

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

This paper was downloaded from TechRxiv (<https://www.techrxiv.org>).

LICENSE

CC BY 4.0

SUBMISSION DATE / POSTED DATE

25-06-2021 / 01-07-2021

CITATION

Kovalyov, Anton (2021): Joint Calibration and Synchronization of Two Arrays of Microphones and Loudspeakers using Particle Swarm Optimization. TechRxiv. Preprint.  
<https://doi.org/10.36227/techrxiv.14849742.v1>

DOI

[10.36227/techrxiv.14849742.v1](https://doi.org/10.36227/techrxiv.14849742.v1)

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

Anton Kovalyov, Kashyap Patel, and Issa Panahi

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

**M**ICROPHONE 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]. 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. Wozniak and Kowalczyk [15] assumed sources at far field and individual nodes equipped with a synchronized microphone array capable of reliable DOA estimation. Their method uses DOAs observed at individual nodes and TDOAs observed between the microphones of different nodes to jointly estimate the WASN geometry, the relative positions of the sources with respect to the WASN, and the synchronization offsets between nodes.

On the other hand, work on active calibration includes [11]–[13], [16]. Peng et al. [16] proposed a system called “Beep-Beep” 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

The authors are with the Department of Electrical and Computer Engineering, University of Texas at Dallas, Richardson, TX-75080, USA.

Corresponding authors:

Anton Kovalyov; e-mail: anton.kovalyov@utdallas.edu

Kashyap Patel; e-mail: patelkashyap@utdallas.edu

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. 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 microphones and loudspeakers 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 simulation experiments. Results show that the proposed method attains the CRLB except for a few boundary 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 benchmarked against the CRLB 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.  $(\cdot)$  denotes a known estimate.  $\mathbb{E}[\cdot]$  denotes expectation.  $(\hat{\cdot})$  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  $i = 1, \dots, I$  and  $j = 1, \dots, J$ . 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  $k = I + 1, \dots, I + K$  and  $l = J + 1, \dots, J + L$ . Consequently, the position of SA with respect to PA is modeled by

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

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.

## 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  $p = 1, \dots, I+K$  and  $q = 1, \dots, J+L$ . Assuming direct path between loudspeakers and microphones, corresponding noise-free TOAs are given by

$$\tau_{p,j} = c^{-1} \|\mathbf{s}_p - \mathbf{m}_j\| + \tau_p + \delta_m \quad (2)$$

$$\tau_{p,l} = c^{-1} \|\mathbf{s}_p - \mathbf{m}_l\| + \tau_p + \delta_s, \quad (3)$$

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 [17]. 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 (2) and (3). Here, we assume direct knowledge of noisy TOA estimates modeled by

$$\tilde{\tau}_{p,q} = \tau_{p,q} + \Delta\tau_{p,q}. \quad (4)$$

## C. TDOA and Synchronization Offset

The parameters  $\tau_p$ ,  $\delta_m$  and  $\delta_s$  are of no interest to us. Subtracting (2) and (3) results in the following system of nonlinear TDOA equations

$$r_{p,j,l} = c^{-1} (\|\mathbf{s}_p - \mathbf{m}_j\| - \|\mathbf{s}_p - \mathbf{m}_l\|) + \xi, \quad (5)$$

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<sup>1</sup> 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 related to the rotation matrix  $\mathbf{R}$

<sup>1</sup>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 [18], [19].

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  $i = 1, \dots, I$ ,  $j = 1, \dots, J$ ,  $k = I+1, \dots, I+K$ ,  $l = J+1, \dots, J+L$ ,  $p = 1, \dots, I+K$ , and  $q = 1, \dots, J+L$ .

## III. PROBLEM SOLUTION

Since the transformation of SA's LCS in (1) does not change relative distances between points, the TDOA representation in (5) can be conveniently rewritten as

$$r_{i,j,l} = c^{-1} (\|\mathbf{s}_i - \mathbf{m}_j\| - \|\mathbf{s}_i - \mathbf{R}\mathbf{m}_l^\circ - \mathbf{t}\|) + \xi \quad (6)$$

$$r_{k,j,l} = c^{-1} (\|\mathbf{R}\mathbf{s}_k^\circ + \mathbf{t} - \mathbf{m}_j\| - \|\mathbf{s}_k^\circ - \mathbf{m}_l^\circ\|) + \xi, \quad (7)$$

where the only unknowns on the right-hand side of (6) and (7) are the parameters of interest. An estimate of true TDOA is given by

$$\begin{aligned} \tilde{r}_{p,j,l} &= \tilde{\tau}_{p,j} - \tilde{\tau}_{p,l} \\ &= r_{p,j,l} + \Delta\tau_{p,j} - \Delta\tau_{p,l} \\ &= r_{p,j,l} + \Delta r_{p,j,l}. \end{aligned} \quad (8)$$

