Self-Localization of Distributed Microphone Arrays Using Directional Statistics with DoA Estimation Reliability

Szymon Woźniak
AGH University of Science and Technology
Kraków, Poland
szymon.wozniak@agh.edu.pl

Konrad Kowalczyk
AGH University of Science and Technology
Kraków, Poland
konrad.kowalczyk@agh.edu.pl

Maximo Cobos
Universitat de València
València, Spain
maximo.cobos@uv.es

***Abstract*—This paper addresses the problem of self-localization of distributed microphone arrays from microphone recordings by following a two-step optimization procedure. In the first step, the relative geometry of the sources and arrays is inferred by the proposed maximum likelihood estimator. It is derived under the assumption that the acquired unit-norm vectors pointing towards the unknown source positions follow a von Mises-Fisher distribution in a D-dimensional space. In the second step, the absolute positions and synchronization offsets between the arrays are estimated from the inferred relative geometry by using the Least Squares procedure. To improve the accuracy of the method, we propose as well the use of a reliability measure for the estimated Directions of Arrival based on the presented directional statistics model. The results of numerical experiments confirm the validity of the proposed approach.**

***Index Terms*—microphone arrays, wireless acoustic sensor networks, distributed sensor networks, geometry calibration, maximum likelihood, directional statistics, circular statistics**

I. Introduction

Joint processing of the microphone signals acquired by distributed devices allows to localize sound sources in 2D/3D or to enhance signals coming from a set of desired positions [1], [2]. Information about the geometrical arrangement of the recording devices enables the exploitation of the spatial information present in the microphone signals. Unfortunately, in practice, the set-up is created in an ad-hoc fashion and the actual positions of the distributed devices are unknown.

The localization of scattered microphones using controlled sources was performed in [3] based on Time of Flight (ToF) measurements and in [4], [5] based on Time of Arrival (ToA) measurements. For an unknown source, self-localization of scattered microphones based on Time Difference of Arrival (TDoA) measurements was presented in [6], [7], [8]. The relative geometry of the distributed arrays up to a scaling constant can be found using Direction of Arrival (DoA) estimates only [9], [10], [11]. However, inference of the absolute microphone array positions requires the estimation of the scaling factor from the TDoA information, e.g. as done in [9], [10], [11], [12]. Furthermore, exploiting inter-array TDoAs, the synchronization offsets between the arrays can be found in addition to the geometry scaling factor [12].

In this paper, we present a two-step optimization method for estimating the positions of the distributed microphone arrays from the recorded microphone signals of unknown and uncontrolled sound sources. We focus on increasing the accuracy of the relative geometry estimation based on the DoAs estimated at each array, while the second optimization step in which the scaling factor and synchronization offsets are found is performed as described in [12]. The proposed maximum likelihood (ML) estimator for the relative geometry inference is derived under the assumption that the unit-length direction vectors pointing towards the unknown source positions follow a von Mises-Fisher (vMF) distribution [13]. The proposed cost function enables incorporating a reliability measure of the DoA measurements, which is shown here to increase the accuracy of the relative geometry inference, and in turn also the accuracy of the absolute array positions. The suitability of directional statistics for addressing the microphone array self-calibration problem has already been suggested in other works closely related to this contribution. The cost function in [11] was derived assuming a von Mises distribution of DoAs. Interestingly, a similar cost function was obtained heuristically from basic geometrical relations in [9], [10]. The contribution of this paper includes (i) the formulation of the ML cost function with directional statistics, which is derived using the von Mises-Fisher distribution, and (ii) the proposal of an adaptive DoA estimation reliability method motivated by the Cramer-Rao Bound (CRB) for DoAs.

