Geometric calibration of distributed microphone arrays from acoustic source correspondences

S.D.Valente, M.Tagliasacchi, F. Antonacci, P. Bestagini, A.Sarti, S.Tubaro

Dipartimento di Elettronica ed Informazione,Politecnico di Milano
piazza Leonardo da Vinci 32
20133 Milano, Italy
[1] valente/tagliasa/antonacc/bestagini/sarti/tubaro@elet.polimi.it

***Abstract*—This paper proposes a method that solves the problem of geometric calibration of microphone arrays. We consider a distributed system, in which each array is controlled by separate acquisition devices that do not share a common synchronization clock. Given a set of probing sources, e.g. loudspeakers, each array computes an estimate of the source locations using a conventional TDOA-based algorithm. These observations are fused together by the proposed method, in order to estimate the position and pose of one array with respect to the other. Unlike previous approaches, we explicitly consider the anisotropic distribution of localization errors. As such, the proposed method is able to address the problem of geometric calibration when the probing sources are located both in the near- and far-field of the microphone arrays. Experimental results demonstrate that the improvement in terms of calibration accuracy with respect to state-of-the-art algorithms can be substantial, especially in the far-field .**

I. Introduction

Microphone arrays can be used for sampling the space-time structure of an acoustic field. Therefore, they are routinely adopted in the field of auditory field analysis, e.g. for acoustic source localization and tracking. In such applications, the deployed system might consist of several microphone arrays, which are distributed in the environment in order to achieve a better coverage of the area under analysis. In principle, all microphones of the various arrays might be thought of as composing a single array. This setup implicitly assumes that all signals are sampled in a synchronous manner with respect to a centralized clock. In this case, intra-array calibration techniques [1][2][3] can be used to determine the position of the individual microphones in space.

Unfortunately, such system setup can be costly with the current technology, due to the inherent difficulty of designing and manufacturing devices able to acquire more than 8-16 channels. Therefore, an alternative option consists of deploying distinct microphone arrays, each governed by its own acquisition device. In this case, the geometric calibration of the system proceeds according to two steps. First, intra-array calibration is performed, independently for each array, e.g. using one of the aforementioned methods. Second, inter-array calibration determines the location and the pose of each array with respect to a selected coordinate reference system.

MMSP'10, October 4-6, 2010, Saint-Malo, France. 978-1-4244-8112-5/10/$26.00 ©2010 IEEE.

The problem of inter-array calibration has been investigated in the recent literature. In [4] and [5] the authors have presented methods based on estimated Directions of Arrival (DOAs). In [4], DOAs are obtained from the analysis of acoustic maps generated using a delay and sum beamformer. Calibration is performed using well established computer vision techniques, based on sets of correspondences between pairs of acoustic maps. Each correspondence is obtained by activating an acoustic probe, e.g. a loudspeaker, in an arbitrary and unknown location in space. The method in [4] proves to be efficient only when probing sources are in the far-field of the microphone arrays, since acoustic maps do not convey information about their range.

The main contribution of this paper is a geometric calibration method that extends our previous work in [4]. Instead of exploiting measurements related to DOAs, the key tenet is to leverage, whenever possible, the estimated locations of the probing sources in the three-dimensional space. A first step in this direction has been undertaken in [6], where we solved the problem by estimating the rigid body transformation between pairs of microphone arrays. First, each array localizes a number of acoustic sources in space using a Global Coherence Field (GCF) approach [7]. Then, the rigid body transformation is computed based on a sufficient number of point correspondences, solving a least squares problem. The algorithm in [6] exhibits a complementary behavior with respect to that of [4], in the sense that it performs well when sources are in the near-field. This is due to the fact that GCF is not able to accurately estimate the location of acoustic sources in the far-field of the microphone array [7].