Let

$$\mathbf{x} = [\mathbf{x}_R^T \quad \mathbf{x}_t^T \quad \xi]^T \quad (9)$$

group the unknown parameters, where  $\mathbf{x}_R$  and  $\mathbf{x}_t$  group parameters associated to  $\mathbf{R}$  and  $\mathbf{t}$ , respectively. Let  $\mathbf{r}(\mathbf{x})$  group noise-free linearly independent TDOAs constructed as a function of  $\mathbf{x}$  using (6) and (7). Let  $\tilde{\mathbf{r}}$  group the corresponding linearly independent TDOA measurements defined in (8). Assuming additive zero-mean Gaussian noise, the maximum likelihood (ML) estimate of  $\mathbf{x}$  becomes a nonlinear weighted least squares (NWLS) problem given by

$$\hat{\mathbf{x}} = \arg \min_x \mathcal{L}(\mathbf{x}) \quad (10)$$

$$\mathcal{L}(\mathbf{x}) = (\tilde{\mathbf{r}} - \mathbf{r}(\mathbf{x}))^T \mathbf{Q}^{-1} (\tilde{\mathbf{r}} - \mathbf{r}(\mathbf{x})), \quad (11)$$

where  $\mathbf{Q} = \mathbb{E} [\Delta\mathbf{r}\Delta\mathbf{r}^T]$  is the covariance matrix of the noise in  $\tilde{\mathbf{r}}$ . Knowing the second order statistics of noise in TOA measurements,  $\mathbf{Q}$  can be easily found by noting that

$$\begin{aligned} &\mathbb{E} [\Delta r_{p,j,l} \Delta r_{p',j',l'}] = \\ &\mathbb{E} [\Delta\tau_{p,j} \Delta\tau_{p',j'}] + \mathbb{E} [\Delta\tau_{p,l} \Delta\tau_{p',l'}] - \\ &\mathbb{E} [\Delta\tau_{p,j} \Delta\tau_{p',l'}] - \mathbb{E} [\Delta\tau_{p,l} \Delta\tau_{p',j'}], \end{aligned} \quad (12)$$

where we used the relationship  $\Delta r_{p,j,l} = \Delta\tau_{p,j} - \Delta\tau_{p,l}$  defined in (8).

Note that the total number of TDOAs is  $(I+K)JL$  but 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 is given by

$$\tilde{R}_p = \{\tilde{r}_{p,1,J+1}, \dots, \tilde{r}_{p,1,J+L}, \tilde{r}_{p,2,J+1}, \dots, \tilde{r}_{p,J,J+1}\}. \quad (13)$$

Consequently, letting  $D = D_R + D_t + 1$  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  $(I+K)(J+L-1) \geq D$  should be satisfied for the estimator in (10) to work.

#### IV. NUMERICAL OPTIMIZATION METHOD

Minimization of (11) 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, LMA 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], [16]. Here we prefer not to make additional assumptions and propose solving the nonlinear problem in (10) directly using PSO.

PSO is a well-known population-based metaheuristic (PM) applied in a wide variety of fields [20]. 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) [21], Differential Evolution (DE) [22] and Artificial Bee Colony (ABC) [23]. 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 (10). 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 [24] with a slight modification to allow parallelization of computations.

##### A. Solution Search with PSO

PSO is applied to solve the optimization problem in (10) as follows. Let  $N_i$  be the number of iterations. Let  $N_p$  be the number of particles. A particle  $n$ , for  $n = 0, \dots, N_p - 1$ , is represented by four vectors: its position  $\mathbf{x}_n$ , i.e., candidate solution to (10); 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

$$\mathbf{v}_n = \chi \left[ \mathbf{v}_n + c_1 \mathbf{a}_{1,n} \odot (\mathbf{x}_n - \mathbf{p}_n) + c_2 \mathbf{a}_{2,n} \odot (\mathbf{x}_n - \mathbf{g}_n) \right] \quad (14)$$

$$\mathbf{x}_n = \mathbf{x}_n + \mathbf{v}_n \quad (15)$$

$$\mathbf{p}_n = \arg \min_{\mathbf{x}} \{ \mathcal{L}(\mathbf{x}) \mid \mathbf{x} \in \{ \mathbf{p}_n, \mathbf{x}_n \} \} \quad (16)$$

$$\mathbf{g}_n = \arg \min_{\mathbf{x}} \{ \mathcal{L}(\mathbf{x}) \mid \mathbf{x} \in G_n \} , \quad (17)$$

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 (10) is given by

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \{ \mathcal{L}(\mathbf{x}) \mid \mathbf{x} \in \{ \mathbf{p}_0, \dots, \mathbf{p}_{N_p-1} \} \} . \quad (18)$$

