ITERATIVE GEOMETRY CALIBRATION FROM DISTANCE ESTIMATES FOR WIRELESS ACOUSTIC SENSOR NETWORKS

Tobias Gburrek, Joerg Schmalenstroeer, Reinhold Haeb-Umbach

Department of Communications Engineering, Paderborn University, Germany
{gburrek, schmalen, haeb}@nt.uni-paderborn.de

ABSTRACT

In this paper we present an approach to geometry calibration in wireless acoustic sensor networks, whose nodes are assumed to be equipped with a compact microphone array. The proposed approach solely works with estimates of the distances between acoustic sources and the nodes that record these sources. It consists of an iterative weighted least squares localization procedure, which is initialized by multidimensional scaling. Alongside the sensor node locations, also the positions of the acoustic sources are estimated. Furthermore, we derive the Cramer-Rao lower bound (CRLB) for source and sensor position estimation, and show by simulation that the estimator is efficient.

Index Terms— Geometry calibration, CRLB, MDS

1. INTRODUCTION

In our scenario, the wireless acoustic sensor network (WASN) hardware consists of sensor nodes, each equipped with a microphone array. The nodes are at random positions in a room, and they are connected via a WiFi network. The task to be solved, called geometry calibration, is to estimate the positions of all nodes from the observed acoustic signals in the room. Knowledge of the sensor network's geometry is required, if the WASN is to be used to localize speakers, or to enable proximity-based recording and handover functionality [1].

Geometry calibration can be solved by optimizing a cost function that describes the difference between an assumed geometry and one computed from measurements [2, 3] or by a maximum likelihood method using a probabilistic model [4,5]. An overview about different geometry calibration methods can be found in [6].

In previous publications, direction-of-arrival (DoA) information was employed to solve the geometry calibration task [2,7]. From this, however, no distance information can be obtained, leaving the scaling of the found geometry undetermined. Thus, additional information, e.g., gleaned from another modality, such as video [8], is required to fix the scaling.

In recent publications [9–11] it was shown how an estimate of the distance between the acoustic source and the microphone array can be obtained from the recorded audio signal. Distances as inputs allow to employ techniques for geometry calibration which were previously used to localize mobile devices based on signal strength information. In [12], for example, a system of equations based on the distance estimates is set up, where the unknown parameter vector consists of the Cartesian coordinates of the mobile device and a range variable which depended on the Cartesian coordinates. Hence, the system of equations is solved approximately by a constrained weighted least squares (CWLS) approach. The authors of [13] established also a system of equations but subtracted one equation from the system such resulting in an optimization problem that can be solved by unconstrained least squares (LS) without resorting to approximations. However, both approaches require the knowledge of the base station positions. This would correspond to the knowledge about the acoustic sources' position, which is not available in our scenario.

Besides the LS based approaches, also Multi Dimensional Scaling (MDS) based methods have been proposed for location estimation and geometry calibration. The authors of [14] modified MDS for mobile location estimation using time-of-arrival measurements of a signal emitted from the mobile station. In [15] MDS is extended such that it can handle missing distance information by applying a projection.

In this paper we present a geometry calibration algorithm that uses source-node distance estimates from a deep neural network (DNN) based acoustic distance estimator [11]. The distance estimation method does not require the nodes of the sensor network to be time synchronized, which is a major advantage over many other geometry calibration approaches. Our approach combines ideas from mobile phone localization, e.g., [13], and iterative optimization techniques. Additionally, we show by simulation that the used position estimator is efficient, since it reaches the CRLB.

The paper is organized as follows: In Sec. 2 distance-based geometry calibration is explained, followed by the CRLB derivation of the variance of the corresponding position estimates in Sec. 3. Implementation issues are discussed in Sec. 4, before experimental results are presented in Sec. 5 and conclusions are drawn in Sec. 6.

2. GEOMETRY CALIBRATION

We assume a setup with N sensor nodes at positions $\Omega_{\boldsymbol{P}} \coloneqq \{\boldsymbol{P}_1, \ldots, \boldsymbol{P}_N\}$ and K spatially distributed acoustic sources at positions $\Omega_{\boldsymbol{O}} \coloneqq \{\boldsymbol{O}_1, \ldots, \boldsymbol{O}_K\}$. Our geometry calibration procedure, named **G**eometry c**A**libration f**R**om **D**istance **E**stimates (GARDE), assumes that all nodes can observe the same local audio sources and that the node positions remain fixed.