II. Problem formulation

Consider a D-dimensional Euclidean space $\mathbb{R}^D$ spanned by standard basis $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_D)$ as an ambient space, in which N distributed microphone arrays, interchangeably referred to as nodes, record S sound events generated by a source at different locations. The position of the i-th node is denoted by vector $\mathbf{n}_i \in \mathbb{R}^D$, and the positions of all nodes are grouped into the matrix $\mathbf{N} = [\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times D}$, where $(\cdot)^{\mathrm{T}}$ denotes the transposition operator. Orientation of the i-th node in space is determined by its local orthonormal basis $(\mathbf{b}_{i,1}, \mathbf{b}_{i,2}, \ldots, \mathbf{b}_{i,D})$. The transformation matrix defined as $\mathbf{B}_i = [\mathbf{b}_{i,1}, \mathbf{b}_{i,2}, \ldots, \mathbf{b}_{i,D}]^{\mathrm{T}} \in \mathbb{R}^{D\times D}$ performs an isometric linear mapping from the standard basis to the local basis of the i-th node. Moreover, we assume that $\mathbf{B}_i$ is an element from a special orthogonal group $\mathrm{SO}(D)$. The position of the j-th sound event is denoted as vector $\mathbf{s}_j \in \mathbb{R}^D$, and the positions of all events are represented by matrix $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_S]^{\mathrm{T}} \in \mathbb{R}^{S\times D}$. The position of the j-th sound event in reference to the position of the i-th node in its local basis can be expressed by vector $\mathbf{p}_{i,j} \in \mathbb{R}^D$ defined as

$$\mathbf{p}_{i,j} = \mathbf{B}_i \left(\mathbf{s}_j - \mathbf{n}_i\right). \quad (1)$$

Figure 1 depicts the geometric relation for three distributed nodes and a single sound event in a $D = 2$ dimensional space.

Each node measures two types of quantities per sound event. The first one is the DoA of the j-th event at the i-th node and it is denoted by vector $\mathbf{d}_{i,j}$, while the TDoA of the j-th event between the i-th node and the k-th node is denoted as $\tau_{(i,k),j}$. The DoA and TDoA models are given, respectively, by

$$\mathbf{d}_{i,j} = \frac{\mathbf{p}_{i,j}}{\left\|\mathbf{p}_{i,j}\right\|} \quad \text{and} \quad \tau_{(i,k),j} = \frac{\left\|\mathbf{p}_{i,j}\right\| - \left\|\mathbf{p}_{k,j}\right\|}{c^{-1}} + \delta_i - \delta_k, \quad (2)$$

where $\|\cdot\|$ denotes the L_2 norm, c is the speed of sound, and δ_i and δ_k are the synchronization offsets for the i-th and k-th nodes between the local and reference timeline. Offsets of all arrays are grouped into a vector $\boldsymbol{\delta} = [\delta_1, \delta_2, \dots, \delta_N]^{\mathrm{T}} \in \mathbb{R}^N$.

The aim is to estimate the positions and orientations of all arrays stored in matrices $\mathbf{N}$ and $\mathbf{B}$, and the positions of all sound events $\mathbf{S}$. Since the main focus of this work is to increase the accuracy of the relative geometry estimation based on the DoA information, in the next section we derive the proposed maximum likelihood (ML) estimator based on directional statistics with an adaptive DoA estimation reliability.

III. Proposed method

In this section, we present a two-step ML optimization for microphone array self-localization. The ML estimator for finding the relative array geometry using directional statistics is derived in Sec. III-A. The method to incorporate robustness against DoA estimation errors into the proposed framework is presented in Sec. III-B. The final optimization step to obtain the absolute geometry and synchronization offsets is briefly discussed in Sec. III-C.