Given the availability of the methods in [4] and [6], one might decide to switch between them, depending on the problem at hand. Here, we propose an alternative method, whose solution seamlessly shifts between that of [4] in the far-field, and [6] in the near-field. More specifically, we inform the calibration algorithm about the uncertainty of the source location estimation. This is modeled in terms of the covariance matrix of the estimated location. Indeed, in the near-field, the

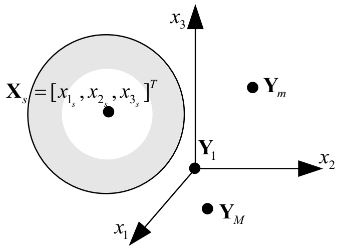


Fig. 1. Receivers are located in $\mathbf{Y}_m = [y_{1_m}, y_{2_m}, y_{3_m}]^T, m = 1, \dots, M$ and the source is in $\mathbf{X}_s = [x_{1_s}, x_{2_s}, x_{3_s}]^T$.

source location can be accurately estimated. Conversely, in the far-field, only the direction of arrival can be reliably estimated. We demonstrate by means of numerical simulations that the proposed method attains good results both in near- and far-field conditions.

The rest of the paper is organized as follows. Section II describes the algorithm used to localize acoustic sources in space. Section III illustrates the proposed geometric calibration method. Finally, Section IV describes simulation results, demonstrating the improvements related to the use of the proposed method.

II. Source localization

Similarly to [6], the proposed method requires the localization of a set of acoustic sources in the three-dimensional space. To this end, we adopt the spherical least squares algorithm [8], which is briefly summarized in this section.

Let $\mathbf{X}_s : [x_{1_s}, x_{2_s}, x_{3_s}]^T$ denote the location of an acoustic source and $\mathbf{Y}_m : [y_{1_m}, y_{2_m}, y_{3_m}]^T$, $m = 1, \dots, M$ the position of the microphones in an array, as illustrated in Figure 1. All microphones in the same array are controlled by the same acquisition device and share a synchronous clock. Let $\mathbf{Y}_1$ denote the position of the reference microphone, which indicates the origin of the coordinate system of the array. We compute the cross-correlation functions $R_m(k)$, $m = 2, \dots, M$, between the signals acquired by the reference microphone and the m-th microphone of the array. The time lag $\hat{k}_m$ corresponding to the peak of the cross-correlation provides an estimate of the Time Difference of Arrival (TDOA), expressed in number of samples, between the signals acquired by the two microphones.

Let Δ_m indicate the range difference, i.e. the difference of the distances between the source location $\mathbf{X}_s$ and the microphone positions $\mathbf{Y}_m$ and $\mathbf{Y}_1$

$$\Delta_m = ||\mathbf{Y}_m - \mathbf{X}_s|| - ||\mathbf{Y}_1 - \mathbf{X}_s|| \; . \tag{1}$$

TDOA measurements provide a noisy estimate $\hat{\Delta}_m$ of the true range difference Δ_m,

$$\hat{\Delta}_m = \hat{k}_m \frac{c}{F_s} \; , \tag{2}$$

where F_s is the sampling frequency and c is the sound speed. Estimated range differences are stacked together in the vector $\hat{\boldsymbol{\Delta}} = [\hat{\Delta}_2, \dots, \hat{\Delta}_M]^T$.

The spherical least squares error function for the m-th microphone is defined as the difference between the estimated and true distances from the source $\mathbf{X}_s$

$$e_m(\mathbf{X}_s) = \frac{1}{2}\left(\hat{d}_m^2 - ||\mathbf{Y}_m - \mathbf{X}_s||^2\right) \; , \tag{3}$$

where $\hat{d}_m = ||\mathbf{X}_s - \mathbf{Y}_1|| + \hat{\Delta}_m$.