##### 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 (17). 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. 1.a), and as such

$$G_n = \{ \mathbf{p}_0, \dots, \mathbf{p}_{N_p-1} \} . \quad (19)$$

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. 1.b). It then follows that

$$G_n = \{ \mathbf{p}_{\text{mod}(n-1, N_p)}, \mathbf{p}_n, \mathbf{p}_{\text{mod}(n+1, N_p)} \} . \quad (20)$$

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 (10), 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 (10) 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 a uniform distribution. Let  $\mathbf{b}_l$  and  $\mathbf{b}_u$  be vectors representing the

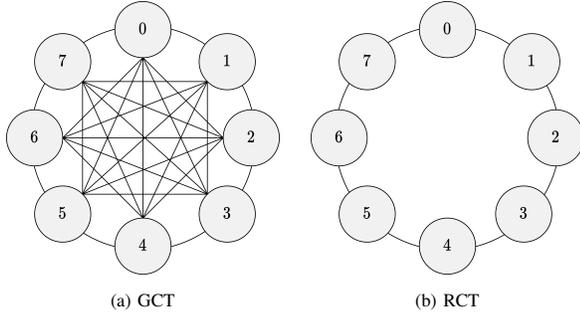


Fig. 1. Particle swarm communication topologies.

lower and upper boundaries of the search space, respectively. Initially, we let

$$\begin{aligned} \mathbf{x}_n(d) &\sim \mathcal{U}(\mathbf{b}_l(d), \mathbf{b}_u(d)) \\ \mathbf{v}_n(d) &\sim \mathcal{U}(-\mathbf{v}_{max}(d), \mathbf{v}_{max}(d)) \end{aligned} \quad (21)$$

$$d = 1, \dots, D,$$

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

#### D. Particles Travelling Outside the Search Space

The swarm update rule in (15) 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 [24], 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 (14) 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 (11) 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  $N_{FE} = 10^5$ , a good choice is to let  $N_p = 100$  and consequently  $N_i = 999$ . 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 [24]. Proof of convergence of PSO can be found in [25].

#### F. Simultaneous vs. Sequential Swarm Update

The swarm update rules in (14)-(17) can be interpreted in two ways: the update is sequential, that is, particles are updated

sequentially, one at a time, using (14)-(17); the update is simultaneous, that is, all particles are updated simultaneously using (14), followed by (15) 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 [24] used as reference in this paper. We did not find noticeable difference in estimation performance when applying either approach to solve (10). Hence, due to its considerable computational speedup, in this work we use simultaneous swarm update.

#### G. Algorithm Summary

The proposed algorithm for solving (10), 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 (21).
- Step 4: Initialize all  $\mathbf{p}_n$  as  $\mathbf{x}_n$ .
- Step 5: Compute FEs for all  $\mathbf{x}_n$ , using (11).
- Step 6: Initialize all  $\mathbf{g}_n$ , using (17) and (19).
- Step 7: Increment  $t$ .
- Step 8: Update all  $\mathbf{v}_n$ , using (14).
- Step 9: Update all  $\mathbf{x}_n$ , using (15).
- Step 10: Compute FEs for all  $\mathbf{x}_n$ , using (11).
- Step 11: Update all  $\mathbf{p}_n$ , using (16).
- Step 12: Update all  $\mathbf{g}_n$ , using (17) and (19).
- Step 13: If  $t < N_i$ , return to Step 7, otherwise, give final solution, using (18).

#### V. CRLB

The CRLB places a lower bound on the variance of an unbiased estimator [26]. 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. The Fisher information matrix is found by

$$\mathcal{I}(\mathbf{x}) = -\mathbb{E} \left[ \frac{\partial^2 \ln f(\tilde{\mathbf{r}}|\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T} \right], \quad (22)$$

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 (9) and  $\tilde{\mathbf{r}}$  groups the TDOA measurements in (8) as follows

$$\tilde{\mathbf{r}} = \begin{bmatrix} \tilde{\mathbf{r}}_m^T & \tilde{\mathbf{r}}_s^T \end{bmatrix}^T, \quad (23)$$

where  $\tilde{\mathbf{r}}_m$  groups all linearly independent  $\tilde{r}_{i,j,l}$  and  $\tilde{\mathbf{r}}_s$  groups all linearly independent  $\tilde{r}_{k,j,l}$ . 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  $\mathcal{N}(\mathbf{r}_s, \mathbf{Q}_s)$ , respectively, where  $\mathbf{r}_m$  groups all linearly independent  $r_{i,j,l}$ , modeled by (6),  $\mathbf{r}_s$  groups all linearly independent  $r_{k,j,l}$ , modeled by (7), and  $\mathbf{Q}_m = \mathbb{E}[\Delta\mathbf{r}_m\Delta\mathbf{r}_m^T]$  and  $\mathbf{Q}_s = \mathbb{E}[\Delta\mathbf{r}_s\Delta\mathbf{r}_s^T]$  are the corresponding noise covariance matrices, whose elements are computed using the relationship of TOA and TDOA noise second order statistics in (12). It follows that the PDF of  $\tilde{\mathbf{r}}$  conditioned on  $\mathbf{x}$  is given by

$$f(\tilde{\mathbf{r}}|\mathbf{x}) = f(\tilde{\mathbf{r}}_m|\mathbf{x})f(\tilde{\mathbf{r}}_s|\mathbf{x}). \quad (24)$$

