Robust Self-Localization of Microphone Arrays Using a Minimum Number of Acoustic Sources

1st Matthias Schrammen
Institute of Communication Systems
RWTH Aachen University
Aachen, Germany
schrammen@iks.rwth-aachen.de

2nd Ahmad Hamad
ZESS - Center for Sensorsystems
University of Siegen
Siegen, Germany
hamad@zess.uni-siegen.de

3rd Peter Jax
Institute of Communication Systems
RWTH Aachen University
Aachen, Germany
jax@iks.rwth-aachen.de

***Abstract*—Multi-microphone signal processing is becoming increasingly popular in applications such as distant speech recognition or communication in adverse environments. To deploy source localization or signal enhancement algorithms like beamforming the locations of the microphones must be known. One well-studied approach to retrieve the relative positions of the microphones is based on time-difference-of-arrival (TDoA) measurements. However, current approaches are restricted to scenarios with a large number of sources or specific coherence assumptions. In this paper a non-iterative approach based on orthogonal geometric projection (OGP), which is able to perform a blind self-localization of the array in 2D with only two sources at arbitrary positions, is presented and extended to estimate a 3D array shape with only three sources. Furthermore, an efficient method for outlier correction in the pairwise distance (PD) estimates is proposed, that significantly reduces the position error.**

***Index Terms*—array shape estimation, self-localization, geometry calibration, time-difference-of-arrival, acoustic sensor networks**

I. Introduction

In scenarios where distant speech acquisition is required, ad-hoc microphone arrays and self-calibrating multi-microphone devices become increasingly popular. The performance of most signal enhancement algorithms used in that context relies on an accurate estimation of the positions of the microphones and drops if the estimates are erroneous. There exist numerous approaches in literature to retrieve the microphone positions blindly from acoustic sources. They can be categorized by the localization scenario that they are designed for and the estimation method they use [1]. One popular method measures the coherence between the signals of two microphones to estimate the pairwise distance (PD) between them [2], [3]. Its applicability is restricted to environments where the prevailing signal coherence matches a specific model, for example, that of an ideal diffuse sound field. Probably the most widely used method measures the time-difference-of-arrival (TDoA) between the signals recorded at two microphones [1], [4]–[7]. Using the speed of sound the PDs between the microphones can be estimated, if the source is located in the endfire direction of two microphones. For both methods classical multidimensional scaling (CMDS) is usually used to retrieve a coordinate representation of the microphone locations from the estimated PDs [2], [6]. Alternatively, iterative optimization methods can be used to find a set of microphone positions that best explains the observed TDoA measurements [8], [9]. However, the latter usually need careful initialization to converge to the correct solution [1], [10]. Recently, closed-form solutions for the more general problem of joint source and microphone localization using the low rank property of the matrix of TDoA measurements have been proposed [11]–[13]. They provide attractive results if constraints on the minimum number of microphones M and sources S are fulfilled. For a small number of sources and microphones (e.g. $M{+}S{<}13$ for 3D, $M{+}S{<}11$ for 2D) they fail to provide a solution for the array geometry. For calibrating a microphone array of arbitrary shape in 2D, [14] shows that the minimum number of microphones and sources is $M = 3$ and $S = 3$, if an arbitrarily low error is desired.

In this contribution we achieve a further reduction of the number of required sources, by imposing a regularity constraint on the array shape and derive a new non-iterative method to calibrate a 2D array with $M{=}3$ and $S{=}2$ and a 3D array with $M{=}4$ and $S{=}3$ only. After that we discuss the robustness of the method if the regularity constraint is violated. The method is based on an orthogonal geometric projection, that only depends on the relative angle between two sources. The main contribution is the estimation of this relative angle by measuring the similarity of the vector of TDoA measurements obtained from each source. Orthogonal projection techniques have been proposed for many applications such as direction-of-arrival estimation (e.g. MUSIC [15]) and subspace-based auto-calibration of microphone arrays [16]. However, in our contribution the projection is applied to TDoA measurements and maps them to microphone coordinates.

In practical applications one major source of errors in TDoA measurements is caused by reverberation. Acoustic reflections add additional peaks to the cross-correlation function that can wrongly be considered as the direct path peak, especially if noise is present. If a wrong peak is selected in the cross-correlation function, the resulting error of the TDoA estimate can be very high with a significant impact on the accuracy of the estimated microphone positions. To tackle this error source algorithms for matrix completion were proposed [17],

This work was funded by Intel Mobile Communications.