The overall error function $\mathbf{e}(\mathbf{X}_s)$ considers all the error functions together. Based on [8], it is possible to write

$$\mathbf{e}(\mathbf{X}_s) = \mathbf{A}\boldsymbol{\theta} - \mathbf{b} \; , \tag{4}$$

where

$$\mathbf{A} = [\mathbf{S}|\hat{\boldsymbol{\Delta}}] \; ,$$

$$\mathbf{S} = [\mathbf{Y}_2, \dots, \mathbf{Y}_M]^T \; ,$$

$$\boldsymbol{\theta} = [\mathbf{X}_s^T, ||\mathbf{X}_s - \mathbf{Y}_1||]^T \; ,$$

$$\mathbf{b} = 0.5 \begin{bmatrix} ||\mathbf{Y}_2||^2 - \hat{d}_2^2 \\ \dots \\ ||\mathbf{Y}_M||^2 - \hat{d}_M^2 \end{bmatrix} \tag{5}$$

Solving for the unknown vector $\boldsymbol{\theta}$, we obtain

$$\hat{\boldsymbol{\theta}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b} \; , \tag{6}$$

Finally, the estimated source location $\hat{\mathbf{X}}_s$ is given by the first three elements of the vector $\hat{\boldsymbol{\theta}}$.

III. Calibration from source correspondences

In this section we formulate the problem of geometric calibration and we propose a solution based on a maximum likelihood estimation approach.

A. Problem formulation

We consider a system composed of two microphone arrays. Figure 2 depicts an exemplary configuration, which consists of two planar microphone arrays. The arrays are internally calibrated, i.e. the position of each microphone in the array is assumed to be known. For each array, a reference microphone is defined, and the local coordinate system is referred to it, as illustrated in Figure 2. Furthermore, we assume that the two arrays are internally synchronized, but mutually asynchronous, i.e. the clock is not shared among them.

The goal of geometric calibration is to determine the location and pose of the second array with respect to the first. Generalizing to multiple microphone arrays is straightforward, e.g. performing calibration of each array with respect to the first. Hereafter, we do not consider collaborative calibration, whereby the calibration problem is solved considering all the possible pairs of microphone arrays.

A probing acoustic source (e.g. a loudspeaker) is moved at S different positions in space. Each array localizes the source at, respectively, $\hat{\mathbf{X}}'_s$ and $\hat{\mathbf{X}}''_s$, $s = 1, \dots, S$, where the positions are referred to the local coordinate systems of each array. The relationship between the two coordinate systems can be expressed as a roto-translation

$$\mathbf{X}'' = \mathbf{R}(r, p, y)\mathbf{X}' + \mathbf{t} \tag{7}$$

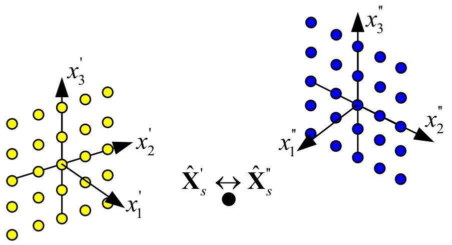


Fig. 2. An acoustic source is activated in a arbitrary position and is localized by two microphone arrays. The estimated positions of the source, as expressed in the two local coordinate systems, are given by $\hat{\mathbf{X}}'_s$ and $\hat{\mathbf{X}}''_s$. Note that local coordinate systems are centered in the reference microphones of each array.

where $\mathbf{R}(r,p,y)$ is a 3×3 rotation matrix, $\mathbf{t}$ is the translation vector that brings the origin of the coordinate system of the first array to the origin of the second array. More specifically, the rotation matrix $\mathbf{R}$ is expressed in terms of three parameters, r, p, y, representing, respectively, roll, pitch and yaw angles. Thus, the transformation in (7) has six degrees of freedom.