Taking the natural logarithm of both sides of (24) and expanding, we get

$$\ln f(\tilde{\mathbf{r}}|\mathbf{x}) = -\frac{1}{2}[(\tilde{\mathbf{r}}_m - \mathbf{r}_m)^T \mathbf{Q}_m^{-1}(\tilde{\mathbf{r}}_m - \mathbf{r}_m) + (\tilde{\mathbf{r}}_s - \mathbf{r}_s)^T \mathbf{Q}_s^{-1}(\tilde{\mathbf{r}}_s - \mathbf{r}_s)] + C, \quad (25)$$

where  $C$  is some constant. The expectation of the double partial in (22) is then found to be

$$\mathbb{E} \left[ \frac{\partial^2 \ln f(\tilde{\mathbf{r}}|\mathbf{x})}{\partial \mathbf{x} \partial \mathbf{x}^T} \right] = - \left[ \left( \frac{\partial \mathbf{r}_m}{\partial \mathbf{x}} \right)^T \mathbf{Q}_m^{-1} \left( \frac{\partial \mathbf{r}_m}{\partial \mathbf{x}} \right) + \left( \frac{\partial \mathbf{r}_s}{\partial \mathbf{x}} \right)^T \mathbf{Q}_s^{-1} \left( \frac{\partial \mathbf{r}_s}{\partial \mathbf{x}} \right) \right], \quad (26)$$

where computation of the right-hand side of (26) is straightforward given the following element-wise partials

$$\begin{aligned} \frac{\partial r_{i,j,l}}{\partial \mathbf{x}_R(a)} &= \frac{1}{c} \left( \frac{\partial \mathbf{R}}{\partial \mathbf{x}_R(a)} \mathbf{m}_i^\circ \right)^T \frac{(\mathbf{s}_i - \mathbf{m}_i)}{\|\mathbf{s}_i - \mathbf{m}_i\|} \\ \frac{\partial r_{k,j,l}}{\partial \mathbf{x}_R(a)} &= \frac{1}{c} \left( \frac{\partial \mathbf{R}}{\partial \mathbf{x}_R(a)} \mathbf{s}_k^\circ \right)^T \frac{(\mathbf{s}_k - \mathbf{m}_j)}{\|\mathbf{s}_k - \mathbf{m}_j\|} \\ a &= 1, \dots, D_R \\ \\ \frac{\partial r_{i,j,l}}{\partial \mathbf{x}_t(b)} &= \frac{1}{c} \left( \frac{\partial \mathbf{t}}{\partial \mathbf{x}_t(b)} \right)^T \frac{(\mathbf{s}_i - \mathbf{m}_i)}{\|\mathbf{s}_i - \mathbf{m}_i\|} \\ \frac{\partial r_{k,j,l}}{\partial \mathbf{x}_t(b)} &= \frac{1}{c} \left( \frac{\partial \mathbf{t}}{\partial \mathbf{x}_t(b)} \right)^T \frac{(\mathbf{s}_k - \mathbf{m}_j)}{\|\mathbf{s}_k - \mathbf{m}_j\|} \\ b &= 1, \dots, D_t \\ \\ \frac{\partial r_{i,j,l}}{\partial \xi} &= \frac{\partial r_{k,j,l}}{\partial \xi} = 1. \end{aligned} \quad (27)$$

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 (6), and group them into  $\mathbf{r}_m$ . Compute all linearly independent  $r_{k,j,l}$ , using (7), and group them into  $\mathbf{r}_s$ . Assuming knowledge of second order TOA noise statistics, compute  $\mathbf{Q}_m$  and  $\mathbf{Q}_s$ , using (12). Compute the element-wise partials in (27). Substitute the values of the element-wise partials in (27) into the corresponding entries of the vectorized partials in (26). The CRLB is then given by negation followed

TABLE I  
PA MICROPHONE COORDINATES

$j$	1	2	3	4	5	6
$x$	-9.29	3.57	-2.16	4.12	-9.08	3.90
$y$	6.98	5.15	3.11	-9.36	-8.06	-3.66
$z$	8.68	4.86	-6.58	-4.46	6.47	9.00

TABLE II  
PA LOUDSPEAKER COORDINATES

$i$	1	2	3	4	5	6
$x$	6.29	8.27	-4.43	9.30	9.14	-7.16
$y$	8.12	2.65	0.94	-6.85	-0.29	-1.56
$z$	-7.46	-8.05	9.15	9.41	6.01	8.31

by inverse of (26). 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

A series of simulation experiments were conducted to compare the performance of the proposed estimator with the CRLB using a randomly generated calibration scenario.