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

[18]. In this paper a simpler method for outlier detection and correction based on the zero-sum condition [19] is presented in Sec. III-B.

II. Signal Model and Pairwise Distance Estimation

In a reverberant environment the signal $u_i^{(l)}(t)$ at microphone i, which originates from the signal $s_l(t)$ at source l, can be modeled by convolving the source signal with the impulse response $h_{il}(t)$ between the source and the microphone, where t is the time index. The additive noise $v_i^{(l)}(t)$ is present at the i-th microphone. The resulting signal model is as follows

$$u_i^{(l)}(t) = h_{il}(t) * s_l(t) + v_i^{(l)}(t), \tag{1}$$

with $i = 1, 2, ..., M$ and $l = 1, 2, ..., S$, where M and S is the total number of microphones and sources, respectively. It is assumed that only one source is active at the same time and that all sources are located in the far-field of the array. Furthermore, it is assumed that all microphone signals are sampled with a synchronized clock.

To estimate the TDoA $\hat{T}_{ij}^{(l)}$ between two microphones i and j with respect to source l the generalized cross-correlation with phase transform (GCC-PHAT) is used according to

$$\hat{T}_{ij}^{(l)} = \arg\max_T \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{\Phi_{u_i^{(l)} u_j^{(l)}}(\Omega)}{|\Phi_{u_i^{(l)} u_j^{(l)}}(\Omega)|} e^{j\Omega T} \, \mathrm{d}\Omega, \tag{2}$$

where $\Phi_{u_i^{(l)} u_j^{(l)}}(\Omega)$ is the cross-power spectral density between $u_i^{(l)}$ and $u_j^{(l)}$ and Ω is the normalized radian frequency. From (2) a signed estimate for the PD between microphones i and j can be calculated via $\hat{d}_{ij}^{(l)} = c \cdot \hat{T}_{ij}^{(l)}$, where c is the speed of sound in air.

III. Orthogonal Geometric Projection

Our proposed approach for array shape estimation is based on two steps. At first we reduce the problem of array shape estimation to the estimation of the angle between two sources. This is done by a transformation of the coordinate system which we call *orthogonal geometric projection* (OGP). After that we introduce a new method to estimate the angle between two sources without knowledge of the microphone positions in Sec. III-A. We show that the method delivers exact results for array shapes forming a regular polygon and discuss the robustness of the method for irregular arrays.

The idea behind the OGP approach is that the measured PDs between each microphone and a reference microphone can be interpreted as an orthogonal projection of the true coordinates of the microphones on a line connecting the array center ($\times$) with one source, see Fig. 1. Any microphone can be chosen as the reference, i.e. it is located at the origin of the coordinate system. This projection cannot be inverted directly, because the corresponding projection matrix is singular. Therefore, we need a second source to span a 2D space and retrieve the coordinates of a planar array. Consequently, we need three sources for estimating the shape of a 3D array.

To retrieve an estimate $\hat{\boldsymbol{m}}_{ir}$ for the coordinates of microphone i relative to a fixed but arbitrary reference microphone

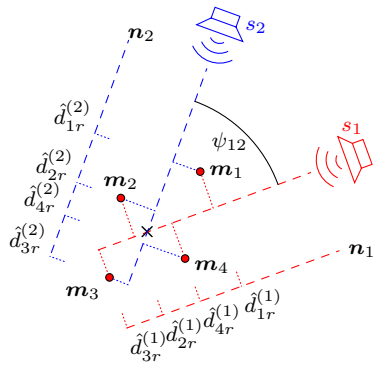


Fig. 1. Orthogonal projection of the positions $\boldsymbol{m}_i$ of microphones $i = 1, ..., 4$ to the relative PD estimates $\hat{d}_{ir}^{(1)}$ and $\hat{d}_{ir}^{(2)}$ generated by sources 1 and 2 for the 2D case. $\times$ indicates the array center.