Considering localization errors, the estimated positions of the sources are related by the following expression

$$\hat{\mathbf{X}}''_s = \mathbf{R}(r,p,y)\hat{\mathbf{X}}'_s + \mathbf{t} + \nu_s, \quad l = 1, \dots, L \tag{8}$$

where ν_s is a random variable that models localization uncertainty. Hence, geometric calibration aims at estimating $\mathbf{R}(r,p,y)$ and $\mathbf{t}$, given a set of (noisy) correspondences $\hat{\mathbf{X}}'_s \leftrightarrow \hat{\mathbf{X}}''_s$, $l = 1, \dots, L$. This problem is common in the literature of computer vision and it is known as *calibrated reconstruction* [9].

A closed-form solution to this problem has been presented in [10], based on the Singular Value Decomposition (SVD) of a matrix computed from point correspondences. The work in [6] combines this method with RANSAC [11] in order to gain robustness in face of outliers that might arise when localizing acoustic sources, and it is used to initialize the iterative algorithm described in the next section.

B. Proposed solution

The method described in [6] is accurate when the probing sources are in the near-field of the microphone array, since robust point correspondences can be estimated. Conversely, inaccurate results are obtained when sources are in the far-field. This is due to the fact that TDOA-based localization algorithms do not accurately estimate the range of a source when its distance is much greater than the array baseline. Indeed, the algorithms presented in [4], [5] recognize this fact, and perform calibration based on DOA measurements only.

In the following, we show that we can achieve accurate calibration results both in the near-field and far-field conditions, by properly modeling the anisotropic characteristic of TDOA-based localization error. Let us consider that a probing source is localized by the first (second) array at $\hat{\mathbf{X}}'_s = \mathbf{X}'_s + \nu'_s$ ($\hat{\mathbf{X}}''_s = \mathbf{X}''_s + \nu''_s$), where $\mathbf{X}'_s$ ($\mathbf{X}''_s$) represents the true source location in the coordinate system of the first (second) array. The positions $\mathbf{X}'_s$ and $\mathbf{X}''_s$ are deterministically related by the expression (7). The term ν'_s (ν''_s) captures localization errors, that we model as a zero mean Gaussian random variable, i.e. $\nu'_s \sim N(\mathbf{0}, \mathbf{\Sigma}'_s)$ ($\nu''_s \sim N(\mathbf{0}, \mathbf{\Sigma}''_s)$. We assume here that the localization errors at the two arrays are independent, thus we can write the likelihood of observing a set of point correspondences $\hat{\mathbf{X}}'_s \leftrightarrow \hat{\mathbf{X}}''_s$, given the (unknown) parameters $\mathbf{X}'_s$, $\mathbf{\Sigma}'_s$, $\mathbf{\Sigma}''_s$, $\mathbf{R}$, $\mathbf{t}$, $l = 1, \dots, S$ as

$$\begin{aligned} L(\mathbf{X}'_1, \dots, \mathbf{X}'_S, \mathbf{\Sigma}'_1, \dots, \mathbf{\Sigma}'_S, \mathbf{\Sigma}''_1, \dots, \mathbf{\Sigma}''_S, \mathbf{R}, \mathbf{t}) = \\ = \prod_{l=1}^{S} p(\hat{\mathbf{X}}'_s; \mathbf{X}'_s, \mathbf{\Sigma}'_s) p(\hat{\mathbf{X}}''_s; \mathbf{X}'_s, \mathbf{\Sigma}''_s, \mathbf{R}, \mathbf{t}) \end{aligned} \tag{9}$$

where

$$p(\hat{\mathbf{X}}'_s; \mathbf{X}'_s, \mathbf{\Sigma}'_s) = N(\hat{\mathbf{X}}'_s; \mathbf{X}'_s, \mathbf{\Sigma}'_s) \tag{10}$$

$$\begin{aligned} p(\hat{\mathbf{X}}''_s; \mathbf{X}'_s, \mathbf{\Sigma}''_s, \mathbf{R}, \mathbf{t}) &= N(\hat{\mathbf{X}}''_s; \mathbf{X}''_s, \mathbf{\Sigma}''_s) \\ &= N(\hat{\mathbf{X}}''_s; \mathbf{R}\mathbf{X}'_s + \mathbf{t}, \mathbf{\Sigma}''_s) \end{aligned} \tag{11}$$

and

$$N(\mathbf{X}; \mu, \mathbf{\Sigma}) = \frac{1}{\sqrt{2\pi |\mathbf{\Sigma}|}} e^{\left[-\frac{1}{2}(\mathbf{X}-\mu)^T \mathbf{\Sigma}^{-1} (\mathbf{X}-\mu)\right]} \tag{12}$$