### A. Calibration Scenario

The calibration scenario was defined as follows. Microphone and loudspeaker coordinates were drawn from  $\mathcal{U}(-10, 10)$ . Tables I and II show the microphone and loudspeaker array geometries given to PA. For the sake of simplicity, SA geometry was set equal to that of PA. The position of SA with respect to PA was simulated by applying a 3D rotation and 3D translation to the LCS of SA as in (1). The 3D rotation matrix was defined, using Tait-Bryan representation, as

$$\mathbf{R}(\alpha, \beta, \gamma) = \mathbf{R}_z(\alpha)\mathbf{R}_y(\beta)\mathbf{R}_x(\gamma), \quad (28)$$

where the *yaw* angle  $\alpha$ , the *pitch* angle  $\beta$  and the *roll* angle  $\gamma$  are rotation parameters whose corresponding matrices represent rotation about the  $z$ ,  $y$ , and  $x$  axes, respectively. The 3D translation vector was defined, using spherical representation, as

$$\mathbf{t}(r, \theta, \phi) = [r \cos \theta \cos \phi \quad r \sin \theta \cos \phi \quad r \sin \theta \sin \phi]^T, \quad (29)$$

where  $r$ ,  $\theta$  and  $\phi$  are translation parameters denoting range, azimuth angle and elevation angle, respectively. Unless otherwise specified, in all experiments, we let  $\alpha = \beta = \gamma = \theta = \phi = \pi/6$  and  $r = \xi = 100$ . These seven parameters define the calibration parameters that need to be estimated by the proposed method using a set of noisy TDOA measurements. We will refer to this calibration scenario as 7D.

### B. Noisy TDOA measurements

A set of linearly independent noisy TDOA measurements were simulated using (5), (8) and the linear independence set definition in (13). Additive white Gaussian noise (AWGN) was used to simulate the TOA noise in (8). Since no actual acoustic signals were used in the simulations, for the sake of simplicity, we let  $c = 1$ .

### C. Performance Metrics

Considering that the proposed method is being compared to the CRLB, the performance metrics used are based on MSE. Four metrics were defined: rotation MSE, given by  $\mathbb{E}[(\hat{\alpha} - \alpha)^2] + \mathbb{E}[(\hat{\beta} - \beta)^2] + \mathbb{E}[(\hat{\gamma} - \gamma)^2]$ ; ranging MSE, given by  $\mathbb{E}[(\hat{r} - r)^2]$ ; direction MSE, given by  $\mathbb{E}[(\hat{\theta} - \theta)^2] + \mathbb{E}[(\hat{\phi} - \phi)^2]$ ; and synchronization MSE, given by  $\mathbb{E}[(\hat{\xi} - \xi)^2]$ . In the case of CRLB, the MSE of each parameter was computed as the corresponding diagonal entry of the inverse of the Fisher information matrix derived in Section V. In the case of the proposed estimator based on PSO, the MSE of each parameter was estimated through a set of 1000 MC simulations. Each MC simulation consisted in rerunning PSO with a different random seed for: simulation of the AWGN noise in TOA measurements used in (8); stochastic initialization of the swarm positions and velocities in (21); stochastic update of the swarm velocities in (14).

### D. PSO Implementation Details

The calibration parameters were estimated using the PSO procedure described in Section IV-G. The five parameters of PSO were set according to Section IV-E, that is,  $N_p = 100$ ,  $N_i = 999$ ,  $\chi = 0.72894$  and  $c_1 = c_2 = 2.05$ . PSO search boundaries for  $\alpha$ ,  $\gamma$  and  $\theta$  were set to  $[-\pi, \pi]$ . PSO search boundaries for  $\beta$  and  $\phi$  were set to  $[-\pi/2, \pi/2]$ . PSO search boundaries for  $r$  and  $\xi$  were set to  $[0, 1000]$  and  $[-1000, 1000]$ , respectively. Note that these search boundaries define all possible 3D positions and 3D orientations of SA with respect to PA within a sphere of radius 1000.

### E. Simulation Results

In the first experiment, we compare the performance of the proposed estimator with the CRLB as the noise power in TOA measurements varies in  $[-20\text{dB}, 0\text{dB}]$ . The numbers of microphones and loudspeakers in PA and SA, that is  $I$ ,  $J$ ,  $K$  and  $L$ , were all fixed at 3, meaning that only the first 3 columns of Tables I and II were used to define the array geometries of PA and SA. The results in Fig. 2 show that PSO attains the CRLB in most metrics except for rotation MSE at higher noise powers, where PSO deviates slightly.