r from the PD estimates $\hat{d}_{ir}^{(1)}$ and $\hat{d}_{ir}^{(2)}$, the angle ψ_{12} between the sources 1 and 2 must be known. For the special case of $\psi_{12} = 90°$, the PD estimates can directly be used as an estimate for the relative positions:

$$\hat{\boldsymbol{m}}_{ir} = [x_{ir}, y_{ir}]^T = \left[\hat{d}_{ir}^{(1)}, \hat{d}_{ir}^{(2)}\right]^T, \text{ for } \psi_{12} = 90° \tag{3}$$

For $\psi_{12} \neq 90°$, the positions of the microphones calculated via (3) are skewed versions $\hat{\boldsymbol{m}}'_{ir}$ of the true coordinates $\boldsymbol{m}_{ir}$, which can be concluded from Fig. 1. To retrieve an unskewed estimate $\hat{\boldsymbol{m}}_{ir}$, a matrix operation can be applied to $\hat{\boldsymbol{m}}'_{ir}$. The matrix is constructed from the normal vectors pointing from the center of the array to each source. Because we are only interested in the *relative* angle between the sources, we can assume without loss of generality that the normal vector $\boldsymbol{n}_1$ pointing to the first source lies in the x-axis, i.e. $\boldsymbol{n}_1 = [n_{1x}, n_{1y}]^T = [1, 0]^T$. Furthermore we can assume without loss of generality that the normal vector pointing towards the second source lies in the xy-plane. Using the properties $||\boldsymbol{n}_2|| = 1$ and $\boldsymbol{n}_1^T \boldsymbol{n}_2 = \cos(\psi_{12})$ this normal vector can be constructed as $\boldsymbol{n}_2 = [n_{2x}, n_{2y}]^T = [\cos(\psi_{12}), \sin(\psi_{12})]^T$. Now we have all elements to formulate the projection matrix for the 2D case as

$$\boldsymbol{P}_{\text{2D}} = \begin{pmatrix} n_{1x} & n_{1y} \\ n_{2x} & n_{2y} \end{pmatrix}. \tag{4}$$

For the 3D case the dimension of $\hat{\boldsymbol{m}}'_{ir}$ is extended to $\hat{\boldsymbol{m}}'_{ir} = [x_{ir}, y_{ir}, z_{ir}]^T = [\hat{d}_{ir}^{(1)}, \hat{d}_{ir}^{(2)}, \hat{d}_{ir}^{(3)}]^T$. Furthermore, with a third source, we can define three relative source angles ψ_{12}, ψ_{13} and ψ_{23}. Using the properties $\boldsymbol{n}_1^T \boldsymbol{n}_3 = \cos(\psi_{13})$, $\boldsymbol{n}_2^T \boldsymbol{n}_3 = \cos(\psi_{23})$ and $||\boldsymbol{n}_3|| = 1$ the normal vector $\boldsymbol{n}_3 = [n_{3x}, n_{3y}, n_{3z}]^T$ pointing towards the third source can be constructed as

$$n_{3x} = \cos(\psi_{13}) \tag{5}$$

$$n_{3y} = \frac{\cos(\psi_{23}) - \cos(\psi_{13}) \cos(\psi_{12})}{\sin(\psi_{12})} \tag{6}$$

$$n_{3z} = \sqrt{1 - n_{3x}^2 - n_{3y}^2}. \tag{7}$$

The projection matrix for the 3D case is defined as

$$\boldsymbol{P}_{\text{3D}} = \begin{pmatrix} n_{1x} & n_{1y} & 0 \\ n_{2x} & n_{2y} & 0 \\ n_{3x} & n_{3y} & n_{3z} \end{pmatrix}. \tag{8}$$

Applying the inverse of the projection matrix to the skewed

microphone positions results in the final position estimate

$$\hat{\boldsymbol{m}}_{ir} = \boldsymbol{P}_{dim}^{-1}\hat{\boldsymbol{m}}'_{ir},\ dim \in \{2\mathrm{D}, 3\mathrm{D}\}. \tag{9}$$