In principle, the estimated parameters can be found by maximizing the likelihood function in (9). In order to make the problem mathematically tractable, we perform an iterative procedure that alternates between the following two steps:

1) Fix $\mathbf{R}$ and $\mathbf{t}$, and find $\mathbf{X}'_s, \mathbf{\Sigma}'_s, \mathbf{\Sigma}''_s$.
2) Fix $\mathbf{X}'_s, \mathbf{\Sigma}'_s, \mathbf{\Sigma}''_s$ and find $\mathbf{R}, \mathbf{t}$.

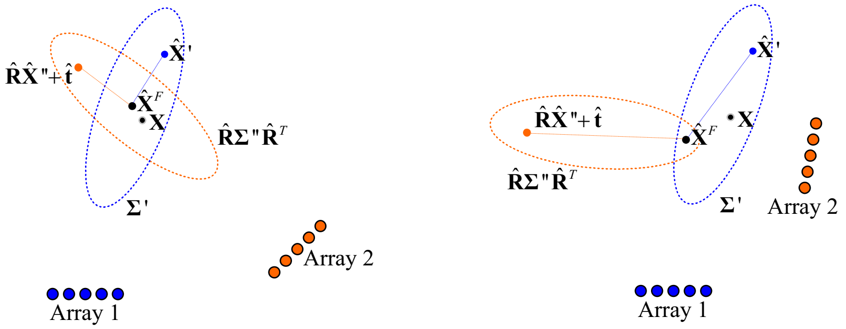
In order to initialize the iterative procedure, we set $\mathbf{R}$ and $\mathbf{t}$ equal to the result obtained with the method described in [6]. Then, the process is repeated until convergence, i.e. when the variation in the estimated values between two successive steps falls below a threshold value.

Step 1: The problem can be solved separately for each source $\mathbf{X}'_s$. Thus, we omit the subscript s in the remainder of the description of the step 1. In order to further simplify the problem solution, we recognize that the covariance matrix of the localization errors can be approximated by the inverse of the Fisher Information Matrix (FIM). The latter is determined based on: i) the geometry of the array; ii) the location of the source; iii) the noise that affects the TDOA measurements. More specifically, let $\bar{\mathbf{\Sigma}}'$ denote the Cramer-Rao lower bound (CRLB) on the covariance matrix of the localization error. That is, for any unbiased estimator of the source location $\mathbf{X}'$, let $\mathbf{\Sigma}'$ denote its covariance matrix, i.e. $\mathbf{\Sigma}' = E[(\hat{\mathbf{X}}' - \mathbf{X}')(\hat{\mathbf{X}}' - \mathbf{X}')]$. It holds that $\mathbf{\Sigma}' - \bar{\mathbf{\Sigma}}'$ is positive semidefinite. The matrix $\bar{\mathbf{\Sigma}}'$ can be found as the inverse of the Fisher Information Matrix (FIM). For TDOA-based estimators, the FIM is given by

$$\mathbf{I}(\mathbf{X}') = \frac{\mathbf{G}(\mathbf{X}')\mathbf{G}^T(\mathbf{X}')}{(\sigma c)^2}\ , \tag{13}$$

where σ is the standard deviation of the TDOA measurements (expressed in samples), and c is the sound speed. The elements of the $M \times M$ matrix $\mathbf{G}$ are defined as

$$\mathbf{g}_{ij}(\mathbf{X}') = \mathbf{g}_i(\mathbf{X}') - \mathbf{g}_j(\mathbf{X}')$$