In the second experiment, we compare the performance of the proposed estimator with the CRLB as the range  $r$  varies in  $[100, 500]$ . The noise power in TOA measurements was fixed at  $-10\text{dB}$ . The numbers of microphones and loudspeakers of PA and SA were all fixed at 3. The results in Fig. 3 show that the overall performance of PSO relative to CRLB is again satisfactory. The results also show that increased range has no significant impact on all metrics except for rotation MSE. Note that the fact that direction MSE does not change implies that overall localization error increases when range is high.

In the third experiment, we compare the performance of the proposed estimator with the CRLB as the numbers of microphones and loudspeakers of PA and SA vary jointly in  $[2, 6]$ . The noise power in TOA measurements was fixed at  $-10\text{dB}$ . The results in Fig. 4 show that PSO attains the CRLB as long as  $I, J, K, L > 2$ . When  $I, J, K, L = 2$  the overall

gap between PSO and CRLB is significant, although it must be noted that the MSE given by CRLB is also high, especially for the rotation and ranging parameters.

The performance of the proposed estimator can be further improved if the number of degrees of freedom is reduced. In practice, it may be often assumed that the gravitational orientation of PA and SA is fixed, or an estimate is available, e.g., each array may include a tilt sensor. In either case, the gravitational orientation of SA with respect to PA would be known. Including this information in the two LCSs implies that estimation of the pitch angle  $\beta$  and roll angle  $\gamma$  can be skipped and the rotation matrix in (1) can be modeled as simply  $\mathbf{R}_z(\alpha)$ . Hence, the number of degrees of freedom would be reduced to 5. We will refer to this calibration scenario as 5D.

In the fourth and final experiment we use the same problem geometry, that is, the position of SA with respect to PA defined in Section VI-A, and perform the same comparison procedure as in experiment 3, except that the calibration scenario is now 5D. Hence, the rotation matrix in (1) was redefined as  $\mathbf{R}_z(\alpha)$ , and the rotation MSE is now given by  $\mathbb{E}[(\hat{\alpha} - \alpha)^2]$ . Additionally, to ensure that the problem geometry of 5D and 7D calibration scenarios was the same, the LCS of SA given in the 7D scenario was transformed by the rotation  $\mathbf{R}(0, \beta, \gamma)$  in the 5D scenario. The results in Fig. 5 show that although the gap between PSO and CRLB for the boundary case when  $I, J, K, L = 2$  remains noticeable in the 5D scenario when compared to the 7D scenario in Fig. 4, the MSE for the same boundary case is now reduced dramatically in all metrics.

## 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. Results showed that PSO attains the CRLB within  $10^5$  FEs, except for a few boundary cases. As expected, results also showed that the overall MSE for the boundary cases can be further decreased 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

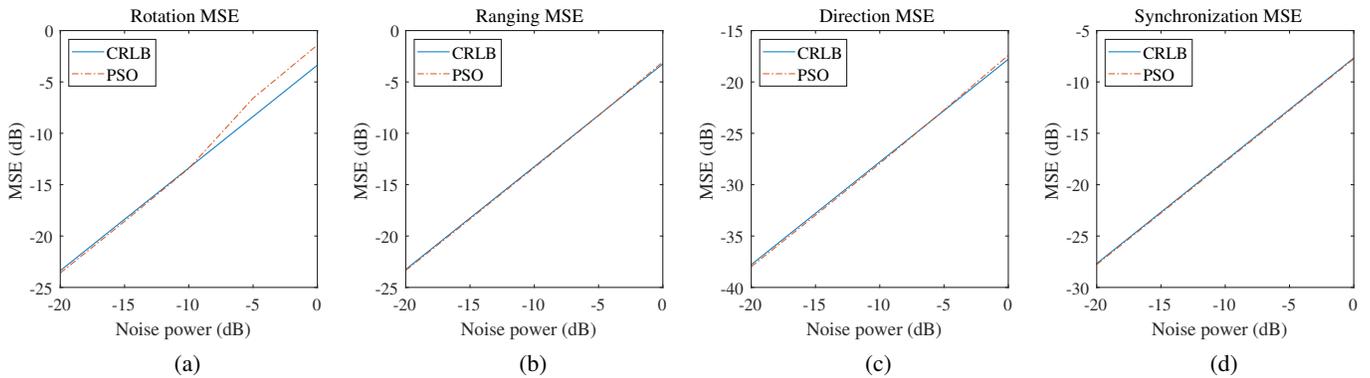


Fig. 2. Experiment 1. 7D calibration scenario. Performance of PSO vs. CRLB as noise power in TOA measurements varies in  $[-20\text{dB}, 0\text{dB}]$ . The numbers of microphones and loudspeakers in PA and SA were all fixed at 3.