The GARDE task can be summarized as follows: Based on the estimates $\hat{d}_{n,k}$, i.e., the distances between acoustic sources at unknown positions $\boldsymbol{O}_k$ and sensor nodes at unknown positions $\boldsymbol{P}_n$, the sensor nodes' positions should be inferred. The $\hat{d}_{n,k}$ stem from the DNN-based distance estimator that we proposed in [11]. GARDE delivers estimates of all unknown positions $\Omega = \Omega_{\boldsymbol{P}} \cup \Omega_{\boldsymbol{O}}$ by solving the following optimization problem:

$$\hat{\Omega} = \underset{\Omega}{\operatorname{argmin}} \underbrace{\sum_{k=1}^{K} \sum_{n=1}^{N} \left(\hat{d}_{n,k}^2 - \|\boldsymbol{P}_n - \boldsymbol{O}_k\|_2^2 \right)^2}_{J(\Omega)}. \tag{1}$$

Note that $K \geq \frac{2N}{N-2}$ must be fulfilled to arrive at an (over)determined system of equations.

Calculating the gradient of $J(\Omega)$ w.r.t. the g-th unknown position $\boldsymbol{P}_g$ and setting the result equal to zero gives:

$$\frac{\partial J}{\partial \boldsymbol{P}_g}\Big|_{\hat{\Omega}} \stackrel{!}{=} 0 \Leftrightarrow \sum_{k=1}^{K} \left(\hat{\boldsymbol{P}}_g - \hat{\boldsymbol{O}}_k\right) \underbrace{\left(\hat{d}_{g,k}^2 - \left\|\hat{\boldsymbol{P}}_g - \hat{\boldsymbol{O}}_k\right\|_2^2\right)}_{e_{gk}} = 0. \quad (2)$$

Similarly, for the gradient of $J(\Omega)$ w.r.t. the h-th unknown acoustic source position $\boldsymbol{O}_h$ it follows:

$$\frac{\partial J}{\partial \tilde{\boldsymbol{O}}_h}\Big|_{\hat{\Omega}} \stackrel{!}{=} 0 \Leftrightarrow \sum_{n=1}^{N} \left(\hat{\boldsymbol{P}}_n - \hat{\boldsymbol{O}}_h\right) \underbrace{\left(\hat{d}_{n,h}^2 - \left\|\hat{\boldsymbol{P}}_n - \hat{\boldsymbol{O}}_h\right\|_2^2\right)}_{e_{nh}} = 0. \quad (3)$$

For (2) and (3) no closed form solution exists. However, minimizing alternatingly the e_{gk} and e_{nh} should at least minimize the cost function. To this end, an iterative weighted least squares (WLS) approach can be set up, assuming either the positions $\hat{\boldsymbol{O}}_k$ in (2) or the positions $\hat{\boldsymbol{P}}_n$ in (3) to be fixed. The similarity of (2) and (3) allows to state a common solution for both problems.

Following the ideas of [13], the localization procedure has to select one sensor node as reference node, which has to be chosen judiciously, as the resulting localization error depends on the precision of the distance estimate of the selected sensor. Here, we always select the node with the smallest distance estimate to the acoustic source as reference node, as it was shown in [11] that the distance estimation error increases with the distance between audio source and sensor. Informal experiments showed an overall error reduction by approximately a factor of two compared to randomly selecting the reference node.

Let us assume that for some $\boldsymbol{P}_\nu$ $\hat{d}_{\nu,i} \leq \hat{d}_{j,i}$ $\forall i,j$ holds and that $\boldsymbol{P}_\nu$ is chosen to be the center of the used coordinate system ($\tilde{\boldsymbol{P}}_n = \boldsymbol{P}_n - \boldsymbol{P}_\nu$, $\tilde{\boldsymbol{O}}_k = \boldsymbol{O}_k - \boldsymbol{P}_\nu$). For this choice we get:

$$\left(\tilde{P}_{n,x} - \tilde{O}_{k,x}\right)^2 + \left(\tilde{P}_{n,y} - \tilde{O}_{k,y}\right)^2 = \hat{d}_{n,k}^2 \qquad \forall n \neq \nu \quad (4)$$

$$\tilde{O}_{k,x}^2 + \tilde{O}_{k,y}^2 = \hat{d}_{\nu,k}^2 \qquad n = \nu \quad (5)$$

with $n \in \{1,\ldots,N\}$ and $N > 3$ to guarantee an overdetermined system of equations. Subtracting (5) from all equations of (4) and stacking the resulting equations in a matrix formulation (excluding the ν-th one) gives:

