IEEE Open Journal of Signal Processing

Received 25 June 2021; revised 3 October 2021; accepted 3 October 2021. Date of publication 11 October 2021; date of current version 27 October 2021. The review of this article was arranged by Associate Editor Dr. Hiroshi Sawada.

*Digital Object Identifier 10.1109/OJSP.2021.3118574*

# Joint Calibration and Synchronization of Two Arrays of Microphones and Loudspeakers Using Particle Swarm Optimization

**ANTON KOVALYOV, KASHYAP PATEL, AND ISSA PANAHI**

Department of Electrical and Computer Engineering, University of Texas at Dallas, Richardson, TX 75080 USA

CORRESPONDING AUTHORS: ANTON KOVALYOV; KASHYAP PATEL (e-mail: anton.kovalyov@utdallas.edu; patelkashyap@utdallas.edu)

This work was supported by the National Institute on Deafness and Other Communication Disorders (NIDCD) of the National Institutes of Health (NIH) under Award 5R01DC015430-05.

**ABSTRACT** This work presents a methodology for the joint calibration and synchronization of two arrays of microphones and loudspeakers. The problem is modeled as estimation of the rigid motion of one array with respect to the other, as well as estimation of the synchronization mismatch between the two. The proposed method uses dedicated signals emitted by the loudspeakers of the two arrays to compute a set of time of arrival (TOA) estimates. Through a simple transformation, estimated TOAs are converted into a set of linearly independent time difference of arrival (TDOA) measurements, which are modeled by a system of nonlinear equations in the unknown parameters of interest. A maximum likelihood estimate is then given as the solution to a nonlinear weighted least squares (NWLS) problem, which is optimized applying a parallelizable variant of Particle Swarm Optimization (PSO). In this paper, we also derive the Cramér-Rao lower bound (CRLB), and benchmark it against the proposed method in a series of Monte Carlo (MC) simulations. Results show that the proposed method attains high-performance comparable to the CRLB.

**INDEX TERMS** Microphone array, calibration, localization, synchronization, PSO.

## I. INTRODUCTION

Microphone arrays can be employed to determine the space-time structure of an acoustic field. They have been used in many practical applications, including speech enhancement [1], sound source localization (SSL) [2], direction of arrival (DOA) estimation [3] and tracking [4]. The performance of these applications generally improves as the number of spatially distributed microphones being deployed increases. This is where wireless acoustic sensor networks (WASNs) are of special interest. A WASN simulates an ad-hoc array of spatially distributed microphones using an array of acoustic sensor nodes interconnected by a wireless medium [5]. Each node includes a processor, a wireless transmitter and receiver, an array of one or more microphones, and possibly one or more loudspeakers. These node characteristics are nowadays easily satisfied by many commercial off-the-shelf (COTS) devices, such as laptops, tablets and smartphones.

Most multi-channel signal processing techniques, such as acoustic beamforming [6] and SSL/DOA based on time difference of arrival (TDOA) measurements [7], rely on precise knowledge of microphone array geometry, i.e., relative 3-dimensional (3D) microphone positions, and the assumption that the multiple audio input channels are synchronized. These constraints can be especially hard to achieve in a WASN, where nodes are generally asynchronous, and their relative positions are not necessarily fixed. In these situations, an automatic mechanism for geometric calibration, also known simply as calibration, as well as synchronization of the multiple audio input channels, is desired.

A lot of research has been done on WASN calibration. Approaches in literature often model the problem as estimation of microphone pairwise distances, which are then transformed into relative 3D microphone positions applying multidimensional scaling (MDS) [8]–[13]. Detailed information on MDS can be found in [14]. Calibration methodology can be classified into two types: passive calibration, also known as self-calibration; and active calibration. Passive calibration methods estimate the WASN geometry using acoustic

signals in the environment. Active calibration methods estimate the WASN geometry using dedicated signals generated by built-in loudspeakers within the nodes in the network. The concept of passive calibration is generally preferred in practice since it does not rely on the emission of potentially disruptive signals of active calibration methods. However, passive calibration methods proposed in the literature typically make certain assumptions about the environment which may jeopardize their implementation in some systems.

Work on passive calibration includes [8]–[10], [15]–[17]. Chen *et al.* [8] assumed acoustic sources and microphones laying on a 2D plane and used energy measurements to estimate positions of both simultaneously. McCowan *et al.* [9] assumed synchronous microphones in a diffused noise environment to estimate microphone pairwise distances by fitting measured noise coherence with its theoretical model. Hon *et al.* [10] assumed acoustic sources at end-fire locations to estimate microphone pairwise distances using TDOA measurements. [15]–[17] assumed sources at far field and individual nodes equipped with a synchronized microphone array capable of reliable DOA estimation. These methods use DOAs observed at individual nodes and TDOAs observed between the microphones of different nodes to estimate the WASN geometry.

On the other hand, work on active calibration includes [11]–[13], [18]. Peng *et al.* [18] proposed a system called "BeepBeep" which estimates the distance between two asynchronous nodes. Each node includes a microphone and a loudspeaker conveniently placed near each other. The loudspeakers emit a special "Beep" signal sequentially and a set of TOAs are estimated using the signals acquired by the microphones. Then, an approximation of the distance between the nodes is found by applying a simple algebraic manipulation on the TOA measurements. Cobos *et al.* [11] later expanded upon the BeepBeep system to allow simultaneous emission of Beep signals among two or more nodes to compute approximate pairwise distances, which greatly reduces calibration time. Their method excites individual loudspeakers simultaneously with a specific pseudonoise (PN) sequence, which is known for its high autocorrelation and low cross-correlation properties, followed by applying self-interference cancellation to the captured signals to improve TOA estimation. Raykar *et al.* [12] proposed a method that estimates node pairwise distances using a similar strategy to that of BeepBeep with the addition that pairwise distances are then converted into relative node 3D positions employing MDS, followed by applying the Levenberg-Marquardt algorithm (LMA) to further refine estimated positions of microphones and loudspeakers within nodes. Pertila *et al.* [13] proposed a calibration method for a WASN where individual nodes include an array of synchronized microphones and one loudspeaker. Their method follows similar steps to that of Raykar *et al.* with two main differences: TOA estimation was improved using known microphone array geometry within individual nodes; and instead of estimating individual microphone and loudspeaker positions, DOAs observed at individual nodes were used to find the node orientations.

The aforementioned calibration methods consider nodes including one or more microphones and zero or one loudspeaker, at least when formulating the problem mathematically. However, in practice, acoustic sensor nodes may also include an array of loudspeakers, e.g., many of the aforementioned COTS devices come equipped with a microphone array and stereo loudspeakers. Consequently, it is of interest to develop an efficient joint calibration and synchronization method that uses all microphones and loudspeakers within individual nodes. Such a method can greatly benefit from the following two assumptions generally true in practice: intra-array geometry is known, that is, the relative 3D positions of microphones and loudspeakers within individual nodes are known; and intra-array audio input channels are synchronized. Therefore, this work offers a method for the joint calibration and synchronization of two arrays of microphones and loudspeakers, which, to the best of our knowledge, has not been addressed before. As shown in this paper, an increased number of elements within the two arrays helps improve estimation of inter-array geometry and inter-array synchronization mismatch. Hence, the proposed method can be efficiently applied to jointly calibrate and synchronize a WASN whose individual nodes include an array of microphones and loudspeakers.

The proposed method models the problem of joint calibration and synchronization of two arrays of microphones and loudspeakers as estimation of the rigid motion, that is 3D rotation and 3D translation, of one array with respect to the other, as well as estimation of the synchronization offset between the two. The method uses signals emitted by the loudspeakers of the two arrays to compute a set of TOA measurements. Through a simple transformation, measured TOAs are then converted into a set of linearly independent TDOA estimates, which, applying the assumptions of known intra-array geometry and synchronized intra-array audio input channels, are modeled by a system of nonlinear equations in the unknown parameters of interest. A maximum likelihood (ML) estimate is then derived as the solution to a nonlinear weighted least square (NWLS) problem. Then, a parallelizable variant of Particle Swarm Optimization (PSO) is proposed to optimize the NWLS problem. In this work, we also include derivation of the Cramér-Rao lower bound (CRLB), which is benchmarked against the proposed method in a series of Monte Carlo (MC) simulations. Results show that the proposed method attains the CRLB in most cases.