A. Relative geometry estimation using directional statistics

In a D-dimensional Euclidean space, a unit-length vector $\mathbf{d}_{i,j} \in \mathbb{R}^D$ given in (2) defines the direction of the j-th event in the local basis of the i-th microphone array. Assuming that the DoAs follow a von Mises-Fisher (vMF) distribution [13], the probability density function (PDF) of the measured DoA given by the unit-length vector $\check{\mathbf{d}}_{i,j}$ can be written as

$$f_p\left(\check{\mathbf{d}}_{i,j};\ \mathbf{d}_{i,j}, \kappa_{i,j}\right) \equiv Z(\kappa_{i,j})\, e^{\kappa_{i,j}\, \mathbf{d}_{i,j}^{\mathrm{T}}\, \check{\mathbf{d}}_{i,j}}, \quad (3)$$

where $Z(\kappa_{i,j})$ denotes the normalization parameter given by

$$Z(\kappa_{i,j}) = \frac{(\kappa_{i,j})^{\frac{D}{2}-1}}{(2\pi)^{\frac{D}{2}}\, I_{\frac{D}{2}-1}(\kappa_{i,j})}, \quad (4)$$

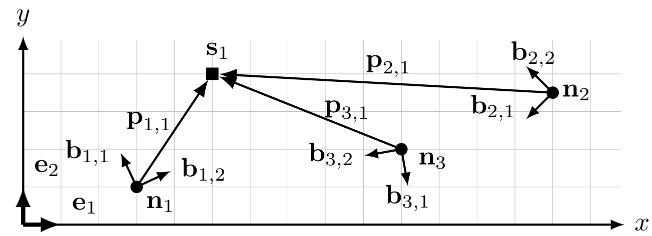

Fig. 1. Example geometry with three nodes and one sound event in a two-dimensional space.

$\kappa_{i,j}$ is the concentration of the PDF around the mean direction $\mathbf{d}_{i,j}$, and $I_m(\kappa)$ denotes the m-order modified Bessel function of the first kind. Vector $\mathbf{d}_{i,j} \in \mathbb{R}^D$ denotes the mean unit-length direction vector, computed by inserting (1) into (2). Note that the von Mises-Fisher distribution is isotropic on the $(D-1)$-dimensional sphere. Based on the PDF for a single measurement $\check{\mathbf{d}}_{i,j}$ given by (3), the negative log-likelihood function for all measurements can be written as

$$\mathcal{L}_{\mathrm{vMF}} = -\sum_{i=1}^{N}\sum_{j=1}^{S}\left[\ \kappa_{i,j}\ \mathbf{d}_{i,j}^{\mathrm{T}}\ \check{\mathbf{d}}_{i,j} + \ln Z(\kappa_{i,j})\ \right]. \quad (5)$$

The problem of finding the relative geometry of the distributed arrays and sound events can be solved by minimizing the following cost function

$$\hat{\mathbf{B}}, \tilde{\mathbf{N}}, \tilde{\mathbf{S}} = \underset{\mathbf{B},\mathbf{N},\mathbf{S}}{\arg\min}\ \mathcal{L}_{\mathrm{vMF}} \text{ subject to } \mathbf{B}_i \in \mathrm{SO}(D). \quad (6)$$

Since incorporation of the constraint $\mathbf{B}_i \in \mathrm{SO}(D)$ in an optimization procedure is non-trivial, one can exploit rotation matrices as well known generators of $\mathrm{SO}(D)$ for $D = 2$ and $D = 3$, and use the following substitution $\mathbf{B}_i = \mathbf{R}(\boldsymbol{\theta}_i)$, where $\mathbf{R}(\boldsymbol{\theta}_i)$ denotes the rotation matrix by angle vector $\boldsymbol{\theta}_i$. As a result, we search for unknown generators $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N]$ instead of the basis matrix $\mathbf{B}$, and the final problem can be formulated as

$$\hat{\boldsymbol{\Theta}}, \tilde{\mathbf{N}}, \tilde{\mathbf{S}} = \underset{\boldsymbol{\Theta},\mathbf{N},\mathbf{S}}{\arg\min}\ -\sum_{i=1}^{N}\sum_{j=1}^{S} \mathcal{G}_{i,j}, \quad (7)$$

$$\mathcal{G}_{i,j} = \kappa_{i,j}\frac{(\mathbf{s}_j - \mathbf{n}_i)^{\mathrm{T}}\mathbf{R}\left(\boldsymbol{\theta}_i\right)^{\mathrm{T}}\check{\mathbf{d}}_{i,j}}{\left\|\mathbf{s}_j - \mathbf{n}_i\right\|} + \ln Z(\kappa_{i,j}). \quad (8)$$

Matrices $\tilde{\mathbf{N}}$ and $\tilde{\mathbf{S}}$ define the relative positions of the distributed arrays and sound events up to an unknown scaling factor γ. In order to infer an absolute geometry, i.e. to find the final positions of sound events and distributed arrays, an estimate of the scaling factor $\hat{\gamma}$ is required.