$$\underbrace{\begin{bmatrix} 2\tilde{P}_{1,x} & 2\tilde{P}_{1,y} \\ \vdots & \vdots \\ 2\tilde{P}_{n,x} & 2\tilde{P}_{n,y} \end{bmatrix}}_{\boldsymbol{R}} \begin{bmatrix} \tilde{O}_{k,x} \\ \tilde{O}_{k,y} \end{bmatrix} = \underbrace{\begin{bmatrix} \hat{d}_{\nu,k}^2 + \tilde{P}_{1,x}^2 + \tilde{P}_{1,y}^2 - \hat{d}_{1,k}^2 \\ \vdots \\ \hat{d}_{\nu,k}^2 + \tilde{P}_{n,x}^2 + \tilde{P}_{n,y}^2 - \hat{d}_{N,k}^2 \end{bmatrix}}_{\boldsymbol{b}}$$

The DNN-based distance estimates are corrupted by an error which tends to be heteroscedastic (see [11]). So WLS is employed to estimate the location $\hat{\boldsymbol{O}}_k$ with

$$\hat{\boldsymbol{O}}_k = \left(\boldsymbol{R}^T \boldsymbol{W} \boldsymbol{R}\right)^{-1} \boldsymbol{R}^T \boldsymbol{W} \boldsymbol{b} + \boldsymbol{P}_\nu, \quad (6)$$

where the elements of the diagonal weighting matrix are chosen as $W_{n,n} = 1/\hat{d}_{n,k}^2$, reflecting the earlier mentioned observation that small distances exhibit a lower estimation error. The concrete definition of the distances $\tilde{d}_{n,k}$ can be found in Sec. 4.

The WLS solution from (6) can also be used to estimate the position of the sensor nodes from the distance estimates by exchanging the roles of $\hat{\boldsymbol{P}}$ and $\hat{\boldsymbol{O}}$. This results in an iterative positioning procedure used in GARDE, as it will be explained in Sec. 4. However, an initial guess is needed to start the iterative algorithm which can be either positions of the nodes or the observations. We propose to use a MDS-based initialization for the sensor node positions, which we present in the following.

2.1. MDS-based Initialization

MDS employs a distance matrix $\boldsymbol{D}$ containing all inter-node distances $D_{i,j}$ with $i,j \in [1,N]$ to estimate the positions of the senor nodes. But, the utilized DNN-based distance estimator delivers only a set of distances $\boldsymbol{A}_P = \{\hat{d}_{1,1},\ldots,\hat{d}_{N,K}\}$ between unknown positions of nodes $\boldsymbol{P}_n$ and unknown positions of sources $\boldsymbol{O}_k$. Neither distance estimates between sensors nor distance estimates between sources are available. To overcome this difficulty, we utilize the triangular inequality (compare Fig. 1)

$$\max_l(|\hat{d}_{i,l} - \hat{d}_{j,l}|) \leq D_{i,j} \leq \min_u(\hat{d}_{i,u} + \hat{d}_{j,u}), \quad (7)$$

where i,j are sensor node indices and u,l are acoustic source indices. The bounds can be used to obtain a first estimate for the distance between the nodes i and j:

$$\widehat{D}_{i,j} = [\max_l(|\hat{d}_{i,l} - \hat{d}_{j,l}|) + \min_u(\hat{d}_{i,u} + \hat{d}_{j,u})]/2. \quad (8)$$

Based on $\widehat{\boldsymbol{D}}$ an initial geometry is estimated via classical MDS as described in [16].

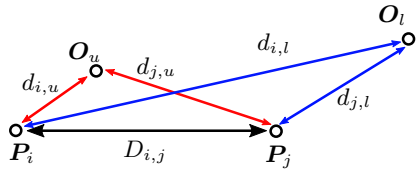

Fig. 1: Inter-node distance estimation example using subtraction (blue distances) and addition (red distances) of node to acoustic source distances for lower and upper bound approximation (see (7)).

3. Cramer-Rao lower bound

In this section the CRLB for the GARDE approach is derived, following the ideas of [17] and [18]. The CRLB for the GARDE position estimator, including acoustic source and sensor node positions, can be stated as follows. The observation errors $e_{n,k}$ which are given by

$$e_{n,k} = \hat{d}_{n,k} - \|\boldsymbol{P}_n - \boldsymbol{O}_k\|_2 \quad (9)$$

are assumed to follow a zero-mean Gaussian with variance σ_d^2, compare Fig. 2.