This paper is structured as follows. The problem of joint calibration and synchronization of two arrays of microphones and loudspeakers is described in Section II. The solution, given by optimization of a NWLS problem, is derived in Section III. A numerical optimization method based on PSO is presented in Section IV. Derivation of the CRLB is shown in Section V. The proposed method is evaluated in a series of simulation experiments in Section VI. Finally, Section VII concludes the paper.

By convention, vectors in this paper are column vectors. They are denoted by lower case bold letters/symbols. Matrices are denoted by upper case bold letters/symbols. $\mathbf{x}(i)$ is the $i$-th element of $\mathbf{x}$. $\mathbf{X}^T$ is the transpose of $\mathbf{X}$. $||\mathbf{x}||$ is the Euclidean

norm of $\mathbf{x}$. $\odot$ is the Schur product. $\mathbb{E}[\cdot]$ denotes expectation. $\text{mod}(a, b)$ denotes $a$ modulo $b$. $\tilde{(\cdot)}$ denotes a known estimate. $\hat{(\cdot)}$ denotes an unknown estimate that needs to be found. Finally, $\Delta(\cdot)$ denotes additive noise modeled as a zero-mean random variable.

## II. PROBLEM FORMULATION

Let us consider two arrays of microphones and loudspeakers. We will refer to one array as *primary array* (PA) and to the other as *secondary array* (SA). It is assumed that each array is properly calibrated and synchronized, i.e., intra-array geometry is known, and intra-array audio input channels are synchronized. However, the position of SA with respect to PA, as well as the synchronization offset between the two, are not known and need to be estimated.

### A. LOCAL COORDINATE SYSTEMS AND RIGID MOTION

Known intra-array geometry is represented here by defining two local coordinate systems (LCSs), one for PA and one for SA. Let $I$ and $J$ be the number of loudspeakers and microphones in PA, respectively. Similarly, let $K$ and $L$ be the number of loudspeakers and microphones in SA, respectively. The 3D positions of PA's $i$-th loudspeaker and $j$-th microphone, associated to PA's LCS, are represented by $\mathbf{s}_i$ and $\mathbf{m}_j$, respectively, where $i = 1, \ldots, I$ and $j = 1, \ldots, J$. On the other hand, the 3D positions of SA's $k$-th loudspeaker and $l$-th microphone, associated to SA's LCS, are represented by $\mathbf{s}_k^\circ$ and $\mathbf{m}_l^\circ$, respectively. For mathematical convenience we let $k = I+1, \ldots, I+K$ and $l = J+1, \ldots, J+L$. Consequently, the position of SA with respect to PA is modeled by

$$\begin{aligned} \mathbf{s}_k &= \mathbf{R}\mathbf{s}_k^\circ + \mathbf{t} \\ \mathbf{m}_l &= \mathbf{R}\mathbf{m}_l^\circ + \mathbf{t}\,, \end{aligned} \tag{1}$$

where $\mathbf{R}$ is a $3 \times 3$ rotation matrix and $\mathbf{t}$ is a $3 \times 1$ translation vector defining the rigid motion that brings SA's LCS to that of PA.

It is well known that the maximum number of degrees of freedom of a rigid body in 3D space is six, that is, three coordinates are required to locate its center of mass and another three to describe its orientation [19]. Consequently, using (1) to model the problem of calibration of SA with respect to PA is especially convenient, since the maximum number of parameters that need to be estimated is fixed regardless of the number of elements in SA. This work is flexible in how $\mathbf{R}$ and $\mathbf{t}$ are parametrized. For the purpose of demonstration, $\mathbf{R}$ can be parametrized, using Tait-Bryan representation, as

$$\mathbf{R} = \mathbf{R}_z(\alpha)\mathbf{R}_y(\beta)\mathbf{R}_x(\gamma)\,, \tag{2}$$

where the *yaw* angle $\alpha \in [-\pi, \pi]$, the *pitch* angle $\beta \in [-\pi/2, \pi/2]$ and the *roll* angle $\gamma \in [-\pi, \pi]$ are rotation parameters whose corresponding matrices represent rotation about the $z$, $y$ and $x$ axes, respectively. Similarly, $\mathbf{t}$ can be parametrized, using spherical representation, as

$$\mathbf{t} = [\rho\cos\theta\cos\phi,\ \rho\sin\theta\cos\phi,\ \rho\sin\phi]^T\,, \tag{3}$$

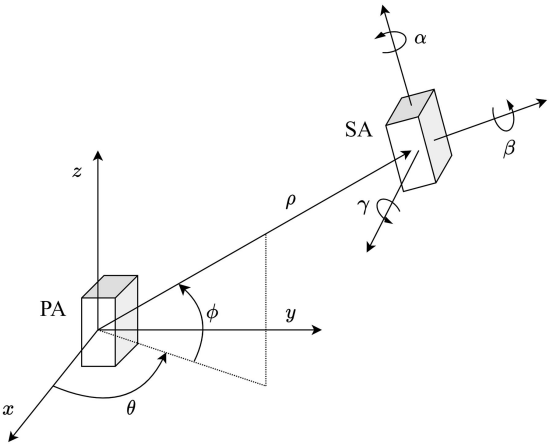


**FIGURE 1.** Parameters describing the 3D position of SA with respect to PA.

where $\rho \in [0, \infty]$, $\theta \in [-\pi, \pi]$ and $\phi \in [-\pi/2, \pi/2]$ are translation parameters denoting range, azimuth angle and elevation angle, respectively. Fig. 1 illustrates the problem of describing the 3D position of SA with respect to PA in terms of the aforementioned parameters.

### B. TOA

As part of the inter-array calibration procedure, each loudspeaker is excited with a known calibration signal such as a chirp signal or a PN sequence. Emitted signals are then captured by each microphone and TOAs are estimated. Let $\tau_{p,q}$ be the noise-free TOA of a signal emitted by loudspeaker $p$, when captured at microphone $q$, where $p = 1, \ldots, I+K$ and $q = 1, \ldots, J+L$. Assuming direct path between loudspeakers and microphones, corresponding noise-free TOAs are given by

$$\tau_{p,j} = c^{-1}||\mathbf{s}_p - \mathbf{m}_j|| + \tau_p + \delta_m \tag{4}$$

$$\tau_{p,l} = c^{-1}||\mathbf{s}_p - \mathbf{m}_l|| + \tau_p + \delta_s\,, \tag{5}$$

where $c$ is the propagation speed of the signal, $\tau_p$ is the physical time with reference to some global clock at which loudspeaker $p$ was excited, and $\delta_m$ and $\delta_s$ are time offsets due to different internal clocks and capture times at PA and SA, respectively. This formulation comes directly from the assumption made earlier that intra-array audio input channels are synchronized.

TOA estimation is similar to the problem of time delay estimation (TDE) in [20]. Hence, assuming no multipath and high signal-to-noise ratio, an estimate of $\tau_{p,q}$ can be found as the peak of the cross-correlation of the known signal emitted by loudspeaker $p$ and the signal captured by microphone $q$. For best results, calibration signals can be emitted sequentially. However, this may be impractical when the number of loudspeakers is large. In that case, the methodology for TOA estimation when loudspeakers emit calibration signals simultaneously, developed by Cobos *et al.* in [11], is of special interest. Moreover, known intra-array geometry can also be exploited to further improve estimation of TOAs using the

methodology developed by Pertila *et al.* in [13]. This work is independent of the methodology applied to estimate the TOAs in (4) and (5). Here, we assume direct knowledge of noisy TOA estimates modeled by

$$\tilde{\tau}_{p,q} = \tau_{p,q} + \Delta\tau_{p,q} \ , \tag{6}$$

where $\Delta\tau_{p,q}$ is, by definition, zero-mean additive noise.