A. Estimation of the angle between sources

In general the angle $\psi_{l_1 l_2}$ between two sources l_1 and l_2 is not known and has to be estimated. No direct information on the angle between the axes in Fig. 1 is available. Therefore, we exploit a relation between the PD estimates and the relative source angle for array shapes forming a regular polygon. After that, we show the robustness of this relation to irregular shapes. The new method measures the similarity between all PD estimates $\boldsymbol{p}_{l_1}$ and $\boldsymbol{p}_{l_2}$ of sources l_1 and l_2 with the inner product. The inner product is then interpreted as an estimate $\hat{\psi}_{l_1 l_2}$ for the relative angle $\psi_{l_1 l_2}$ between the sources according to

$$\hat{\psi}_{l_1 l_2} = \arccos\left(\frac{\boldsymbol{p}_{l_1}^T \boldsymbol{p}_{l_2}}{||\boldsymbol{p}_{l_1}||\,||\boldsymbol{p}_{l_2}||}\right), \tag{10}$$

where $\boldsymbol{p}_q = [\hat{d}_{11}^{(q)}, \hat{d}_{21}^{(q)}, ..., \hat{d}_{ij}^{(q)}, ..., \hat{d}_{MM}^{(q)}]^T$ and $q \in \{l_1, l_2\}$. The estimated PDs $\hat{d}_{ij}^{(q)}$ can be expressed by the true PDs d_{ij} with $\hat{d}_{ij}^{(q)} = d_{ij}\cos(\theta_q - \alpha_{ij})$, where θ_q is the angle between source q and the horizontal axis and α_{ij} is the angle between the edge connecting $\boldsymbol{m}_i$ and $\boldsymbol{m}_j$ and the horizontal axis. Using basic trigonometric relations (10) can be expressed as

$$\hat{\psi}_{l_1 l_2} = \arccos\left(\frac{\cos(\psi_{l_1 l_2})C_0 + C}{\sqrt{C_0 + A_{l_1}}\sqrt{C_0 + A_{l_2}}}\right), \tag{11}$$

with

$$C_0 = \sum_{i=1}^{M}\sum_{j=1}^{M} d_{ij}^2,\ C = \sum_{i=1}^{M}\sum_{j=1}^{M} d_{ij}^2 \cos(\theta_{l_1} + \theta_{l_2} - 2\alpha_{ij}), \tag{12}$$

$$\text{and for } q \in \{l_1, l_2\},\ A_q = \sum_{i=1}^{M}\sum_{j=1}^{M} d_{ij}^2 \cos(2\theta_q - 2\alpha_{ij}). \tag{13}$$

For regular polygons it can be shown straightforwardly that the sum of cosines in (12) and (13) evaluates to zero. Therefore, $A_{l_1,l_2} = C = 0$ for regular polygons. Then, it follows with (11) that $\hat{\psi}_{l_1 l_2} = \psi_{l_1 l_2}$.

If the array shape is not regular the estimation according to (10) introduces an error. To give a quantitative impression of the error in estimating ψ for irregular shapes, a simulation with an array of four microphones and two sources was carried out. The irregularity of the array shape is varied by adding Gaussian noise to the coordinates of the square array shape. The estimation of ψ_{12} is then carried out for 1000 realizations of the irregular shape. In Fig. 2 these realizations of the irregular array are shown for a standard deviation of the noise of $\sigma_{\text{noise}} = 0.1(d_0/6)$ and $\sigma_{\text{noise}} = (d_0/6)$, where $d_0 = 60\,\text{cm}$ is the side length of the undisturbed array. The scaling by $d_0/6$ ensures that 99.7 % ($3\sigma_{\text{noise}}$) of the disturbed coordinates lie inside a radius of $d_0/2$ around the undisturbed corners. This defines a meaningful maximum amount of irregularity, because a change of the orientation of the square, i.e. a swap of the corners, is prevented. For the evaluation of the performance we calculate the absolute relative error of the angle estimation as $\psi_{\text{err, rel}} = |(\psi_{12} - \hat{\psi}_{12})/\psi_{12}|$. Furthermore, we investigate the effect of an erroneous angle estimation on the estimate for the

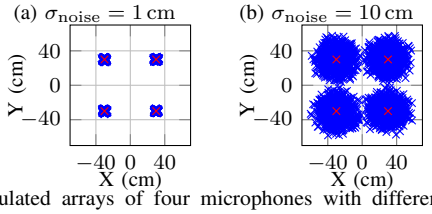


Fig. 2. Simulated arrays of four microphones with different amount of irregularity. The regular square is highlighted in red.