Given the position $\boldsymbol{O}_k$, the distribution of the N observations $\hat{\boldsymbol{d}}_k = \begin{bmatrix}\hat{d}_{1,k} & \cdots & \hat{d}_{N,k}\end{bmatrix}^T$ is

$$p\left(\hat{\boldsymbol{d}}_k | \boldsymbol{O}_k\right) = \xi \cdot \exp\left\{\sum_{n=1}^{N} \frac{-\left(\hat{d}_{n,k} - \tilde{d}_{n,k}\right)^2}{2\sigma_d^2}\right\}, \quad (10)$$

with constant $\xi = \left((2\pi)^{\frac{N}{2}} \prod_{n=1}^{N} \sigma_d\right)^{-1}$ and the ground truth distance from positions

$$\tilde{d}_{n,k} = \|\boldsymbol{P}_n - \boldsymbol{O}_k\|_2 = \sqrt{(P_{n,x} - O_{k,x})^2 + (P_{n,y} - O_{k,y})^2}.$$

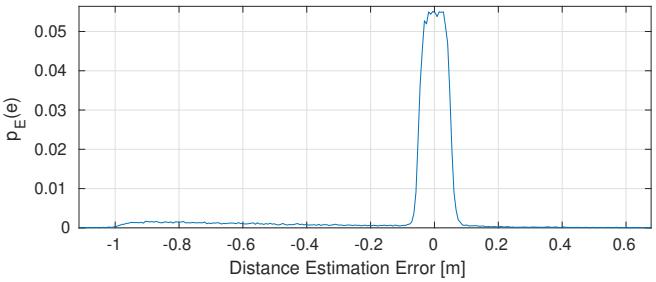

Fig. 2: Histogram of distance estimation errors from 80.000 observations.

The CRLB for the position estimate of $\boldsymbol{O}_k = \begin{bmatrix} O_{k,x} & O_{k,y} \end{bmatrix}^T$ is derived from the log-likelihood L_k with

$$L_k := \mathtt{ln}\left(p(\hat{\boldsymbol{d}}_k|\boldsymbol{O}_k)\right) = \mathtt{ln}(\xi) - \sum_{n=1}^{N} \frac{\left(\hat{d}_{n,k} - \tilde{d}_{n,k}\right)^2}{2\sigma_d^2}. \quad (11)$$

The first and second order derivatives w.r.t. $O_{k,x}$ of the log-likelihood are:

$$\frac{\partial\, L_k}{\partial O_{k,x}} = \sum_{n=1}^{N} \frac{1}{\sigma_d^2}\left(\hat{d}_{n,k} - \tilde{d}_{n,k}\right) \frac{\partial}{\partial O_{k,x}}\left(\tilde{d}_{n,k}\right) \quad (12)$$

$$\frac{\partial^2\, L_k}{(\partial O_{k,x})^2} = \sum_{n=1}^{N} \frac{-1}{\sigma_d^2}\left[\left(\frac{\partial\, \tilde{d}_{n,k}}{\partial O_{k,x}}\right)^2 - \left(\hat{d}_{n,k} - \tilde{d}_{n,k}\right) \frac{\partial^2\left(\tilde{d}_{n,k}\right)}{(\partial O_{k,x})^2}\right] \quad (13)$$

Applying the expectation operator to (13) results in

$$E\left\{\frac{\partial^2\, L_k}{(\partial O_{k,x})^2}\right\} = \sum_{n=1}^{N} \frac{-1}{\sigma_d^2} \frac{(P_{n,x} - O_{k,x})^2}{\|\boldsymbol{P}_n - \boldsymbol{O}_k\|_2^2} \quad (14)$$

Similarly, the derivatives w.r.t. $O_{k,y}$ can be found. For a short hand notation we summarize the terms by:

$$E\left\{\frac{\partial^2\, L_k}{(\partial O_{k,x})^2}\right\} = \sum_{n=1}^{N} \frac{-1}{\sigma_d^2} \frac{(P_{n,x} - O_{k,x})^2}{\|\boldsymbol{P}_n - \boldsymbol{O}_k\|_2^2} := \gamma_{xx} \quad (15)$$

$$E\left\{\frac{\partial^2\, L_k}{(\partial O_{k,y})^2}\right\} = \sum_{n=1}^{N} \frac{-1}{\sigma_d^2} \frac{(P_{n,y} - O_{k,y})^2}{\|\boldsymbol{P}_n - \boldsymbol{O}_k\|_2^2} := \gamma_{yy} \quad (16)$$

$$E\left\{\frac{\partial^2\, L_k}{\partial O_{k,x}\, \partial O_{k,y}}\right\} = E\left\{\frac{\partial^2\, L_k}{\partial O_{k,y}\, \partial O_{k,x}}\right\}$$

$$= -\sum_{n=1}^{N} \frac{1}{\sigma_d^2} \frac{(P_{n,x} - O_{k,x})(P_{n,y} - O_{k,y})}{\|\boldsymbol{P}_n - \boldsymbol{O}_k\|_2^2} := \gamma_{xy} = \gamma_{yx} \quad (17)$$