### C. TDOA AND SYNCHRONIZATION OFFSET

The parameters $\tau_p$, $\delta_m$ and $\delta_s$ are of no interest to us. Subtracting (4) and (5) results in the following system of nonlinear TDOA equations

$$r_{p,j,l} = \tau_{p,j} - \tau_{p,l} \tag{7}$$

$$= c^{-1}\left(||\mathbf{s}_p - \mathbf{m}_j|| - ||\mathbf{s}_p - \mathbf{m}_l||\right) + \xi \ , \tag{8}$$

where $\xi = \delta_m - \delta_s$ is defined here as the synchronization offset between PA and SA that needs to be estimated. Once $\xi$ is found, the two arrays can be synchronized[1] by simply delaying SA's audio input channels by $\xi$.

### D. PROBLEM SUMMARY

The problem is formulated as finding the synchronization offset $\xi$ and the parameters defining the rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$, given the known parameters $c$, $\mathbf{s}_i$, $\mathbf{m}_j$, $\mathbf{s}_k^\circ$, $\mathbf{m}_l^\circ$ and $\tilde{\tau}_{p,q}$, where $i = 1, \ldots, I$, $j = 1, \ldots, J$, $k = I + 1, \ldots, I + K$, $l = J + 1, \ldots, J + L$, $p = 1, \ldots, I + K$, and $q = 1, \ldots, J + L$.

## III. PROBLEM SOLUTION

In this section, we derive the ML estimate of the parameters of interest based on noisy TDOA measurements. Additionally, we discuss the minimum number of microphones and loudspeakers required.

### A. ML ESTIMATE

Since the transformation of SA's LCS in (1) does not change relative distances between points, the TDOA representation in (8) can be conveniently rewritten as

$$r_{i,j,l} = c^{-1}\left(||\mathbf{s}_i - \mathbf{m}_j|| - ||\mathbf{s}_i - \mathbf{R}\mathbf{m}_l^\circ - \mathbf{t}||\right) + \xi \tag{9}$$

$$r_{k,j,l} = c^{-1}\left(||\mathbf{R}\mathbf{s}_k^\circ + \mathbf{t} - \mathbf{m}_j|| - ||\mathbf{s}_k^\circ - \mathbf{m}_l^\circ||\right) + \xi \ , \tag{10}$$

where the only unknowns on the right-hand side of (9) and (10) are the parameters of interest. An estimate of true TDOA $r_{p,j,l}$ is given by

$$\tilde{r}_{p,j,l} = \tilde{\tau}_{p,j} - \tilde{\tau}_{p,l} \tag{11}$$

$$= r_{p,j,l} + \Delta r_{p,j,l} \ , \tag{12}$$

where, using (6), we have

$$\Delta r_{p,j,l} = \Delta\tau_{p,j} - \Delta\tau_{p,l}. \tag{13}$$

[1]This work assumes no clock drift between devices, which is rarely true in practice. However, clock drift can be corrected using the Network Time Protocol (NTP) or the Global Positioning System (GPS). Other solutions specific to WASN can be found in [21], [22].

It then follows that $\Delta r_{p,j,l}$ is a zero-mean random variable whose second order statistics can be conveniently expressed in terms of the, assumed to be known, second order statistics of $\Delta\tau_{p,q}$ as follows

$$\begin{aligned}\mathbb{E}\left[\Delta r_{p,j,l}\Delta r_{p',j',l'}\right] &= \mathbb{E}\left[\Delta\tau_{p,j}\Delta\tau_{p',j'}\right] + \mathbb{E}\left[\Delta\tau_{p,l}\Delta\tau_{p',l'}\right] \\ &\quad - \mathbb{E}\left[\Delta\tau_{p,j}\Delta\tau_{p',l'}\right] - \mathbb{E}\left[\Delta\tau_{p,l}\Delta\tau_{p',j'}\right] \ ,\end{aligned} \tag{14}$$

for $p' = 1, \ldots, I + K$, $j' = 1, \ldots, J$, $l' = J + 1, \ldots, J + L$.

Note that although the total number of TDOA measurements defined in (11) is $(I + K)JL$, only $(I + K)(J + L - 1)$ are linearly independent, that is, there are $J + L - 1$ linearly independent TDOAs associated to a single loudspeaker. For a given loudspeaker $p$, a plausible set of linearly independent TDOA measurements can be grouped in vector form as follows

$$\tilde{\mathbf{r}}_p = \left[\tilde{r}_{p,1,J+1}, \ldots, \tilde{r}_{p,1,J+L},\ \tilde{r}_{p,2,J+1}, \ldots, \tilde{r}_{p,J,J+1}\right]^T \ , \tag{15}$$

where the first $L$ elements of $\tilde{\mathbf{r}}_p$ group TDOAs between a fixed PA microphone and varying SA microphones, i.e., $j = 1$ and $l = 1, \ldots, L$, and, similarly, the next $J - 1$ elements group remaining TDOAs between a fixed SA microphone and varying PA microphones, i.e., $j = 2, \ldots, J$ and $l = J + 1$.

Let

$$\tilde{\mathbf{r}} = \left[\tilde{\mathbf{r}}_1^T, \ldots, \tilde{\mathbf{r}}_{I+K}^T\right]^T \tag{16}$$

be a vector grouping all linearly independent TDOA measurements. Consequently,

$$\tilde{\mathbf{r}} = \mathbf{r} + \Delta\mathbf{r} \tag{17}$$

where $\mathbf{r}$ groups the corresponding noise-free TDOAs defined in (9) and (10), and $\Delta\mathbf{r}$ groups the corresponding noise terms defined in (13). Let

$$\mathbf{x} = \left[\mathbf{x}_R^T,\ \mathbf{x}_t^T,\ \xi\right]^T \ , \tag{18}$$

be a vector grouping the unknown parameters that need to be estimated, where $\mathbf{x}_R$ groups the parameters defining the rotation matrix $\mathbf{R}$, e.g., $\mathbf{x}_R = [\alpha, \beta, \gamma]^T$, and $\mathbf{x}_t$ groups the parameters defining the translation vector $\mathbf{t}$, e.g., $\mathbf{x}_t = [\rho, \theta, \phi]^T$. Let $\mathbf{r}(\mathbf{x})$ be a vector grouping all the corresponding noise-free linearly independent TDOAs constructed as a function of $\mathbf{x}$, more specifically, $\mathbf{R}$, $\mathbf{t}$ and $\xi$ in (9) and (10) are defined using the corresponding parameters in $\mathbf{x}$. Assuming $\Delta\mathbf{r}$ is zero-mean Gaussian, the ML estimate of $\mathbf{x}$ is given by the solution to the following NWLS problem

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}}\ \mathcal{L}(\mathbf{x}) \tag{19}$$

$$\mathcal{L}(\mathbf{x}) = (\tilde{\mathbf{r}} - \mathbf{r}(\mathbf{x}))^T\mathbf{Q}^{-1}(\tilde{\mathbf{r}} - \mathbf{r}(\mathbf{x})) \ , \tag{20}$$

where

$$\mathbf{Q} = \mathbb{E}\left[\Delta\mathbf{r}\Delta\mathbf{r}^T\right] \tag{21}$$

is the covariance matrix of the TDOA noise, whose individual elements are given by (14).