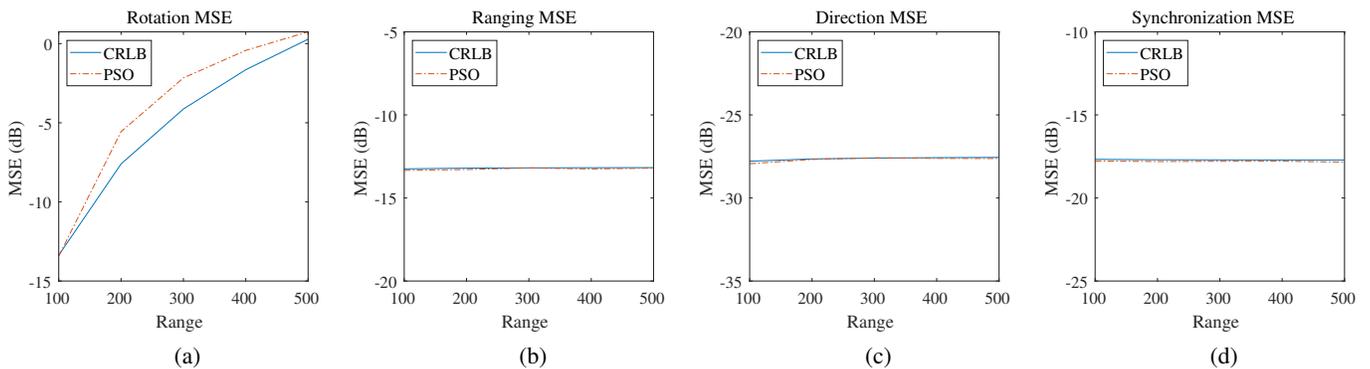


Fig. 3. Experiment 2. 7D calibration scenario. Performance of PSO vs. CRLB as distance of SA with respect to PA varies in  $[100, 500]$ . Noise power in TOA measurements was fixed at  $-10\text{dB}$ . The numbers of microphones and loudspeakers in PA and SA were all fixed at 3.

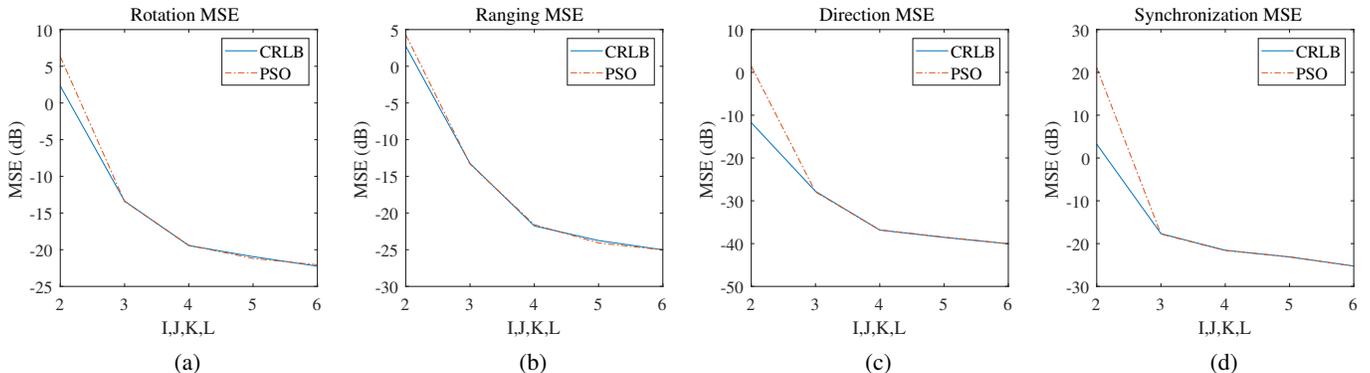


Fig. 4. Experiment 3. 7D calibration scenario. Performance of PSO vs. CRLB as the numbers of microphones and loudspeakers in PA and SA vary jointly in  $[2, 6]$ . Noise power in TOA measurements was fixed at  $-10\text{dB}$ .