Finally, the CRLB can be stated by

$$\mathtt{CRLB}(\boldsymbol{O}_k) = -\left[E\left\{\begin{matrix} \frac{\partial^2 L_k}{\partial O_{k,x}\partial O_{k,x}} & \frac{\partial^2 L_k}{\partial O_{k,x}\partial O_{k,y}} \\ \frac{\partial^2 L_k}{\partial O_{k,y}\partial O_{k,x}} & \frac{\partial^2 L_k}{\partial O_{k,y}\partial O_{k,y}} \end{matrix}\right\}\right]^{-1},$$

whereby the variance of the estimate $\hat{\boldsymbol{O}}_k$ is lower bounded by

$$\mathtt{var}(\hat{\boldsymbol{O}}_{k,x}) \geq \frac{\gamma_{yy}}{\gamma_{xy}^2 - \gamma_{xx}\gamma_{yy}} \quad \text{and} \quad \mathtt{var}(\hat{\boldsymbol{O}}_{k,y}) \geq \frac{\gamma_{xx}}{\gamma_{xy}^2 - \gamma_{xx}\gamma_{yy}}.$$

The CRLB enables the estimation of a lower bound for the Root Mean Square Error (RMSE) value of position $\hat{\boldsymbol{O}}_k$ with

$$\mathtt{RMSE}(\hat{\boldsymbol{O}}_k) \geq \sqrt{\frac{\gamma_{xx} + \gamma_{yy}}{\gamma_{xy}^2 - \gamma_{xx}\gamma_{yy}}}. \quad (18)$$

Since the distances $\hat{d}_{n,k}$ are used both for estimating the positions of sensors and acoustic sources, the result of (18) can be easily adapted to obtain a lower bound for $\mathtt{RMSE}(\hat{\boldsymbol{P}}_n)$ by exchanging the sum index from n to k in γ_{xx}, γ_{xy} and γ_{yy} in (15)-(17).

4. GARDE IMPLEMENTATION DETAILS

The GARDE algorithm for geometry calibration is summarized in Alg. 1. Lines 2-5 correspond to the estimation of initial values of the nodes' positions using MDS. Based on these initial sensor position estimates the WLS localization is used to estimate the positions $\hat{\boldsymbol{O}}$ of the acoustic sources in line 6. To this end, the generic function $\widehat{\Omega}_{\boldsymbol{a}} = \mathtt{LSPos}(\widehat{\Omega}_{\boldsymbol{b}}, \boldsymbol{A}_P, \tilde{d}_{n,k})$ utilizes the distance observations $\boldsymbol{A}_P$ and sensor node positions $\widehat{\Omega}_{\boldsymbol{b}} = \widehat{\Omega}_{\boldsymbol{P}}$ to get source position estimates $\widehat{\Omega}_{\boldsymbol{a}} = \widehat{\Omega}_{\boldsymbol{O}}$. Thereby, $\widehat{\Omega}_{\boldsymbol{a}} = \mathtt{LSPos}(\widehat{\Omega}_{\boldsymbol{b}}, \boldsymbol{A}_P, \tilde{d}_{n,k})$ utilizes (6) with the weightening distances $\tilde{d}_{n,k}$ to determine each element of $\widehat{\Omega}_{\boldsymbol{a}}$ separately.

These first position estimates are refined by additional LS iterations in line 7. Subsequently, the iterative estimation, including annealing steps, is given by the for-loops. Annealing the variance parameter $\mu(g)$ turned out to be important to overcome unfavorable local optima. After finishing an iteration the newly found positions $\widehat{\Omega}$ are compared with the previous best estimate $\widehat{\Omega}_{pos,o}$ via the `OptSelect` function in line 11 with $\widehat{\Omega}_o = \mathtt{OptSelect}(\widehat{\Omega}, \widehat{\Omega}_o, \boldsymbol{A}_P)$. The function selects the geometry with the smallest average error between the observed distances $\boldsymbol{A}_P$ and the distances calculated from the estimated node and acoustic source positions.