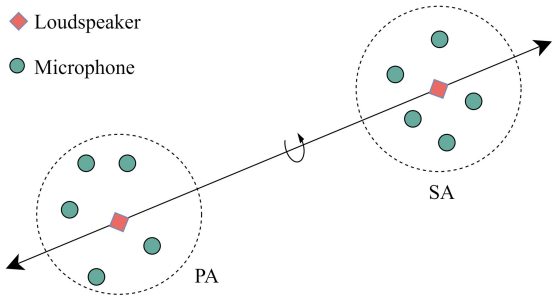


**FIGURE 2.** Ambiguity in 3D positioning of SA with respect to PA.

### B. MINIMUM NUMBER OF MICROPHONES AND LOUDSPEAKERS

Let $D = D_R + D_t + 1$ be the number of parameters in $\mathbf{x}$, where $D_R$ and $D_t$ are the number of parameters in $\mathbf{x}_R$ and $\mathbf{x}_t$, respectively. It follows that the constraint $(I + K)(J + L - 1) \geq D$ is a necessary condition for the estimator in (19) to work. However, this constraint is by no means a sufficient identifiability condition. For instance, let us consider a scenario where PA and SA each include a single loudspeaker and a large number of microphones satisfying the aforementioned constraint. If we allow all six degrees of freedom to SA, it is then easy to verify that the rotation of SA around the axis intersecting both loudspeakers, as illustrated in Fig. 2, will not affect pairwise distances between microphones and loudspeakers. This fact implies that identifiable 3D positioning of SA with respect to PA is not possible when $I = K = 1$ regardless of the number of microphones. On the other hand, letting $I + K > 2$ would break the ambiguity assuming the loudspeakers do not lie along a single line.

## IV. NUMERICAL OPTIMIZATION METHOD

Minimization of (20) is a highly nonlinear problem. A popular method for solving nonlinear problems is LMA. LMA is an iterative optimization method that combines gradient descent and Gauss-Newton methods. The problem with LMA is that it is prone to get stuck in local optimum and, as such, it greatly relies on the initial guess of the solution. Literature on active calibration methodology typically makes certain assumptions about the problem geometry to simplify finding an approximate solution to the particular nonlinear problem. The approximate solution could then be improved with LMA. A common assumption is that of microphones and loudspeakers corresponding to individual nodes being closely spaced [11]–[13], [18]. Here we prefer not to make additional assumptions and propose solving the nonlinear problem in (19) directly using PSO.

PSO is a well-known population-based metaheuristic (PM) applied in a wide variety of fields [23]. PMs involve optimization using a population of candidate solutions or particles that move around the search space in an iterative manner following a predefined update rule. PMs are stochastic in nature and unlike hill climbing approaches are considerably less reliant on initialization. Moreover, computation involving PMs is generally easy to parallelize, and as such, significant speed up can be obtained with the use of a General-Purpose Graphics Processing Unit (GPGPU). These factors make PMs an attractive choice for solving the optimization problem in this paper.

Apart from PSO, there exist many other well-known PMs, such as the Genetic Algorithm (GA) [24], Differential Evolution (DE) [25] and Artificial Bee Colony (ABC) [26]. Many variants of these algorithms were proposed in literature with the aim of improving not only the fitness of the solution for a given optimization problem but also convergence speed, that is, the number of function evaluations (FEs) it takes for the method to converge. Among GA, DE, ABC, and PSO, we found the latter to perform the best for solving (19). Hence, we will focus particularly on PSO.

As an attempt to balance local exploitation and global exploration, in PSO, a swarm of particles moves throughout the search space by means of acceleration towards a weighted combination of the best solution that they individually found and the best solution that other particles within their *neighborhood* found. A neighborhood is known as a predefined set of particles within the swarm that a given particle can communicate with. There exist many variants of PSO in literature. In this paper, we implement the constricted PSO with ring communication topology defined in [27] with a slight modification to allow parallelization of computations.

### A. SOLUTION SEARCH WITH PSO

PSO is applied to solve the optimization problem in (19) as follows. Let $N_i$ be the number of iterations. Let $N_p$ be the number of particles. A particle $n$, for $n = 0, \ldots, N_p - 1$, is represented by four vectors: its position $\mathbf{x}_n$, i.e., candidate solution to (19); its velocity $\mathbf{v}_n$; the best solution it individually found $\mathbf{p}_n$; and the best solution found by particles within its neighborhood $\mathbf{g}_n$. At each iteration of the algorithm, the solution search procedure updates the entire swarm as follows

$$\mathbf{v}_n = \chi \left[\mathbf{v}_n + c_1 \mathbf{a}_{1,n} \odot (\mathbf{x}_n - \mathbf{p}_n) + c_2 \mathbf{a}_{2,n} \odot (\mathbf{x}_n - \mathbf{g}_n)\right] \quad (22)$$

$$\mathbf{x}_n = \mathbf{x}_n + \mathbf{v}_n \quad (23)$$

$$\mathbf{p}_n = \operatorname*{argmin}_{\mathbf{x}} \{\mathcal{L}(\mathbf{x}) \mid \mathbf{x} \in \{\mathbf{p}_n, \mathbf{x}_n\}\} \quad (24)$$

$$\mathbf{g}_n = \operatorname*{argmin}_{\mathbf{x}} \{\mathcal{L}(\mathbf{x}) \mid \mathbf{x} \in G_n\}, \quad (25)$$

where $\chi$ is a parameter known as the constriction factor, $c_1$ and $c_2$ are parameters representing the attraction weights of $\mathbf{x}_n$ towards $\mathbf{p}_n$ and $\mathbf{g}_n$, respectively, $\mathbf{a}_{*,n}$ is a vector of size $D$ whose elements are drawn from $\mathcal{U}(0, 1)$, and $G_n$ is a set of best solutions found by particles within the neighborhood of particle $n$. Definition of $G_n$ depends on the communication topology used, which is explained in Section IV-B. The algorithm terminates once the final iteration is completed. Then, the PSO solution to (19) is given by

$$\hat{\mathbf{x}} = \operatorname*{argmin}_{\mathbf{x}} \left\{\mathcal{L}(\mathbf{x}) \mid \mathbf{x} \in \left\{\mathbf{p}_0, \ldots, \mathbf{p}_{N_p-1}\right\}\right\}. \quad (26)$$

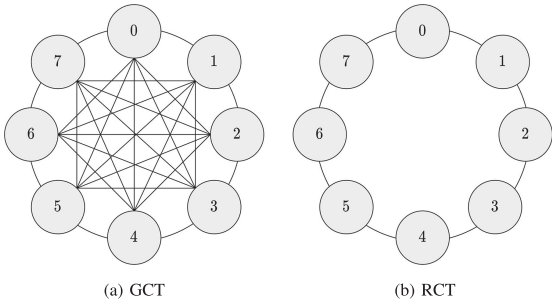


**FIGURE 3.** Particle swarm communication topologies.

### B. COMMUNICATION TOPOLOGY

Communication among particles is a common feature to all PMs, where particles collaborate to find the global optimum. In the context of PSO, the term communication topology gives definition to $G_n$ in (25). There are two well-known communication topologies in literature. One is the *Global Communication Topology* (GCT) and the other is the *Ring Communication Topology* (RCT). In GCT, a particle can communicate with all other particles including itself (see Fig. 3(a)), and as such

$$G_n = \left\{\mathbf{p}_0, \ldots, \mathbf{p}_{N_p-1}\right\} \ . \tag{27}$$

GCT implies that particles are directly attracted by the global best solution found by the algorithm, which in turn results in fast convergence rate. RCT, on the other hand, only allows a particle to communicate with itself and two adjacent particles in a ring structure (see Fig. 3(b)). It then follows that

$$G_n = \left\{\mathbf{p}_{\mathrm{mod}(n-1,N_p)}, \mathbf{p}_n, \mathbf{p}_{\mathrm{mod}(n+1,N_p)}\right\} \ . \tag{28}$$

Due to the overlapping nature of RCT, particles are still attracted by the global best solution found by the algorithm, but unlike GCT, the attraction is not direct, which in turn results in slower convergence rate. However, when implementing both to solve (19), we found that the slower convergence rate of RCT made PSO significantly less likely to converge prematurely when compared to the implementation with GCT. Hence, we propose solving (19) using PSO with RCT.

### C. SWARM INITIALIZATION

As is common in most implementations of PSO, the positions and velocities of the entire swarm are initialized using uniform distribution. Let $\mathbf{b}_l$ and $\mathbf{b}_u$ be vectors representing the lower and upper boundaries of the search space, respectively. Initially, we let

$$\begin{aligned} \mathbf{x}_n(d) &\sim \mathcal{U}(\mathbf{b}_l(d), \mathbf{b}_u(d)) \\ \mathbf{v}_n(d) &\sim \mathcal{U}(-\mathbf{v}_{max}(d), \mathbf{v}_{max}(d)) \\ & d = 1, \ldots, D \ , \end{aligned} \tag{29}$$

where $\mathbf{v}_{max} = \mathbf{b}_u - \mathbf{b}_l$.

