Pattipati, and Y. Bar-Shalom, "A generalized S-D assignment algorithm for multisensor-multitarget state estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 33, no. 2, pp. 523–538, 1997.
- [76] R. L. Popp, K. R. Pattipati, and Y. Bar-Shalom, "M-best S-D assignment algorithm with application to multitarget tracking,"

- IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 1, pp. 22–39, 2001.
- [77] J. Reed, C. da Silva, and R. Buehrer, “Multiple-source localization using line-of-bearing measurements: approaches to the data association problem,” in *Proceedings of the IEEE Military Communications Conference (MILCOM '08)*, pp. 1–7, November 2008.
- [78] M. Swartling, B. Sällberg, and N. Grbić, “Source localization for multiple speech sources using low complexity non-parametric source separation and clustering,” *Signal Processing*, vol. 91, no. 8, pp. 1781–1788, 2011.
- [79] M. Cobos and J. J. Lopez, “Maximum a posteriori binary mask estimation for underdetermined source separation using smoothed posteriors,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 20, no. 7, pp. 2059–2064, 2012.
- [80] S. Schulz and T. Herfet, “On the window-disjoint-orthogonality of speech sources in reverberant humanoid scenarios,” in *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx '08)*, pp. 241–248, September 2008.
- [81] S. T. Roweis, “Factorial models and refiltering for speech separation and denoising,” in *Proceedings of the European Conference on Speech Communication (EUROSPEECH '03)*, 2003.
- [82] M. Taseska and E. A. P. Habets, “Informed spatial filtering for sound extraction using distributed microphone arrays,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 22, no. 7, pp. 1195–1207, 2014.
- [83] A. Alexandridis and A. Mouchtaris, “Multiple sound source location estimation and counting in a wireless acoustic sensor network,” in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '15)*, pp. 1–5, October 2015.
- [84] A. Alexandridis, N. Stefanakis, and A. Mouchtaris, “Towards wireless acoustic sensor networks for location estimation and counting of multiple speakers in real-life conditions,” in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '17)*, pp. 6140–6144, New Orleans, LA, USA, March 2017.
- [85] J. H. DiBiase, *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays [Ph.D. dissertation]*, Brown University, Providence, RI, USA, 2000.
- [86] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, “Robust localization in reverberant rooms,” in *Microphone Arrays: Techniques and Applications*, pp. 157–180, Springer, Berlin, Germany, 2001.
- [87] M. Cobos, A. Marti, and J. 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, 2011.
- [88] M. Omologo, P. Svaizer, and R. DeMori, “Acoustic transduction,” in *Spoken Dialogue with Computers*, chapter 2, pp. 23–64, Academic Press, San Diego, California, USA, 1998.
- [89] S. Astapov, J. Berdnikova, and J.-S. Preden, “Optimized acoustic localization with SRP-PHAT for monitoring in distributed sensor networks,” *International Journal of Electronics and Telecommunications*, vol. 59, no. 4, pp. 383–390, 2013.
- [90] S. Astapov, J.-S. Preden, and J. Berdnikova, “Simplified acoustic localization by linear arrays for wireless sensor networks,” in *Proceedings of the 18th International Conference on Digital Signal Processing (DSP '13)*, pp. 1–6, July 2013.
- [91] S. Astapov, J. Berdnikova, and J.-S. Preden, “A two-stage approach to 2D DOA estimation for a compact circular microphone array,” in *Proceedings of the 4th International Conference on Informatics, Electronics and Vision (ICIEV '15)*, pp. 1–6, June 2015.
- [92] H. Do, H. F. Silverman, and Y. Yu, “A real-time SRP-PHAT source location implementation using stochastic region contraction (SRC) on a large-aperture microphone array,” in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '07)*, vol. 1, pp. 1–121–1–124, April 2007.
- [93] H. Do and H. F. Silverman, “A fast microphone array SRP-PHAT source location implementation using coarse-to-fine region contraction (CFRC),” in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA '07)*, pp. 295–298, October 2007.
- [94] L. O. Nunes, W. A. Martins, M. V. Lima et al., “A steered-response power algorithm employing hierarchical search for acoustic source localization using microphone arrays,” *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5171–5183, 2014.
- [95] A. Marti, M. Cobos, J. J. Lopez, and J. Escolano, “A steered response power iterative method for high-accuracy acoustic source localization,” *Journal of the Acoustical Society of America*, vol. 134, no. 4, pp. 2627–2630, 2013.
- [96] T. Lee, S. Chang, and D. Yook, “Parallel SRP-PHAT for GPUs,” *Computer Speech & Language*, vol. 35, pp. 1–13, 2016.
- [97] V. P. Minotto, C. R. Jung, L. G. Da Silveira Jr., and B. Lee, “GPU-based approaches for real-time sound source localization using the SRP-PHAT algorithm,” *International Journal of High Performance Computing Applications*, vol. 27, no. 3, pp. 291–306, 2013.
- [98] J. A. Belloch, A. Gonzalez, A. M. Vidal, and M. Cobos, “On the performance of multi-GPU-based expert systems for acoustic localization involving massive microphone arrays,” *Expert Systems with Applications*, vol. 42, no. 13, pp. 5607–5620, 2015.
- [99] J. Velasco, C. J. Martín-Arguedas, J. Macias-Guarasa, D. Pizarro, and M. Mazo, “Proposal and validation of an analytical generative model of SRP-PHAT power maps in reverberant scenarios,” *Signal Processing*, vol. 119, pp. 209–228, 2016.
- [100] M. Cobos, M. Garcia-Pineda, and M. Arevalillo-Herraez, “Steered response power localization of acoustic passband signals,” *IEEE Signal Processing Letters*, vol. 24, no. 5, pp. 717–721, 2017.
- [101] F. J. Álvarez, T. Aguilera, and R. López-Valcarce, “CDMA-based acoustic local positioning system for portable devices with multipath cancellation,” *Digital Signal Processing*, vol. 62, pp. 38–51, 2017.
- [102] V. C. Raykar and R. Duraiswami, “Automatic position calibration of multiple microphones,” in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '04)*, vol. 4, pp. 69–72, IEEE, May 2004.
- [103] J. M. Sachar, H. F. Silverman, and W. R. Patterson, “Microphone position and gain calibration for a large-aperture microphone array,” *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 1, pp. 42–52, 2005.
- [104] S. T. Birchfield and A. Subramanya, “Microphone array position calibration by basis-point classical multidimensional scaling,” *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1025–1034, 2005.