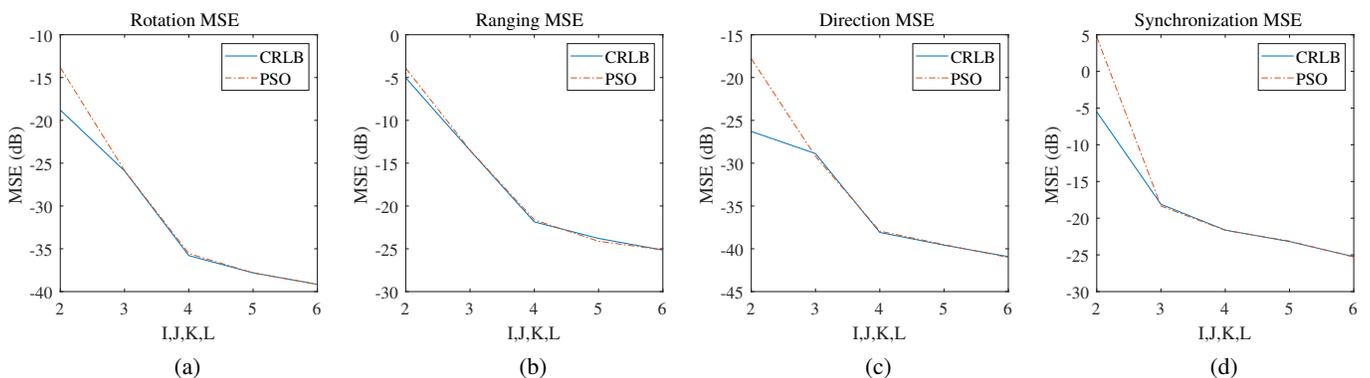


Fig. 5. Experiment 4. 5D calibration scenario. Performance of PSO vs. CRLB as the numbers of microphones and loudspeakers in PA and SA vary jointly in  $[2, 6]$ . Noise power in TOA measurements was fixed at  $-10\text{dB}$ .

where precise localization and/or synchronization between two devices is necessary.

#### ACKNOWLEDGMENT

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. 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 *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 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 *Proceedings 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2003)(Cat. No. 03CH37453)*, vol. 2. IEEE, 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 *2020 IEEE Workshop on Signal Processing Systems (SiPS)*. IEEE, 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," *Robotics and Autonomous Systems*, 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 Communications and Mobile Computing*, vol. 2017, 2017.
- [6] X. Anguera, C. Wooters, and J. Hernando, "Acoustic beamforming for speaker diarization of meetings," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 7, pp. 2011–2022, 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 Transactions on Signal Processing*, vol. 67, no. 2, pp. 320–335, 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 *2007 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*. IEEE, 2007, pp. 22–25.
- [9] 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.
- [10] T.-K. Hon, L. Wang, J. D. Reiss, and A. Cavallaro, "Audio fingerprinting for multi-device self-localization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 10, pp. 1623–1636, 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 Transactions on Signal Processing*, vol. 64, no. 22, pp. 5993–6004, 2016.
- [12] 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, 2004.
- [13] P. Pertilä, M. Mieskolainen, and M. S. Hämäläinen, "Closed-form self-localization of asynchronous microphone arrays," in *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays*. IEEE, 2011, pp. 139–144.
- [14] N. Saeed, H. Nam, M. I. U. Haq, and D. B. Muhammad Saqib, "A survey on multidimensional scaling," *ACM Computing Surveys (CSUR)*, vol. 51, no. 3, pp. 1–25, 2018.
- [15] S. Woźniak and K. Kowalczyk, "Passive joint localization and synchronization of distributed microphone arrays," *IEEE Signal Processing Letters*, vol. 26, no. 2, pp. 292–296, 2018.
- [16] 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 international conference on Embedded networked sensor systems*, 2007, pp. 1–14.
- [17] C. Knapp and G. 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.
- [18] D. Budnikov, I. Chikalov, I. Kozintsev, and R. Lienhart, "Distributed array of synchronized sensors and actuators," in *2004 12th European Signal Processing Conference*, 2004, pp. 2243–2246.
- [19] S. Sur, T. Wei, and X. Zhang, "Autodirective audio capturing through a synchronized smartphone array," in *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*, 2014, pp. 28–41.
- [20] Eberhart and Y. Shi, "Particle swarm optimization: developments, applications and resources," in *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, vol. 1, 2001, pp. 81–86 vol. 1.
- [21] T. Li, G. Shao, W. Zuo, and S. Huang, "Genetic algorithm for building optimization: State-of-the-art survey," in *Proceedings of the 9th international conference on machine learning and computing*, 2017, pp. 205–210.
- [22] K. R. Opara and J. Arabas, "Differential evolution: A survey of the theoretical analyses," *Swarm and evolutionary computation*, vol. 44, pp. 546–558, 2019.
- [23] D. Karaboga, "An idea based on honey bee swarm for numerical optimization," Citeseer, Tech. Rep., 2005.
- [24] D. Bratton and J. Kennedy, "Defining a standard for particle swarm optimization," in *2007 IEEE swarm intelligence symposium*. IEEE, 2007, pp. 120–127.
- [25] M. Clerc and J. Kennedy, "The particle swarm-explosion, stability, and convergence in a multidimensional complex space," *IEEE transactions on Evolutionary Computation*, vol. 6, no. 1, pp. 58–73, 2002.
- [26] S. M. Kay, *Estimation theory*. Prentice Hall PTR, 1998.