B. Incorporating reliability of DoA estimates

In this section, we present a method to incorporate DoA estimate robustness into the directional statistics-based relative geometry estimator proposed in Sec. III-A. Room reverberation, noise, and array imperfections are the main factors affecting the accuracy of DoA measurements in real acoustic conditions. With an increasing distance between the sound event and the array, the ratio between the power of the direct-path event signal and the power of room reverberation plus noise decreases, resulting in increased DoA measurement errors. To take this consideration into our proposed framework, we propose to relate the reliability of the DoA estimates with the distance between the array and the sound events.

The distance-dependent concentration of the von Mises-Fisher distribution [13] can be expressed using the following mapping function

$$\kappa_{i,j} = \psi_\kappa \, \Gamma\left(-\beta_\kappa \left\|\mathbf{p}_{i,j}\right\|^2 - C_\kappa\right) + M_\kappa, \tag{9}$$

where $\Gamma(x) = [1 + \exp(-x)]^{-1}$ is a sigmoid function, parameter β_κ defines the steepness of the transition region, C_κ defines the sigmoid shift, and parameters ψ_κ and M_κ determine the maximum and minimum sigmoid values.

In order to determine the range of values in (9), the relation between the variance and concentration of the von Mises-Fisher distribution needs to be determined. The theoretical relation for variance is given by [13], [14]

$$\sigma^2 = 2\left(1 - \frac{I_{D/2}(\kappa)}{I_{D/2-1}(\kappa)}\right). \tag{10}$$

Note that due to numerical instabilities equation (10) can be computed only for a limited range of κ and obtaining an inverse equation for concentration as a function of variance is not straightforward. To this end, in this work we apply an asymptotic approximation for κ derived in [14] which was shown to match well as inverse function to (10). An approximate solution for the concentration is given by [14]

$$\hat{\kappa} = \frac{\bar{r}D - \bar{r}^3}{1 - \bar{r}^2}\ , \tag{11}$$

with

$$\bar{r} = \frac{I_{D/2}(\kappa)}{I_{D/2-1}(\kappa)} = \frac{2 - \sigma^2}{2} \tag{12}$$

which enables conveniently expressing concentration as a function of variance σ^2. Figure 2(a) and (b) depict the standard deviation σ as a function of the concentration parameter κ for $D = 2$ and $D = 3$ spatial dimensions, and the approximated concentration $\hat{\kappa}$ as a function of σ in 2D and 3D.

We can next set the maximum and minimum values of the sigmoid such that the concentration κ in (9) takes the values from within the desired range. As observed in Fig. 2, for low concentration values, e.g. $\kappa \ll 1$, a nearly uniform distribution is obtained, which is undesired. On the other hand, very high values of concentration would indicate very high confidence about the estimated direction vectors, which is usually not the case in practice. To this end, we propose to set the sigmoid parameters M_κ and ψ_κ based on the desired minimum and maximum standard deviation values using equations

$$M_\kappa = \frac{\bar{r}_{\min} D - \bar{r}^3_{\min}}{1 - \bar{r}^2_{\min}} \tag{13}$$

and

$$\psi_\kappa = \frac{\bar{r}_{\max} D - \bar{r}^3_{\max}}{1 - \bar{r}^2_{\max}} - M_\kappa\ , \tag{14}$$