- [105] I. McCowan, M. Lincoln, and I. Himawan, "Microphone array shape calibration in diffuse noise fields," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 3, pp. 666–670, 2008.
- [106] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "BeepBeep: a high accuracy acoustic ranging system using COTS mobile devices," in *Proceedings of the 5th ACM International Conference on Embedded Networked Sensor Systems (SenSys'07)*, pp. 1–14, November 2007.
- [107] M. H. Hennecke and G. A. Fink, "Towards acoustic self-localization of ad hoc smartphone arrays," in *Proceedings of the 3rd Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA '11)*, pp. 127–132, Edinburgh, UK, May–June 2011.
- [108] P. Stoica and J. Li, "Source Localization from Range-Difference Measurements," *IEEE Signal Processing Magazine*, vol. 23, no. 6, pp. 63–66, 2006.
- [109] F. J. Álvarez and R. López-Valcarce, "Multipath cancellation in broadband acoustic local positioning systems," in *Proceedings of the 9th IEEE International Symposium on Intelligent Signal Processing (WISP '15)*, pp. 1–6, IEEE, Siena, Italy, May 2015.
- [110] C. Peng, G. Shen, and Y. Zhang, "BeepBeep: a high-accuracy acoustic-based system for ranging and localization using COTS devices," *Transactions on Embedded Computing Systems*, vol. 11, no. 1, pp. 4:1–4:29, 2012.





**Hindawi**

Submit your manuscripts at  
<https://www.hindawi.com>