In each iteration two auxiliary functions are used. The function $\widetilde{\Omega}_{\boldsymbol{P},fit} = \mathtt{Map2Ref}(\widehat{\Omega}_{\boldsymbol{P},fit} \rightarrow \widehat{\Omega}_{\boldsymbol{P}})$ maps the set of positions $\widehat{\Omega}_{\boldsymbol{P},fit}$ to a set of reference positions $\widehat{\Omega}_{\boldsymbol{P}}$ via rotation and translation operations. This processing step is necessary to enable the merge steps in line 20 and 22, where the results of the current iteration are combined with results of previous iterations. The second function $\widehat{\Omega}_{\boldsymbol{O},fit} = \mathtt{FitSelect}(\widehat{\Omega}, \boldsymbol{A}_P)$ selects a set of observations $\widehat{\Omega}_{\boldsymbol{O},fit}$ from $\widehat{\Omega}_{\boldsymbol{O}}$ that best fit the distances $\boldsymbol{A}_P$ and the assumed sensor positions $\widehat{\Omega}$ in terms of minimum average error.

5. EXPERIMENTS

For our experiments we utilize the image source method [19] from the implementation of [20] to simulate room impulse responses (RIRs) for 40 different setups with randomly placed nodes (4 nodes in a room) and 500 observations. The reverberation time T_{60} is selected to be 200 ms or 400 ms. The acoustic sources are simulated using speech signals from the TIMIT database [21], which are convolved with the RIRs to obtain the microphone signals. The DNN approach from [11] is used to estimate the distance between sensor node and acoustic source.

In Fig. 3 the CDF of the maximum distance error in $\boldsymbol{A}_p$ is compared against the maximum error after 30 iterations and annealing rounds for the fitness based selected observations (Alg. 1, line 10), i.e., the set of observations used to estimate the final geometry. The experiment shows that GARDE is able to identify outliers and to significantly reduce the maximum distance error.

Algorithm 1: GARDE algorithm

Data: Observed distances $\boldsymbol{A}_P$;
Init: $\alpha = 0.2$; $\beta = 0.2$;
for $i = 1 \rightarrow N$, $j = 1 \rightarrow N$ **do**
  $\widehat{D}_{ij} = \frac{1}{2}\max_l(|\hat{d}_{il} - \hat{d}_{jl}|) + \frac{1}{2}\min_k(\hat{d}_{iu} + \hat{d}_{ju})$
end
$\widehat{\Omega}_{\boldsymbol{P}} = \texttt{MultiDimensionalScaling}(\widehat{\boldsymbol{D}})$;
$\widehat{\Omega}_{\boldsymbol{O}} = \texttt{LSPos}\left(\widehat{\Omega}_{\boldsymbol{P}}, \boldsymbol{A}_P, \hat{d}_{n,k}\right)$;
$\widehat{\Omega} = \texttt{Iterate}(\widehat{\Omega}, \boldsymbol{A}_P)$;
$\widehat{\Omega}_o = \widehat{\Omega}$;
for $g = 1 \rightarrow \texttt{NumAnnealing}$ **do**
  $\widehat{\Omega} = \texttt{Iterate}(\widehat{\Omega}, \boldsymbol{A}_P)$;
  $\widehat{\Omega}_o = \texttt{OptSelect}(\widehat{\Omega}, \widehat{\Omega}_o, \boldsymbol{A}_P)$;
  $\hat{\boldsymbol{P}}_n = \hat{\boldsymbol{P}}_{n,o} + \mu(g) \cdot \texttt{randn}(),\ \forall n \in \{1, \ldots, N\}$;
  $\hat{\boldsymbol{O}}_k = \hat{\boldsymbol{O}}_{k,o} + \mu(g) \cdot \texttt{randn}(),\ \forall k \in \{1, \ldots, K\}$;
end
Result: $\widehat{\Omega}_o$;
Function $\texttt{Iterate}(\widehat{\Omega}, \boldsymbol{A}_P)$**:**
  for $i = 1 \rightarrow \texttt{NumIterations}$ **do**
    $\widehat{\Omega}_{\boldsymbol{O},fit} = \texttt{FitSelect}(\widehat{\Omega}, \boldsymbol{A}_P)$;
    $\widehat{\Omega}_{\boldsymbol{P},fit} = \texttt{LSPos}(\widehat{\Omega}_{\boldsymbol{O},fit}, \boldsymbol{A}_P, ||\hat{\boldsymbol{P}}_n - \hat{\boldsymbol{O}}_k||_2)$;
    $\widetilde{\Omega}_{\boldsymbol{P},fit} = \texttt{Map2Ref}(\widehat{\Omega}_{\boldsymbol{P},fit} \rightarrow \widehat{\Omega}_{\boldsymbol{P}})$;
    $\hat{\boldsymbol{P}}_n = \alpha \cdot \hat{\boldsymbol{P}}_n + (1-\alpha) \cdot \widetilde{\boldsymbol{P}}_{n,fit},\ \forall n \in \{1, \ldots, N\}$;
    $\widetilde{\Omega}_{\boldsymbol{O}} = \texttt{LSPos}(\widehat{\Omega}_{\boldsymbol{P}}, \boldsymbol{A}_P, \hat{d}_{n,k})$;
    $\hat{\boldsymbol{O}}_k = \beta \cdot \hat{\boldsymbol{O}}_k + (1-\beta) \cdot \widetilde{\boldsymbol{O}}_k,\ \forall k \in \{1, \ldots, K\}$;
  end
  return $\widehat{\Omega}$;