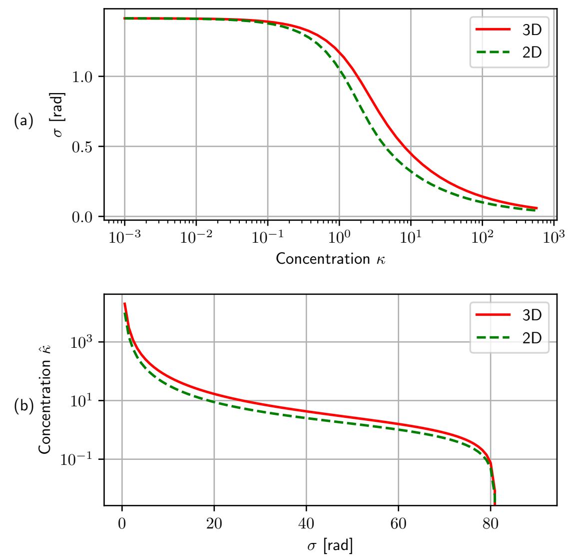

Fig. 2. (a) Theoretical standard deviation σ as a function of concentration κ and (b) approximated concentration $\hat{\kappa}$ as a function of standard deviation σ for the 2D (dashed line) and 3D (solid line) von Mises-Fisher distribution.

where $\bar{r}_{\min} = 0.5(2 - \sigma^2_{\min})$ and $\bar{r}_{\max} = 0.5(2 - \sigma^2_{\max})$. In order to select $\sigma^2_{\max}$ and $\sigma^2_{\min}$ values which match well the DoA measurement conditions, we propose to relate them to the Cramer-Rao bound (CRB) for DoA estimates described in the Appendix, which depends on several array parameters and signal-to-noise ratio (SNR). The minimum variance value is set here based on the CRB given by (21) as $\sigma^2_{\min} = \text{CRB}$, while its maximum value is set empirically as $\sigma^2_{\max} = A\,\sigma^2_{\min}$, where A is a heuristically found constant.

Having determined the maximum and minimum sigmoid values, the sigmoid shape is controlled using parameters β_κ and C_κ. These two parameters can be conveniently set based on the maximum and minimum values of $\left\|\mathbf{p}_{i,j}\right\|$ for all i, j pairs. These equations are given as

$$\beta_\kappa(\varepsilon) = \frac{\varepsilon}{\max\limits_{i,j}\left\|\mathbf{p}_{i,j}\right\|^2 - \min\limits_{i,j}\left\|\mathbf{p}_{i,j}\right\|^2}\ , \tag{15}$$

$$C_\kappa(\varepsilon, \varphi) = \varepsilon\left(\frac{\min\limits_{i,j}\left\|\mathbf{p}_{i,j}\right\|^2}{\max\limits_{i,j}\left\|\mathbf{p}_{i,j}\right\|^2 - \min\limits_{i,j}\left\|\mathbf{p}_{i,j}\right\|^2} - \varphi\right). \tag{16}$$

Figure 3 depicts example sigmoid shapes for different values of sigmoid parameters as a function of the shaping variables ε and φ.

Incorporation of the proposed reliability into (7) is implemented as follows. The ML function (7) is solved using a selected gradient-based optimizer, where $\mathcal{O}$ denotes a single iteration of the optimization. Every G-th iteration of optimization $\kappa_{i,j}$ is reweighted using (9). To this end, $\beta_\kappa(\varepsilon)$ and $C_\kappa(\varepsilon, \varphi)$ are recomputed every G-th iteration based on $\mathbf{p}_{i,j}$ and sigmoid shape parameters ε and φ. In this work, the value

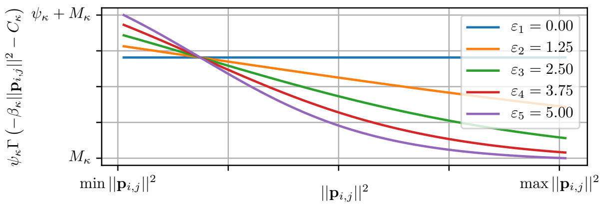

Fig. 3. Equation (9) for different parameters ε and $\varphi = 0.2$ using distances $\left\|\mathbf{p}_{i,j}\right\|$.

of ε is increased with time to introduce robustness against unreliable position estimates found shortly after random position initialization, while φ is kept constant. For convenience, the proposed reweighting method is summarized in Algorithm 1.