### D. PARTICLES TRAVELLING OUTSIDE THE SEARCH SPACE

The swarm update rule in (23) does not prevent particles from travelling outside the search space, which is especially likely to happen at the first few iterations of the algorithm. As recommended in [27], to prevent a bias towards the center of the search space, the proposed variant of PSO puts little restriction on the trajectory of particles and allows them to travel outside the search boundaries. The idea is that the weighted attraction towards known optima in (22) will anyways pull particles back within the search space regardless of their current position. However, to avoid unfeasible solutions, whenever a particle is found to be outside the search boundaries, its FE, defined by (20) is not computed. Additionally, to prevent particles from developing excessively large velocities, their speeds are bounded by $\mathbf{v}_{max}$.

### E. PARAMETERS

The proposed optimization method includes five parameters whose values need to be specified, namely, the number of iterations $N_i$, the number of particles $N_p$, the constriction factor $\chi$, and the attraction weights $c_1$ and $c_2$. The first two parameters define the maximum number of FEs computed by the algorithm as $N_{FE} = N_p(N_i + 1)$. Although no exhaustive search was made, simulation results showed that when restricting the algorithm to $N_{FE} = 5 \times 10^5$, a good choice is to let $N_p = 100$ and consequently $N_i = 4999$. The last three parameters, on the other hand, are rarely tuned in PSO, instead, they are given the default assignments $\chi = 0.72894$ and $c_1 = c_2 = 2.05$. These default assignments are known to satisfy a constraint that guarantees convergence of PSO [27]. Proof of convergence of PSO can be found in [28].

### F. SIMULTANEOUS VS. SEQUENTIAL SWARM UPDATE

The swarm update rules in (22)-(25) can be interpreted in two ways: the update is sequential, that is, particles are updated sequentially, one at a time, using (22)-(25); the update is simultaneous, that is, all particles are updated simultaneously using (22), followed by (23) and so on. In sequential update, particles will tend to steer towards the best known solution faster when compared to simultaneous update, hence the trajectories of the swarms in these two approaches will be different. However, the overall behavior of the algorithm will remain the same. The simultaneous swarm update should be preferred in practice since computation can be parallelized/vectorized for all particles, which is especially convenient due to current trend on GPGPU computing and use of array programming languages such as MATLAB and Python. Nonetheless, perhaps for the sake of simplicity, PSO is generally described in literature using sequential swarm update, including the constricted PSO in [27] used as reference in this paper. We did not find noticeable difference in estimation performance when applying either approach to solve (19). Hence, due to its considerable computational speedup, in this work we use simultaneous swarm update.

### G. ALGORITHM SUMMARY

The proposed algorithm for solving (19), based on constricted PSO with RCT and simultaneous swarm update, is summarized as follows:

Step 1 Initialize the parameters $N_p$, $N_i$, $\chi$, $c_1$ and $c_2$ using the values defined in Section IV-E.

Step 2 Let $t = 0$.

Step 3 Initialize all $\mathbf{x}_n$ and $\mathbf{v}_n$, using (29).
Step 4 Initialize all $\mathbf{p}_n$ as $\mathbf{x}_n$.
Step 5 Compute FEs for all $\mathbf{x}_n$, using (20).
Step 6 Initialize all $\mathbf{g}_n$, using (25) and (28).
Step 7 Increment $t$.
Step 8 Update all $\mathbf{v}_n$, using (22).
Step 9 Update all $\mathbf{x}_n$, using (23).
Step 10 Compute FEs for all $\mathbf{x}_n$, using (20).
Step 11 Update all $\mathbf{p}_n$, using (24).
Step 12 Update all $\mathbf{g}_n$, using (25) and (28).
Step 13 If $t < N_i$, return to Step 7, otherwise, give final solution, using (26).

## V. CRLB

The CRLB places a lower bound on the variance of an unbiased estimator [29]. Hence, it is of interest to derive the CRLB for the problem in this paper to be later used as a benchmark against the proposed estimator based on PSO. The CRLB is given by the inverse of the Fisher information matrix (FIM). The FIM is found by

$$\mathcal{I}(\mathbf{x}) = -\mathbb{E}\left[\frac{\partial^2 \ln f(\tilde{\mathbf{r}}|\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T}\right], \tag{30}$$

where $f(\tilde{\mathbf{r}}|\mathbf{x})$ is the probability density function (PDF) of the measurements vector $\tilde{\mathbf{r}}$ conditioned on the parameters vector $\mathbf{x}$. In our context, $\mathbf{x}$ groups the true values of the unknown parameters in (18) and $\tilde{\mathbf{r}}$ groups the TDOA measurements in (11). The definition of $\tilde{\mathbf{r}}$ in (16) allows us to split it into

$$\tilde{\mathbf{r}}_m = [\tilde{\mathbf{r}}_1^T, \ldots, \tilde{\mathbf{r}}_I^T]^T, \tag{31}$$

which groups all linearly independent $\tilde{r}_{i,j,l}$, and

$$\tilde{\mathbf{r}}_s = [\tilde{\mathbf{r}}_{I+1}^T, \ldots, \tilde{\mathbf{r}}_{I+K}^T]^T, \tag{32}$$

which groups all linearly independent $\tilde{r}_{k,j,l}$. Consequently,

$$\tilde{\mathbf{r}}_m = \mathbf{r}_m + \Delta\mathbf{r}_m \tag{33}$$

and

$$\tilde{\mathbf{r}}_s = \mathbf{r}_s + \Delta\mathbf{r}_s \tag{34}$$

where $\mathbf{r}_m$ and $\mathbf{r}_s$ group the corresponding noise-free TDOAs $r_{i,j,l}$ and $r_{k,j,l}$, respectively, while $\Delta\mathbf{r}_m$ and $\Delta\mathbf{r}_s$ group the corresponding zero-mean noise terms $\Delta r_{i,j,l}$ and $\Delta r_{k,j,l}$, respectively. Let us assume that $\tilde{\mathbf{r}}_m$ and $\tilde{\mathbf{r}}_s$ are independent random vectors in $\mathcal{N}(\mathbf{r}_m, \mathbf{Q}_m)$ and $\mathcal{N}(\mathbf{r}_s, \mathbf{Q}_s)$, where

$$\mathbf{Q}_m = \mathbb{E}\left[\Delta\mathbf{r}_m \Delta\mathbf{r}_m^T\right] \tag{35}$$

and

$$\mathbf{Q}_s = \mathbb{E}\left[\Delta\mathbf{r}_s \Delta\mathbf{r}_s^T\right] \tag{36}$$

are the corresponding noise covariance matrices, whose elements are computed using the relationship of TOA and TDOA noise second order statistics in (14). It follows that the PDF of $\tilde{\mathbf{r}}$ conditioned on $\mathbf{x}$ is given by

$$f(\tilde{\mathbf{r}}|\mathbf{x}) = f(\tilde{\mathbf{r}}_m|\mathbf{x}) f(\tilde{\mathbf{r}}_s|\mathbf{x}) \,. \tag{37}$$