Simulated annealing is part of GARDE to overcome local minima. Its effect is illustrated in Fig. 4. For different number of iterations ($0 \rightarrow 30$), results with activated annealing ("A") and without annealing are compared. Overall, annealing significantly reduces the average geometry calibration error, i.e., the minimum RMSE between estimated geometry and ground truth after applying an optimal mapping including translation and rotation operations. The depicted results with 0 iterations reflect the performance of the MDS initialization.

Tab. 1 compares GARDE (30 iterations and annealing rounds) with the DoA-based geometry calibration method from [2]. The latter is combined with the scaling approach, we have presented in [11]. GARDE outperforms the DoA-based method and is less affected by reverberation.

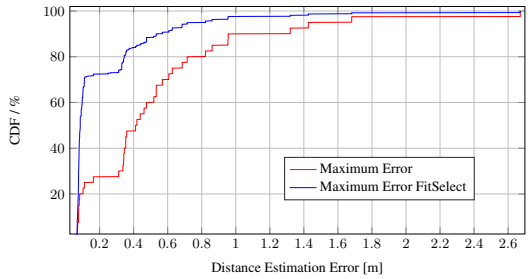

Fig. 3: Cumulative distribution function (CDF) of maximum distance error in $\boldsymbol{A}_p$ and selected subset of $\boldsymbol{A}_p$ by FitSelect function.

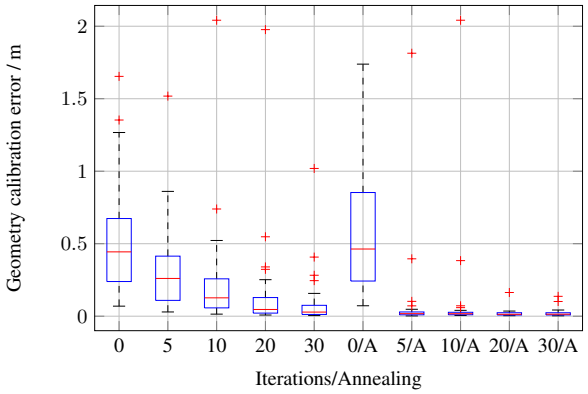

Fig. 4: Geometry calibration error for different number of iterations and optionally 30 annealing rounds (results marked with "/A").

Table 1: RMSE of the sensor positions for different geometry calibration methods

Method	$T_{60} = 200\,\text{ms}$	$T_{60} = 400\,\text{ms}$
DoA [2] + Scaling [11]	0.043 m	0.103 m
GARDE	0.017 m	0.032 m

In a final experiment the RMSEs of the sensor and acoustic source position estimates are compared against the RMSE bounds predicted by the CRLB. Our estimator reaches for both the predicted bounds as shown in Fig. 5.

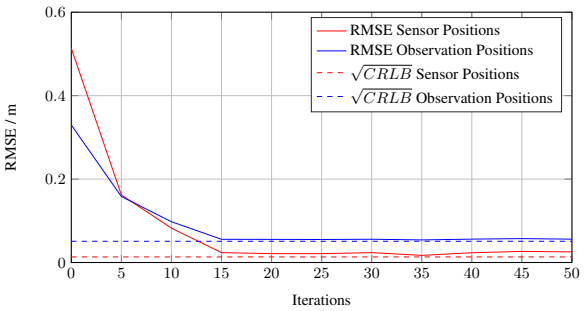

Fig. 5: Comparison between CRLB of estimator and RMSE of observation and sensor positions. Number of iterations and annealing rounds were equally chosen.