Algorithm 1: Proposed reweighting method

```
Data: ď_ij, φ, ε_1, ε_2, ..., ε_W,
Result: Θ̂, Ñ, S̃
Initialize: t = 1; w = 1; Θ̂_1, Ñ_1, S̃_1
while convergence is not achived do
    if t mod G = 1 and w ≤ W then
        recompute β_κ(ε_w) using (15)
        recompute C_κ(ε_w, φ) using (16)
        recompute κ_{i,j} for each pair (i,j) using (9)
        w ← w + 1
    end
    Θ̂_{t+1}, Ñ_{t+1}, S̃_{t+1} ← O(Θ̂_t, Ñ_t, S̃_t)
    t ← t + 1
end
```

C. Absolute geometry and synchronization offset estimation

The absolute geometry can be retrieved from the relative geometry using relations

$$\hat{\mathbf{N}} = \hat{\gamma}\,\tilde{\mathbf{N}} \qquad \text{and} \qquad \hat{\mathbf{S}} = \hat{\gamma}\,\tilde{\mathbf{S}}\,, \tag{17}$$

where $\hat{\gamma}$ is the estimated geometry scaling factor. The required scaling factor $\hat{\gamma}$ and the synchronization offsets $\hat{\boldsymbol{\delta}} = [\hat{\delta}_1, \hat{\delta}_2, \hat{\delta}_3, \ldots, \hat{\delta}_N]^{\mathrm{T}}$ between the timelines at the arrays and a reference timeline can be estimated by solving the following linear least squares (LS) problem [12]

$$\hat{\boldsymbol{\delta}}, \hat{\gamma} = \underset{\boldsymbol{\delta},\gamma}{\arg\min} \sum_{j=1}^{S} \sum_{i=1}^{N} \sum_{k=1}^{N} \left(\gamma\,\tilde{\tau}_{(i,k),j} + \delta_i - \delta_k - \hat{\tau}_{(i,k),j}\right)^2, \tag{18}$$

where the modelled TDoA $\tau_{(i,k),j}$ is given by

$$\tau_{(i,k),j} = \gamma\,\tilde{\tau}_{(i,k),j} + \delta_i - \delta_k \tag{19}$$

and the TDoA of the j-th event between the i-th and the k-th array for the relative geometry is given by

$$\tilde{\tau}_{(i,k),j} = c^{-1}\left(\left\|\tilde{\mathbf{s}}_j - \tilde{\mathbf{n}}_i\right\| - \left\|\tilde{\mathbf{s}}_j - \tilde{\mathbf{n}}_k\right\|\right). \tag{20}$$

Note that (18) can be conveniently solved using the closed-form solution proposed in [12].