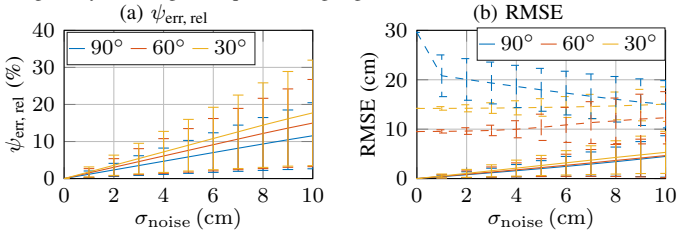


Fig. 3. Relative error $\psi_{\text{err, rel}}$ of the angle estimation (a) and RMSE of the position estimate (b) for $\psi_{12} = \{90°, 60°, 30°\}$. Solid and dashed lines indicate average results for OGP and CMDS, respectively. Vertical error bars indicate ± one standard deviation over 1000 trials.

microphone positions calculated by (9). The performance of the position estimation is evaluated with the Root-Mean-Squared-Error (RMSE) between the true ($\boldsymbol{m}_i$) and the estimated ($\hat{\boldsymbol{m}}_i$) microphone positions. For the RMSE we also present results of a CMDS approach, that uses the same PD estimates as the OGP approach but does not involve an angle estimation. For each PD two estimates (one for each source) are available. Similar to the approach in [6] our CMDS implementation takes the maximum of these estimates before calculating the coordinate representation. Fig. 3(a) and Fig. 3(b) show the results for $\psi_{\text{err, rel}}$ and for the RMSE, respectively, over a varying amount of irregularity. Solid and dashed lines indicate the average performance for OGP and CMDS, respectively. The vertical bars indicate ± one standard deviation over 1000 realizations. Three true source angles $\psi_{12} \in \{90°, 60°, 30°\}$ were evaluated. It can be seen in Fig. 3(a) that the error in estimating ψ_{12} increases linearly with larger irregularity for all source angles. However, the larger ψ_{12} the slower the increase of the error. This can be explained by the fact that sources spanning a larger relative angle provide a larger amount of independent TDoA observations, which is beneficial for the estimation of ψ_{12} via (10). Fig. 3(b) shows that the RMSE of the OGP approach (solid) behaves proportional to $\psi_{\text{err, rel}}$. The mean RMSE of CMDS (dashed) only weakly depends on the amount of irregularity and is significantly larger than for OGP. This is because not for all true PDs a source in endfire position is available, which results in biased PD estimates. It can be concluded that the violation of the regularity assumption introduces an error in the OGP estimates. However, this error is smaller than the error of a conventional CMDS algorithm even for very irregular shapes. For a correct estimation of ψ in the 2D case the array shape must have a two dimensional structure. Thus, the minimum number of microphones is three. Accordingly, in the 3D case the minimum number of microphones is four.

B. Detection and correction of outliers

For $M > 4$ there is a high redundancy in the PD estimates originating from *one* source. Therefore, it is possible to detect

and correct erroneous PD estimates. The detection of erroneous values is achieved by checking the zero-sum condition for each triple of PDs. This sum should be zero if the three PD estimates under test are correct [19]. This is because the PDs generated by the signal from one source lie on a line (cf. Fig. 1). Note that in Fig. 1 only the PD estimates relative to the reference microphone are shown. For the proposed outlier correction method *all* PD estimates generated by the same source l are used. In practice, for each PD estimate $\hat{d}_{ij}^{(l)}$, the zero-sum condition is checked with a small constant ϵ according to

$$|\hat{d}_{ik}^{(l)} + \hat{d}_{kj}^{(l)} - \hat{d}_{ij}^{(l)}| \leq \epsilon, \tag{14}$$