Taking the natural logarithm of both sides of (37) and expanding, we get

$$\ln f(\tilde{\mathbf{r}}|\mathbf{x}) = -\frac{1}{2}[(\tilde{\mathbf{r}}_m - \mathbf{r}_m)^T \mathbf{Q}_m^{-1} (\tilde{\mathbf{r}}_m - \mathbf{r}_m) + (\tilde{\mathbf{r}}_s - \mathbf{r}_s)^T \mathbf{Q}_s^{-1} (\tilde{\mathbf{r}}_s - \mathbf{r}_s)] + C \,, \tag{38}$$

where $C$ is some constant. The expectation of the double partial in (30) is then found to be

$$\mathbb{E}\left[\frac{\partial^2 \ln f(\tilde{\mathbf{r}}|\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T}\right] = -\left[\left(\frac{\partial \mathbf{r}_m}{\partial \mathbf{x}}\right)^T \mathbf{Q}_m^{-1} \left(\frac{\partial \mathbf{r}_m}{\partial \mathbf{x}}\right) + \left(\frac{\partial \mathbf{r}_s}{\partial \mathbf{x}}\right)^T \mathbf{Q}_s^{-1} \left(\frac{\partial \mathbf{r}_s}{\partial \mathbf{x}}\right)\right], \tag{39}$$

where computation of the right-hand side of (39) is straightforward given the following element-wise partials

$$\frac{\partial r_{i,j,l}}{\partial \mathbf{x}_R(a)} = \frac{1}{c}\left(\frac{\partial \mathbf{R}}{\partial \mathbf{x}_R(a)} \mathbf{m}_l^{\circ}\right)^T \frac{(\mathbf{s}_i - \mathbf{m}_l)}{||\mathbf{s}_i - \mathbf{m}_l||}$$

$$\frac{\partial r_{k,j,l}}{\partial \mathbf{x}_R(a)} = \frac{1}{c}\left(\frac{\partial \mathbf{R}}{\partial \mathbf{x}_R(a)} \mathbf{s}_k^{\circ}\right)^T \frac{(\mathbf{s}_k - \mathbf{m}_j)}{||\mathbf{s}_k - \mathbf{m}_j||}$$

$$a = 1, \ldots, D_R$$

$$\frac{\partial r_{i,j,l}}{\partial \mathbf{x}_t(b)} = \frac{1}{c}\left(\frac{\partial \mathbf{t}}{\partial \mathbf{x}_t(b)}\right)^T \frac{(\mathbf{s}_i - \mathbf{m}_l)}{||\mathbf{s}_i - \mathbf{m}_l||}$$

$$\frac{\partial r_{k,j,l}}{\partial \mathbf{x}_t(b)} = \frac{1}{c}\left(\frac{\partial \mathbf{t}}{\partial \mathbf{x}_t(b)}\right)^T \frac{(\mathbf{s}_k - \mathbf{m}_j)}{||\mathbf{s}_k - \mathbf{m}_j||}$$

$$b = 1, \ldots, D_t$$

$$\frac{\partial r_{i,j,l}}{\partial \xi} = \frac{\partial r_{k,j,l}}{\partial \xi} = 1 \,. \tag{40}$$

Hence, derivation of the CRLB is summarized as follows. Define the calibration scenario, that is, define $c$, $\mathbf{s}_i$, $\mathbf{m}_j$, $\mathbf{s}_k^{\circ}$, $\mathbf{m}_l^{\circ}$, $\mathbf{x}_R$, $\mathbf{x}_t$ and $\xi$. Compute $\mathbf{s}_k$ and $\mathbf{m}_l$, using the rigid motion in (1). Let $\mathbf{x}$ group $\mathbf{x}_R$, $\mathbf{x}_t$, and $\xi$. Compute all linearly independent $r_{i,j,l}$, using (9), and group them into $\mathbf{r}_m$. Compute all linearly independent $r_{k,j,l}$, using (10), and group them into $\mathbf{r}_s$. Assuming knowledge of second order TOA noise statistics, compute $\mathbf{Q}_m$ and $\mathbf{Q}_s$, using (14). Compute the element-wise partials in (40). Substitute the values of the element-wise partials in (40) into the corresponding entries of the vectorized partials in (39). The CRLB is then given by negation followed by inverse of (39). The final answer is a $D \times D$ matrix whose diagonal elements represent the lower bound on the variance, i.e., mean squared error (MSE), of an unbiased estimate of $\mathbf{x}$.

## VI. SIMULATION EXPERIMENTS

Three experiments were conducted to analyze the performance of the proposed estimator. In all experiments, the speed of sound $c$ was fixed at 343 m/s. As illustrated in Fig. 4,

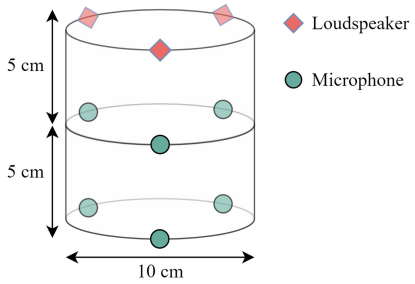


**FIGURE 4.** PA geometry for $I = 3$.

PA was defined as a cylindrical array with radius 0.05 m and height 0.1 m. The coordinates of PA's loudspeakers were generated by

$$\mathbf{s}_i = \begin{bmatrix} 0.05\cos 2\pi(i-1)/I \\ 0.05\sin 2\pi(i-1)/I \\ 0.1 \end{bmatrix} \quad (\text{m}), \tag{41}$$

where we let $I = 3$ for all experiments except experiment 2. The coordinates of PA's microphones, on the other hand, were fixed throughout all experiments. They are given by

$$\mathbf{m}_j = \begin{bmatrix} 0.05\cos 2\pi(j-1)/3 \\ 0.05\sin 2\pi(j-1)/3 \\ 0.05\lfloor (j-1)/3 \rfloor \end{bmatrix} \quad (\text{m}), \tag{42}$$

where we let $J = 6$. For the sake of simplicity, SA's geometry was set equal to that of PA. The degrees of freedom of SA were defined using the parametrization example in Section II-A, that is, $\mathbf{R}$ was parametrized by $\alpha$, $\beta$ and $\gamma$; and $\mathbf{t}$ was parametrized by $\rho$, $\theta$ and $\phi$. In all experiments, the range $\rho$ was constrained within [0,2] (m); and the synchronization offset was constrained within $[-1, 1]$ (s), which is easily achievable in practice. Finally, the five parameters of PSO were set using the values introduced in Section IV-E.

In the first experiment, we compare the performance of the proposed estimator with the CRLB for varying TOA noise magnitude. The position and synchronization mismatch of SA with respect to PA were fixed by letting $\alpha = \beta = \gamma = \theta = \phi = \pi/6$, $\rho = 1$ m and $\xi = 0.1$ s. The TDOA measurements were computed using (12) and (13), where the TOA noise on the right hand side of (13) was simulated using white gaussian noise (WGN) with standard deviation $\sigma$. Consequently, we let

$$\mathbb{E}\left[\Delta\tau_{p,q}\tau_{p,'q'}\right] = \begin{cases} \sigma^2, & \text{if } p = p' \text{ and } q = q' \\ 0, & \text{otherwise} \end{cases}, \tag{43}$$

where $p' = 1, \ldots, I + K$ and $q' = 1, \ldots, J + L$. Considering that the proposed method is being compared to the CRLB, the following performance metrics, based on root mean square error (RMSE), are used

$$\begin{aligned} \text{RMSE}_{\text{ori}} = \mathbb{E}\left[(\hat{\alpha}-\alpha)^2\right]^{1/2} + \mathbb{E}[(\hat{\beta}-\beta)^2]^{1/2} \\ + \mathbb{E}\left[(\hat{\gamma}-\gamma)^2\right]^{1/2} \end{aligned} \tag{44}$$

$$\text{RMSE}_{\text{rng}} = \mathbb{E}\left[(\hat{\rho}-\rho)^2\right]^{1/2} \tag{45}$$

$$\text{RMSE}_{\text{dir}} = \mathbb{E}\left[\left(\hat{\theta}-\theta\right)^2\right]^{1/2} + \mathbb{E}\left[\left(\hat{\phi}-\phi\right)^2\right]^{1/2} \tag{46}$$

$$\text{RMSE}_{\text{syn}} = \mathbb{E}\left[\left(\hat{\xi}-\xi\right)^2\right]^{1/2}, \tag{47}$$

which represent orientation, ranging, direction, and synchronization RMSEs, respectively. In the case of CRLB, the expectation terms in (44-47) were given by the corresponding diagonal elements of the CRLB matrix. In the case of PSO, the expectation terms in (44-47) were estimated using the mean over a set of 100 MC simulations. Each MC simulation consisted in rerunning PSO with a different random seed for: simulation of additive TOA noise $\Delta\tau_{p,q}$; stochastic initialization of swarm positions and velocities in (29); and stochastic update of swarm velocities in (22). Fig. 5 shows that PSO attains the CRLB in all metrics except, perhaps, orientation RMSE at higher values of $\sigma$. However, results show that when $\sigma$ is high, reliable estimation of orientation is not possible.