IV. Experiment description and result evaluation

In this section, we evaluate the performance of the proposed self-localization method for $S = 20$ sound events and $N = 10$ distributed microphone arrays. The positions of events and 5-element circular microphone arrays are randomly selected on a 2D horizontal plane located at a height of 1.4 m in a $10 \times 10 \times 3.5$ [m] room with the reverberation time of 0.3 s simulated using the image-source method [15]. The microphone signals are synthesized as a convolution of anechoic speech of 3 s duration with the simulated room impulse responses (RIRs), and the white Gaussian noise at -60 dB, -50 dB and -40 dB level is added to the microphone signals. In order to focus our investigations on the relative geometry inference, the DoAs are estimated using the Steered-Response Power with Phase Transform (SPR-PHAT) method [16], while the TDoAs are assumed to be known, i.e., they correspond to the true geometry and synchronization offsets. We choose the Limited-memory-BFGS optimization algorithm [17] as a selected optimizer $\mathcal{O}$ to iteratively solve problem (7) with the reliability method described in Algorithm 1. Due to high non-convexity of the investigated DoA-based cost functions, node and event positions are initialized by adding position errors drawn from an isotropic normal distribution with variance of 16 [m] to the 'ground truth' positions in order to avoid Monte Carlo search for good initialization of the positions [18]. Note that in contrast to [10], [9], the Random Sample Consensus (RANSAC) [19] algorithm is not applied in order to be able to verify the gain offered by the proposed DoA reliability measure. However, note that one could additionally apply RANSAC in order to improve the robustness of the presented method to outliers. For incorporating reliability, the following set of parameters is used: $A = 20$, $\varphi = 0.2$, and the value of ε is increased every 50-th iteration of the optimization algorithm by $\varepsilon = 1.25$ from the initial value of $\varepsilon = 0$ to $\varepsilon = 5$. Performance of the proposed method is compared with the method presented in [12] using the root mean square error (RMSE) measure between the 'ground truth' and estimated position and synchronization offset values. We show the errors for the positions of distributed arrays $\mathrm{RMSE}(\hat{\mathbf{N}})$, the positions of sound events $\mathrm{RMSE}(\hat{\mathbf{S}})$, and the synchronization timeline offsets $\mathrm{RMSE}(\hat{\boldsymbol{\delta}})$. Each presented result is computed by averaging over the errors for all arrays (or events) and over 10 repetitions of each experiment for 100 random geometries.

Table IV presents the results for the microphone self-noise levels of -60, -50, and -40 dB in reference to the level of the source signal if it was collocated with the array, i.e., the further the source is from the array the lower the SNR becomes. As can be observed, the proposed method with directional statistics and measurement reliability achieves a nearly one-fourth higher array and source position estimation accuracy and nearly one-third timeline offset accuracy than the existing method [12] without the in-built reliability. The accuracy of direction vector estimation in terms of the event-array distance is shown in Fig. 4 for the proposed method. The depicted cosine distance is shown to increase nearly exponentially with

TABLE I
AVERAGED RESULTS FOR 100 RANDOM GEOMETRIES

	Noise [dB]	Method [12]	Proposed method
RMSE($\hat{\mathbf{N}}$) [m]	-60	0.467	0.315
	-50	0.690	0.537
	-40	0.861	0.708
RMSE($\hat{\mathbf{S}}$) [m]	-60	0.598	0.463
	-50	0.900	0.753
	-40	1.261	1.074
RMSE($\hat{\boldsymbol{\delta}}$) [ms]	-60	1.423	0.783
	-50	2.069	1.746
	-40	2.591	2.258

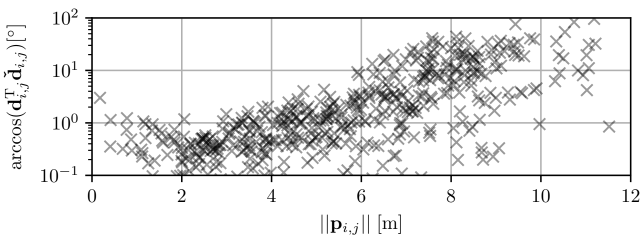

Fig. 4. The cosine distance (error) in degrees between the measured and the ground truth direction vectors as a function of event-array distance obtained in the performed experiment.

an increasing distance, apart from the near-field distances.

V. CONCLUSIONS

In this paper, a method for self-localization of the distributed microphone arrays is discussed. The relative array geometry is found using the proposed method derived using directional statistics with DoA measurement reliability. The results of numerical experiments indicate that an increase in inferred array position accuracy is achieved by the proposed method.

VI. ACKNOWLEDGEMENTS

This research received financial support from the Foundation for Polish Science under grant number First TEAM/2017–3/23 which is co-financed by the European Union under the European Regional Development Fund.