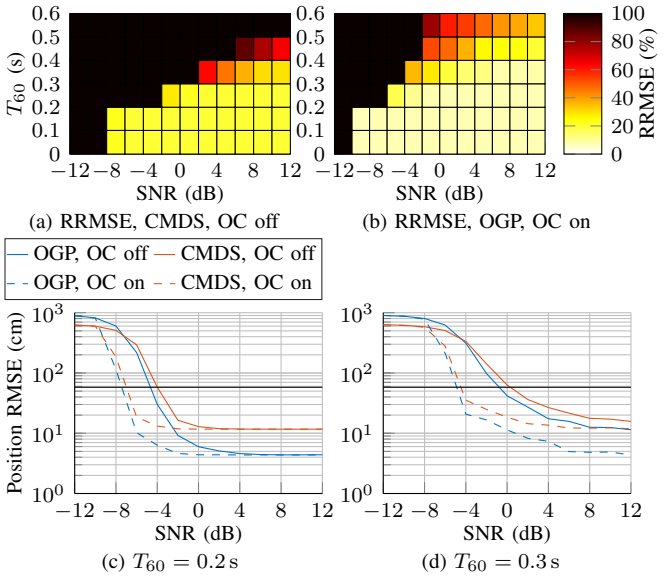
with $k \in \{1, 2, ..., M \,|\, k \neq i, j\}$. If no k is found that satisfies (14), $\hat{d}_{ij}^{(l)}$ is marked as an outlier. After that, for each outlier all possible candidate values are calculated via $|\hat{d}_{ik}^{(l)} + \hat{d}_{kj}^{(l)}|$. Then the most frequent candidate value is chosen as corrected value for $\hat{d}_{ij}^{(l)}$. If two or more candidate values are equally frequent, the smallest one is used. This method can only correct outliers if at least $M-1$ PDs are correct and these $M-1$ PDs provide information on all microphones. If no correction is possible the initial estimate before outlier detection is used. In the following $\epsilon = 3\,\mathrm{mm}$ is used.

IV. Evaluation

A. Simulation setup

To evaluate the performance of the proposed algorithm for array shape estimation, the Root-Mean-Squared-Error (RMSE) between the true ($\boldsymbol{m}_i$) and the estimated ($\hat{\boldsymbol{m}}_i$) microphone positions is used. The RMSE allows a more meaningful interpretation of the performance if it is related to the dimensions of the array under investigation. Therefore, the relative RMSE (RRMSE) is defined as $\mathrm{RRMSE} = \mathrm{RMSE}/||\boldsymbol{M}||$, where $||\boldsymbol{M}|| = \sqrt{\frac{1}{M}\sum_{i=1}^{M}||\boldsymbol{m}_i - \bar{\boldsymbol{m}}||^2}$ and $\bar{\boldsymbol{m}} = \frac{1}{M}\sum_{i=1}^{M}\boldsymbol{m}_i$. Because only the relative and not the absolute position of the microphone array is of interest, a Procrustes analysis is performed. It allows the estimated array shape to be an arbitrary rotated or translated version of the true array shape without increasing the error. However, scaling as a transformation is not allowed [20].

For the simulation the room impulse responses $h_{il}(t)$ from source l to microphone i are generated with the image source method [21]. To allow for a better comparison of the results, the room dimensions and the array geometry are chosen according to the real world setup used by Plinge et al. in their overview article [1]. The room has a dimension of $3.7\,\mathrm{m} \times 6.8\,\mathrm{m} \times 2.6\,\mathrm{m}$. For the source and the noise signals white Gaussian noise is used. Now the microphone signals are calculated with the generated impulse responses for T_{60} between 0 s and 0.6 s in steps of 0.1 s and an SNR from -12 dB to 12 dB in steps of 2 dB. For each SNR the RMSE and RRMSE of the position estimate are averaged over 30 trials to get more meaningful results. We evaluate the performance on a *microphone configuration* calibration scenario where the microphones are distributed across the room. This configuration is probably more relevant than a compact array, because it is very likely that the positions



(a) RRMSE, CMDS, OC off (b) RRMSE, OGP, OC on

(c) $T_{60} = 0.2\,\mathrm{s}$ (d) $T_{60} = 0.3\,\mathrm{s}$

Fig. 4. (a) and (b) show the RRMSE over SNR and T_{60} for CMDS without outlier correction (OC) and for OGP with OC, respectively. (c) and (d) show the RMSE over SNR for $T_{60} = 0.2\,\mathrm{s}$ and $T_{60} = 0.3\,\mathrm{s}$, respectively. The results for OGP (—) and CMDS (—) are shown without (solid) and with (dashed) outlier correction (OC). The horizontal black line indicates the RMSE where the relative error (RRMSE) is 100 %.