In the second experiment, we compare the performance of the proposed estimator with the CRLB as the number of loudspeakers in PA and SA vary jointly. The experimental setup was the same as in the first experiment, except for $\sigma$, which was fixed by letting $10\log_{10}(c^{-1}\sigma) = -30$. As expected, Fig. 6 shows that performance improves as the number of loudspeakers increases. Note that the CRLB for $I = K = 1$ is not included due to the FIM being singular. A singular FIM implies that an unbiased estimator does not exist [30], which is consistent with the ambiguity argument in Section III-B. Again, results show that PSO attains the CRLB in most cases.

Let us now define three practical calibration scenarios which consist in changing the number of degrees of freedom of SA. We will refer to these scenarios as 7D, 5D and 4D. 7D implies that the calibration algorithm is required to estimate seven parameters, that is, all six degrees of freedom of SA, which, in this context, are given by $\alpha$, $\beta$, $\gamma$, $\rho$, $\theta$ and $\phi$; plus the synchronization offset $\xi$. 5D, on the other hand, consists in estimating five parameters by assuming $\beta = \gamma = 0$. Similarly, 4D consists in estimating only four parameters by assuming $\beta = \gamma = \phi = 0$. Note that although 7D is more general, 5D and 4D are very practical, since in many cases the gravitational orientation of devices may be known a priori (5D) and the positioning of devices may be further constrained to a flat surface (4D).

In the third and final experiment, we evaluate the 4D, 5D and 7D calibration performance of PSO when TDOA measurements are estimated using acoustic signals in a simulated $5 \times 5 \times 3$ (m) room and the range $\rho$ varies in [0.2,2] (m). The room was simulated using the image-source method [31] with fixed reverberation time RT60 = 0.5 s. PA was placed in the center of the room. The position and synchronization mismatch of SA with respect to PA were generated randomly for each calibration scenario. Special care had to be taken for the 5D and 7D calibration scenarios to restrict the position of SA within the bounds of the room. The calibration procedure was conducted by exciting each loudspeaker independently with a WGN signal of 1 s length sampled at 48 kHz. The TOA at each microphone was then estimated by applying GCC-PHAT [20]

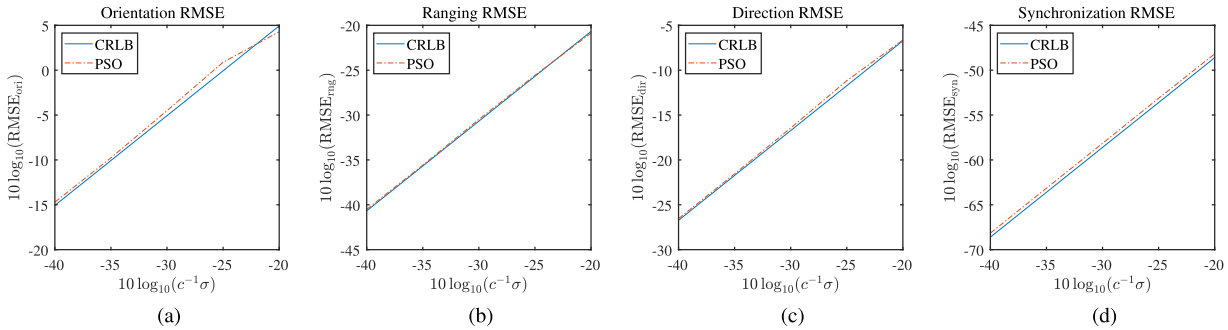


**FIGURE 5.** Experiment 1. Performance of PSO vs. CRLB when varying the magnitude of additive TOA noise $\Delta\tau_{p,q}$.

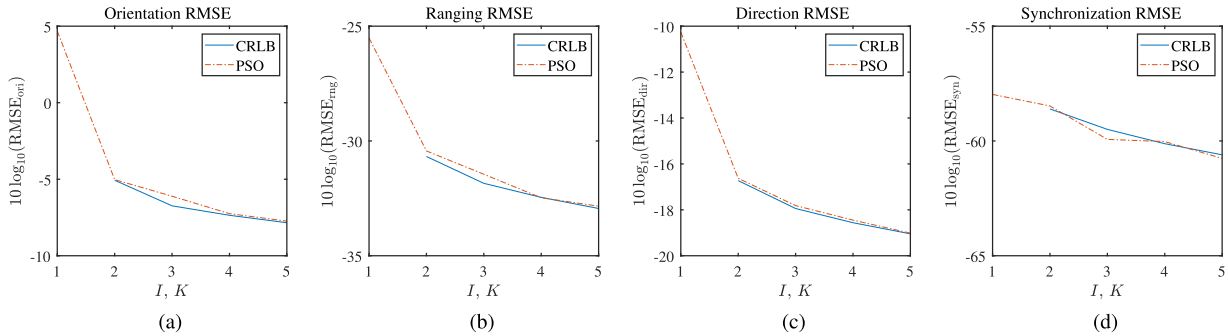


**FIGURE 6.** Experiment 2. Performance of PSO vs. CRLB as the numbers of loudspeakers in PA and SA, $I$, $K$, respectively, vary jointly in [1,5].

to the captured and reference signals. Additionally, quadratic interpolation [32] was used for improved TOA resolution. The second order statistics of TOA noise were approximated using (43), where we let $\sigma^2 = 1$. Two performance metrics are used in this experiment, one is the synchronization RMSE in (47) and the other is the localization RMSE, defined by

$$\mathrm{RMSE}_{\mathrm{loc}} = \mathbb{E}\left[\frac{1}{K+L}\left(\sum_{k=1}^{K}||\hat{\mathbf{s}}_k - \mathbf{s}_k||^2 + \sum_{l=1}^{L}||\hat{\mathbf{m}}_l - \mathbf{m}_l||^2\right)\right]^{1/2}. \tag{48}$$

The expectation terms of (47) and (48) were estimated using the mean over a set of 100 MC simulations. Each MC simulation consisted in rerunning PSO with a different random seed for: simulating the position and synchronization mismatch of SA with respect to PA; stochastic initialization of swarm positions and velocities; and stochastic update of swarm velocities. The results in Fig. 7 show that overall RMSE is exceptionally low at close field but deteriorates with increasing $\rho$ up to a point where it diverges completely due to unreliable TOA estimates caused by reverberation. As expected, results also show that the estimator performance tends to improve when the number of degrees of freedom is reduced. Furthermore, it is shown in Fig. 8 that PSO requires a surprisingly low number of iterations to converge in the 4D and 5D scenarios, while it has difficulties converging in the 7D scenario.

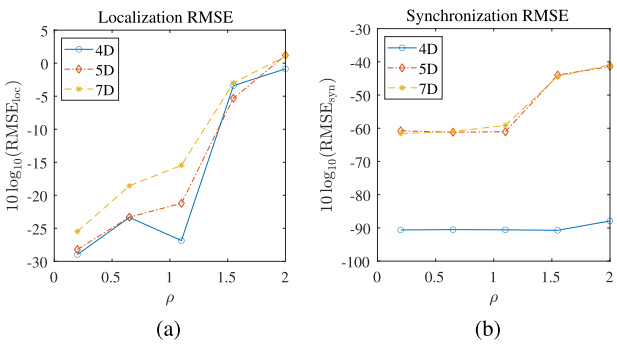


**FIGURE 7.** Experiment 3. Performance of PSO in 4D, 5D, and 7D calibration scenarios as range $\rho$ varies in [0.2,2] (m).