APPENDIX

The Cramer-Rao Bound for estimated azimuth and elevation angles using spherical arrays has recently been derived in [20]. For a circular arrangement, it can be shown that the CRB can be approximated as:

$$\mathrm{CRB}(\mathbf{d}_{i,j}) = \frac{D-1}{2M}\left(\frac{\lambda}{2\pi R}\right)^2 (\mathrm{SNR})^{-1}\left(\sum_{l=1}^{L}\left(\mathbf{d}_{i,j}^T\mathbf{m}_i^{(l)}\right)^2\right)^{-1}, \tag{21}$$

where M is the number of snapshots taken by the array, R is the radius, λ is the wavelength, SNR is the signal-to-noise ratio, L is the number of sensors in the array and $\mathbf{m}_i^{(l)}$ is the unit-norm vector indicating the direction of the l-th microphone within the array.

REFERENCES

[1] M. Cobos, A. Marti, and J. J. Lopez, “A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling,” *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 71–74, Jan 2011.

[2] K. Kowalczyk, O. Thiergart, A. Craciun, and E. A. P. Habets, “Sound acquisition in noisy and reverberant environments using virtual microphones,” in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct. 2013.

[3] M. Crocco and A. Del Bueand V. Murino, “A bilinear approach to the position self-calibration of multiple sensors,” *IEEE Trans. on Signal Process.*, vol. 60, no. 2, pp. 660–673, 2012.

[4] T. Le and N. Ono, “Closed-form and near closed-form solutions for TOA-based joint source and sensor localization,” *IEEE Trans. on Signal Process.*, vol. 64, no. 18, pp. 4751–4766, Sept 2016.

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

[6] S. Thrun, “Affine structure from sound,” in *Proc. Adv. Neural Inf. Process.Syst.*, 2006, pp. 1353–1360.

[7] N. Ono, H. Kohno, N. Ito, and S. Sagayama, “Blind alignment of asynchronously recorded signals for distributed microphone array,” in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct 2009, pp. 161–164.

[8] T. Le and N. Ono, “Closed-form and near closed-form solutions for TDOA-based joint source and sensor localization,” *IEEE Trans. on Signal Process.*, vol. 65, no. 5, pp. 1207–1221, March 2017.

[9] J. Schmalenstroeer, F. Jacob, R. Haeb-Umbach, M. Hennecke, and G. Fink, “Unsupervised geometry calibration of acoustic sensor networks using source correspondences,” in *INTERSPEECH*, 2011, pp. 597–600.

[10] F. Jacob, J. Schmalenstroeer, and R. Haeb-Umbach, “Microphone array position self-calibration from reverberant speech input,” in *Proc. Int. Workshop on Acoust. Signal Ench.*, Sept 2012, pp. 1–4.

[11] 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.* IEEE, 2013, pp. 116–120.

[12] 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.

[13] K. Mardia and P. Jupp, *Directional statistics*, John Wiley & Sons, 2009.

[14] A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra, “Clustering on the unit hypersphere using von Mises-Fisher distributions,” *J. Mach. Learn. Res.*, vol. 6, no. Sep, pp. 1345–1382, 2005.

[15] J. Allen and D. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, 1979.

[16] J. DiBiase, H. Silverman, and M. Brandstein, “Robust localization in reverberant rooms,” in *Microphone Arrays*, pp. 157–180. Springer, 2001.

[17] R. Byrd, P. Lu, J. Nocedal, and C. Zhu, “A limited memory algorithm for bound constrained optimization,” *SIAM J. on Scientific Computing*, vol. 16, no. 5, pp. 1190–1208, 1995.

[18] L. Wang, T. K. Hon, J. D. Reiss, and A. Cavallaro, “Self-localization of ad-hoc arrays using time difference of arrivals,” *IEEE Trans. on Signal Process.*, vol. 64, no. 4, pp. 1018–1033, Feb 2016.

[19] M. Fischler and R. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[20] K. Wong, Z. Morris, and C. Nnonyelu, “Rules-of-thumb to design a uniform spherical array for direction finding—Its Cramér–Rao bounds' nonlinear dependence on the number of sensors,” *J. Acoust. Soc. of Amer.*, vol. 145, no. 2, pp. 714–723, 2019.