of the microphones change over time. This makes online calibration inevitable. However, it is also more challenging, at least for the OGP approach, because the far-field assumption is more violated than for a compact array. The 2D microphone configuration consists of three subarrays that are located at the corners of a triangle with an edge length of $d_0 = 1\,\mathrm{m}$. Each subarray consists of a circular array of five microphones with a radius of 5 cm. This microphone configuration shows a regularity property leading to an exact angle estimation with (10) if no disturbance is present. Therefore, similar to the evaluation in Sec. III-A, we add Gaussian noise with $\sigma_{\mathrm{noise}} = 0.5(d_0/6) = 8.3\,\mathrm{cm}$ to the centers of the subarrays and average the results over 20 realizations of the irregular shape. The sources are located at a distance of 1.5 m from the center of the microphone configuration and span an angle of $\psi_{12} = 70°$. Two algorithms are used to perform the array calibration. The proposed OGP and the CMDS approach. Both, OGP and CMDS, were evaluated with and without using the outlier correction described in Sec. III-B. The outlier correction was applied to the PD estimates generated by the individual sources.

B. Simulation results

The RRMSE in % is shown over SNR and T_{60} for CMDS without outlier correction (OC) in Fig. 4(a) and OGP with OC in Fig. 4(b). An RRMSE larger than 100 % is clipped in the colorbar. In both figures a sharp decrease of the RRMSE can be observed when either decreasing T_{60} or increasing the SNR. With OC in Fig. 4(b) the size of the region with errors above 100 % is significantly smaller than without OC in Fig. 4(a). Furthermore, the minimum achievable error for OGP in Fig. 4(b) is significantly reduced compared to CMDS in Fig. 4(a). A minimum RRMSE of 7 % is achieved for the

OGP approach. This corresponds to an RMSE of $4.1\,\mathrm{cm}$. For the CMDS approach the minimum position RRMSE was $20\,\%$, which corresponds to an RMSE of $11.7\,\mathrm{cm}$. To analyze the performance in more detail, we show the RMSE in cm for two fixed reverberation times. Fig. 4(c) shows the RMSE for $T_{60} = 0.2\,\mathrm{s}$. The horizontal black line indicates the RMSE where the relative error (RRMSE) is $100\,\%$. The OGP method without OC (—) significantly outperforms the CMDS method (—) when the SNR is increased above $-4\,\mathrm{dB}$. If OC is used (dashed lines), the RMSE of both methods is significantly reduced for relevant SNRs between $-8\,\mathrm{dB}$ and $0\,\mathrm{dB}$. For $\mathrm{SNR} \geq 0\,\mathrm{dB}$ CMDS saturates at a higher RMSE than OGP and neither CMDS nor OGP benefit from activated OC. This shows that the performance of CMDS is limited by non-endfire TDoA estimates rather than reverberation and that OGP correctly takes this into account by applying the inverse projection. Fig. 4(d) shows that for $T_{60} = 0.3\,\mathrm{s}$ and without OC (solid lines) OGP only slightly outperforms CMDS for $\mathrm{SNR} \geq -4\,\mathrm{dB}$. However, OGP can benefit from OC to larger extent (dashed lines) than CMDS resulting in a significantly lower RMSE for $\mathrm{SNR} \geq 4\,\mathrm{dB}$ and achieving the same minimum RMSE as with $T_{60} = 0.2\,\mathrm{s}$. This confirms that the OC method proposed in Sec. III-B effectively mitigates errors caused by reverberation. The practical relevance of the proposed method can be confirmed when the achieved RMSE is compared to the tolerable RMSE for typical array applications like source localization. In [1] it is shown, for a similar microphone configuration with five subarrays, that an RMSE below $10\,\mathrm{cm}$ is sufficient to achieve the minimum source localization error that is possible with a reverberation of $T_{60} = 0.5\,\mathrm{s}$.

V. Conclusion

In this paper a new approach for blind estimation of microphone positions based on acoustic sources was presented. Only two acoustic sources for a 2D array geometry and only three sources for a 3D array geometry are needed. It was shown that the new approach based on orthogonal geometric projection is able to estimate the positions with sufficient accuracy, even if the assumed regularity of the array shape is disturbed. This relaxes the constraints for acoustic microphone geometry calibration imposed by other state-of-the-art approaches. Neither many sources nor special coherence characteristics have to be assumed. Another advantage of the proposed estimation algorithm is that it is non-iterative. It does not need a proper initialization as algorithms for non-convex optimization or special assumptions to converge. However, the estimates of the proposed method may be used as a robust initialization of iterative algorithms for refinement or if more than the necessary number of sources are available. Furthermore a simple yet effective method for correcting outliers in the PD estimates was introduced. It provides a significant reduction of the estimation error compared to no outlier correction, especially in reverberant environments.