6. CONCLUSION

In this paper we have presented the GARDE algorithm, an iterative WLS-based algorithm for geometry calibration of WASNs. GARDE estimates the sensor nodes' positions and the positions of acoustic sources based on distance estimates, which are derived from the recorded audio signals. Furthermore, CRLBs for the geometry calibration approach are derived and compared against the simulation results showing the promising precision of the GARDE[1] algorithm.

Acknowledgment Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project 282835863.

[1]The Python software code of GARDE is available in the paderwasn repository, https://github.com/fgnt/paderwasn.

7. REFERENCES

[1] J. Schmalenstroeer, V. Leutnant, and R. Haeb-Umbach, "Audio-visual data processing for ambient communication," in *1st International Workshop on Distributed Computing in Ambient Environments within 32nd Annual Conference on Artificial Intelligence*, 2009.

[2] F. Jacob, J. Schmalenstroeer, and R. Haeb-Umbach, "Microphone array position self-calibration from reverberant speech input," in *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*, 2012, pp. 1–4.

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

[4] C. Taylor, A. Rahimi, J. Bachrach, H. Shrobe, and A. Grue, "Simultaneous localization, calibration, and tracking in an ad hoc sensor network," in *2006 5th International Conference on Information Processing in Sensor Networks*, 2006, pp. 27–33.

[5] O. Dorfan, Y.and Schwartz and S. Gannot, "Joint speaker localization and array calibration using expectation-maximization," *EURASIP Journal on Audio, Speech, and Music Processing*, 2020.

[6] A. Plinge, F. Jacob, R. Haeb-Umbach, and G. A. Fink, "Acoustic microphone geometry calibration: An overview and experimental evaluation of state-of-the-art algorithms," *IEEE Signal Processing Magazine*, vol. 33, no. 4, pp. 14–29, July 2016.

[7] J. Schmalenstroeer, F. Jacob, R. Haeb-Umbach, M.H. Hennecke, and G.A. Fink, "Unsupervised geometry calibration of acoustic sensor networks using source correspondences," *Interspeech*, pp. 2–5, 2011.

[8] F. Jacob and R. Haeb-Umbach, "Absolute geometry calibration of distributed microphone arrays in an audio-visual sensor network," *arXiv:1504.03128*, 2015.

[9] A. Brendel and W. Kellermann, "Distributed source localization in acoustic sensor networks using the coherent-to-diffuse power ratio," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 61–75, 2019.

[10] A. Brendel, A. Regensky, and W. Kellermann, "Probabilistic modeling for learning-based distance estimation," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen, Germany, Sept. 2019.

[11] T. Gburrek, J. Schmalenstroeer, A. Brendel, W. Kellermann, and R. Haeb-Umbach, "Deep neural network based distance estimation for geometry calibration in acoustic sensor networks," in *2020 28th European Signal Processing Conference (EUSIPCO)*, 2021, pp. 196–200.

[12] K. W. Cheung, H. C. So, W. . Ma, and Y. T. Chan, "Least squares algorithms for time-of-arrival-based mobile location," *IEEE Transactions on Signal Processing*, vol. 52, no. 4, pp. 1121–1130, April 2004.

[13] M. A. Spirito, "On the accuracy of cellular mobile station location estimation," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 3, pp. 674–685, May 2001.

[14] K. W. Cheung and H. C. So, "A multidimensional scaling framework for mobile location using time-of-arrival measurements," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 460–470, Feb 2005.

[15] A. Amar, Y. Wang, and G. Leus, "Extending the classical multidimensional scaling algorithm given partial pairwise distance measurements," *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 473–476, May 2010.

[16] I. Borg and P.J.F. Groenen, *Modern Multidimensional Scaling: Theory and Applications*, Springer, 2005.

[17] M. McGuire, K. Plataniotis, and E. Rogers, "A comparison of radiolocation methods for mobile terminals by distance measurements," in *Int. Conf. Wireless Communications*, 2000.

[18] J. Schmalenstroeer and R. Haeb-Umbach, "Investigations into bluetooth low energy localization precision limits," in *Proc. 24th European Signal Processing Conference (EUSIPCO 2016)*, Sep 2016.

[19] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, 1979.

[20] E.A.P. Habets, "Room impulse response generator," *Technische Universiteit Eindhoven, Tech. Rep*, vol. 2, no. 2.4, pp. 1, 2006.

[21] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "TIMIT acoustic-phonetic continuous speech corpus," 1993, Linguistic Data Consortium (LDC).