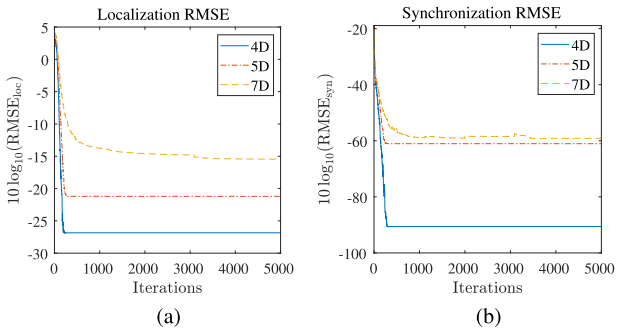


**FIGURE 8.** Experiment 3. $\rho = 1.1$ m. Convergence rate of PSO in 4D, 5D, and 7D calibration scenarios.

## VII. CONCLUSION

A methodology for the joint calibration and synchronization of two arrays of microphones and loudspeakers was presented. The problem is modeled as estimation of the rigid motion of SA with respect to PA, as well as estimation of the synchronization offset between the two. It is assumed that intra-array geometry is known, and intra-array audio input channels are synchronized, assumptions which are generally true in practice. The method consists in using signals emitted by the loudspeakers of PA and SA to compute a set of TOA estimates, which, through a simple transformation, are converted into a set of linearly independent TDOAs. The TDOAs are modeled by a system of nonlinear equations in the unknown parameters of interest, whose ML solution is found by means of optimization of a NWLS problem using a parallelizable variant of constricted PSO with RCT. Additionally, we derived the CRLB for the problem and benchmarked it against PSO in a series of MC simulations. Overall results showed that PSO tends to attain the CRLB. Furthermore, acoustic simulation results showed that the performance of PSO, including convergence rate, can be further improved if the number of degrees of freedom is reduced. Although the presented methodology assumes a network of only two arrays, it can be easily applied to bigger networks. For instance, a WASN could be scaled iteratively as new arrays join the network, or, assuming unambiguous solutions, multiple SAs could be calibrated with respect to a designated PA in parallel. The proposed method can also be used in other applications where precise localization and/or synchronization between two devices is necessary.

## ACKNOWLEDGMENT

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

## REFERENCES

[1] N. Shankar, G. S. Bhat, and I. M. Panahi, "Real-time dual-channel speech enhancement by VAD assisted MVDR beamformer for hearing aid applications using smartphone," in *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2020, pp. 952–955.

[2] J.-M. Valin, F. Michaud, J. Rouat, and D. Létourneau, "Robust sound source localization using a microphone array on a mobile robot," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2003, pp. 1228–1233.

[3] S. Tokgöz, A. Kovalyov, and I. Panahi, "Real-time estimation of direction of arrival of speech source using three microphones," in *Proc. IEEE Workshop Signal Process. Syst.*, 2020, pp. 1–5.

[4] F. Grondin and F. Michaud, "Lightweight and optimized sound source localization and tracking methods for open and closed microphone array configurations," *Robot. Auton. Syst.*, vol. 113, pp. 63–80, 2019.

[5] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," *Wireless Commun. Mobile Comput.*, vol. 2017, pp. 1–2, 2017.

[6] X. Anguera, C. Wooters, and J. Hernando, "Acoustic beamforming for speaker diarization of meetings," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 7, pp. 2011–2022, Sep. 2007.

[7] Y. Sun, K. C. Ho, and Q. Wan, "Solution and analysis of TDOA localization of a near or distant source in closed form," *IEEE Trans. Signal Process.*, vol. 67, no. 2, pp. 320–335, Jan. 2019.

[8] M. Chen, Z. Liu, L.-W. He, P. Chou, and Z. Zhang, "Energy-based position estimation of microphones and speakers for ad hoc microphone arrays," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2007, pp. 22–25.

[9] I. McCowan, M. Lincoln, and I. Himawan, "Microphone array shape calibration in diffuse noise fields," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 3, pp. 666–670, Mar. 2008.

[10] T.-K. Hon, L. Wang, J. D. Reiss, and A. Cavallaro, "Audio fingerprinting for multi-device self-localization," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 10, pp. 1623–1636, Oct. 2015.

[11] M. Cobos, J. J. Perez-Solano, O. Belmonte, G. Ramos, and A. M. Torres, "Simultaneous ranging and self-positioning in unsynchronized wireless acoustic sensor networks," *IEEE Trans. Signal Process.*, vol. 64, no. 22, pp. 5993–6004, Nov. 2016.

[12] V. C. Raykar, I. V. Kozintsev, and R. Lienhart, "Position calibration of microphones and loudspeakers in distributed computing platforms," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 70–83, Jan. 2005.

[13] P. Pertilä, M. Mieskolainen, and M. S. Hämäläinen, "Closed-form self-localization of asynchronous microphone arrays," in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays*, 2011, pp. 139–144.

[14] N. Saeed, H. Nam, M. I. U. Haq, and D. B. M. Saqib, "A survey on multidimensional scaling," *ACM Comput. Surv.*, vol. 51, no. 3, pp. 1–25, 2018.

[15] F. Jacob, J. Schmalenstroeer, and R. Haeb-Umbach, "DOA-based microphone array postion self-calibration using circular statistics," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2013, pp. 116–120.

[16] A. Plinge, G. A. Fink, and S. Gannot, "Passive online geometry calibration of acoustic sensor networks," *IEEE Signal Process. Lett.*, vol. 24, no. 3, pp. 324–328, Mar. 2017.

[17] S. Woźniak and K. Kowalczyk, "Passive joint localization and synchronization of distributed microphone arrays," *IEEE Signal Process. Lett.*, vol. 26, no. 2, pp. 292–296, Feb. 2019.

[18] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "BeepBeep: A high accuracy acoustic ranging system using cots mobile devices," in *Proc. 5th Int. Conf. Embedded Netw. Sensor Syst.*, 2007, pp. 1–14.

[19] F. Scheck, *Mechanics: From Newton's Laws to Deterministic Chaos*. New York, NY, USA: Springer Science & Business Media, 2010, pp. 187–189.

[20] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[21] D. Budnikov, I. Chikalov, I. Kozintsev, and R. Lienhart, "Distributed array of synchronized sensors and actuators," in *Proc. 12th Eur. Signal Process. Conf.*, 2004, pp. 2243–2246.

[22] S. Sur, T. Wei, and X. Zhang, "Autodirective audio capturing through a synchronized smartphone array," in *Proc. 12th Annu. Int. Conf. Mobile Syst., Appl., Serv.*, 2014, pp. 28–41.

[23] Eberhart and Y. Shi, "Particle swarm optimization: developments, applications and resources," in *Proc. Congr. Evol. Comput.*, 2001, pp. 81–86.

[24] T. Li, G. Shao, W. Zuo, and S. Huang, "Genetic algorithm for building optimization: State-of-the-art survey," in *Proc. 9th Int. Conf. Mach. Learn. Comput.*, 2017, pp. 205–210.

[25] K. R. Opara and J. Arabas, "Differential evolution: A survey of theoretical analyses," *Swarm Evol. Comput.*, vol. 44, pp. 546–558, 2019.

[26] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Eng. Faculty, Comput. Eng. Dept., Erciyes Univ., Kayseri, Turkey, Tech. Rep. TR06, 2005.

[27] D. Bratton and J. Kennedy, "Defining a standard for particle swarm optimization," in *Proc. IEEE Swarm Intell. Symp.*, 2007, pp. 120–127.

[28] M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space," *IEEE Trans. Evol. Comput.*, vol. 6, no. 1, pp. 58–73, Feb. 2002.

[29] S. M. Kay, *Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice Hall PTR, 1998, pp. 27–63.

[30] P. Stoica and B. C. Ng, "On the Cramér-Rao bound under parametric constraints," *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, Jul. 1998.

[31] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoustical Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[32] I. J. Tashev, *Sound Capture and Processing: Practical Approaches*. New York, NY, USA: Wiley, 2009, pp. 300–301.