References

[1] 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.

[2] O. Schwartz, A. Plinge, E. A. P. Habets, and S. Gannot, "Blind microphone geometry calibration using one reverberant speech event," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2017, pp. 131–135.

[3] 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, Mar. 2008.

[4] L. Wang, T. K. Hon, J. D. Reiss, and A. Cavallaro, "Self-Localization of Ad-Hoc Arrays Using Time Difference of Arrivals," *IEEE Transactions on Signal Processing*, vol. 64, no. 4, pp. 1018–1033, Feb. 2016.

[5] M. Parviainen, P. Pertilä, and M. S. Hämäläinen, "Self-localization of wireless acoustic sensors in meeting rooms," in *2014 4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)*, May 2014, pp. 152–156.

[6] P. Pertilä, M. Mieskolainen, and M. S. Hämäläinen, "Passive self-localization of microphones using ambient sounds," in *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, Aug. 2012, pp. 1314–1318.

[7] M. Pollefeys and D. Nister, "Direct computation of sound and microphone locations from time-difference-of-arrival data," in *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, Mar. 2008, pp. 2445–2448.

[8] Y. Rockah and P. Schultheiss, "Array shape calibration using sources in unknown locations–Part II: Near-field sources and estimator implementation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 6, pp. 724–735, June 1987.

[9] Johannes Wendeberg and Christian Schindelhauer, "Polynomial time approximation algorithms for localization based on unknown signals," in *International Symposium on Algorithms and Experiments for Sensor Systems, Wireless Networks and Distributed Robotics*. 2012, pp. 132–143, Springer.

[10] M. H. Hennecke and G. A. Fink, "Towards acoustic self-localization of ad hoc smartphone arrays," in *2011 Joint Workshop on Hands-free Speech Communication and Microphone Arrays*, May 2011, pp. 127–132.

[11] Marco Crocco, Alessio Del Bue, and Vittorio Murino, "A Bilinear Approach to the Position Self-Calibration of Multiple Sensors," *IEEE Transactions on Signal Processing*, vol. 60, no. 2, pp. 660–673, Feb. 2012.

[12] Yubin Kuang, Simon Burgess, Anna Torstensson, and Kalle Astrom, "A complete characterization and solution to the microphone position self-calibration problem," May 2013, pp. 3875–3879, IEEE.

[13] T. K. Le and N. Ono, "Closed-Form and Near Closed-Form Solutions for TDOA-Based Joint Source and Sensor Localization," *IEEE Transactions on Signal Processing*, vol. 65, no. 5, pp. 1207–1221, Mar. 2017.

[14] Y. Rockah and P. Schultheiss, "Array shape calibration using sources in unknown locations–Part I: Far-field sources," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 3, pp. 286–299, Mar. 1987.

[15] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[16] M. Viberg and A. L. Swindlehurst, "A Bayesian approach to auto-calibration for parametric array signal processing," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, pp. 3495–3507, Dec. 1994.

[17] Marco Compagnoni, Alessia Pini, Antonio Canclini, Paolo Bestagini, Fabio Antonacci, Stefano Tubaro, and Augusto Sarti, "A Geometrical–Statistical Approach to Outlier Removal for TDOA Measurements," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 3960–3975, Aug. 2017.

[18] Jose Velasco, Daniel Pizarro, Javier Macias-Guarasa, and Afsaneh Asaei, "TDOA Matrices: Algebraic Properties and Their Application to Robust Denoising With Missing Data," *IEEE Transactions on Signal Processing*, vol. 64, no. 20, pp. 5242–5254, Oct. 2016.

[19] Jan Scheuing and Bin Yang, "Disambiguation of TDOA Estimates in Multi-Path Multi-Source Environments (DATEMM)," in *2006 IEEE International Conference on Acoustics Speed and Signal Processing Proceedings*, Toulouse, France, 2006, vol. 4, pp. IV–837–IV–840, IEEE.

[20] John C Gower and Garmt B Dijksterhuis, *Procrustes Problems*, Oxford University Press, Jan. 2004.

[21] Jont B. Allen and David A. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979.