$\hat{\mathbf{R}}\hat{\mathbf{X}}"+\hat{\mathbf{t}}$
$\hat{\mathbf{X}}'$
$\hat{\mathbf{X}}^F$
$\mathbf{X}$
$\hat{\mathbf{R}}\mathbf{\Sigma}"\hat{\mathbf{R}}^T$
$\mathbf{\Sigma}'$
Array 2
Array 1
$\hat{\mathbf{R}}\hat{\mathbf{X}}"+\hat{\mathbf{t}}$
$\hat{\mathbf{X}}'$
$\mathbf{X}$
$\hat{\mathbf{X}}^F$
$\hat{\mathbf{R}}\mathbf{\Sigma}"\hat{\mathbf{R}}^T$
$\mathbf{\Sigma}'$
Array 2
Array 1

(a) Correct Estimation of the Rotation Matrix

(b) Wrong Estimation of the Rotation Matrix

Fig. 3. Computation of the likelihood function in (17). a) Correct estimates of $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$. b) Wrong estimates of $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$. All values are expressed in the coordinate system of the first array.

$$\mathbf{g}_i(\mathbf{X}') = \frac{\mathbf{X}' - \mathbf{X}_i}{||\mathbf{X}' - \mathbf{X}_i||} \, .$$

where $\mathbf{X}_i, i = 1, \ldots, M$ denotes the locations of the microphones within the same array. Finally, $\bar{\mathbf{\Sigma}}'(\mathbf{X}') = \mathbf{I}^{-1}(\mathbf{X}')$.

In the proposed solution, we assume that the estimator is efficient, in the sense that it attains the CRLB, i.e. $\mathbf{\Sigma}'(\mathbf{X}') \simeq \bar{\mathbf{\Sigma}}'(\mathbf{X}')$. Thus, a chicken-and-egg problem arises, since $\mathbf{\Sigma}'$ depends on $\mathbf{X}'$ and, vice-versa, the estimation of $\mathbf{X}'$ depends on $\mathbf{\Sigma}'$. Therefore, we adopt an approximation of $\mathbf{\Sigma}'$ that can be obtained by expressing it as a function of $\hat{\mathbf{X}}'$, i.e. $\mathbf{\Sigma}'(\mathbf{X}') \simeq \bar{\mathbf{\Sigma}}'(\hat{\mathbf{X}}')$, and it is kept the same for all iterations. The same argument holds for $\mathbf{\Sigma}''$.

Hence, the estimated source location $\hat{\mathbf{X}}$ can be obtained by maximizing the likelihood function below

$$L(\mathbf{X}') = p(\hat{\mathbf{X}}'; \mathbf{X}', \mathbf{\Sigma}'(\hat{\mathbf{X}}'))p(\hat{\mathbf{X}}''; \mathbf{X}', \mathbf{\Sigma}''(\hat{\mathbf{X}}'), \hat{\mathbf{R}}, \hat{\mathbf{t}}) \quad (14)$$

where $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$ represent, respectively, the current estimate of the rotation matrix and translation vector. The solution can be found in closed-form [12] to be equal to

$$\hat{\mathbf{X}}^F = (\mathbf{\Sigma}'^{-1} + (\hat{\mathbf{R}}\mathbf{\Sigma}''\hat{\mathbf{R}}^T)^{-1})^{-1} \cdot \\ \cdot (\mathbf{\Sigma}'^{-1}\hat{\mathbf{X}}' + (\hat{\mathbf{R}}\mathbf{\Sigma}''\hat{\mathbf{R}}^T)^{-1}(\hat{\mathbf{R}}\hat{\mathbf{X}}'' + \hat{\mathbf{t}})) \quad (15)$$

where the superscript ^{F} refers to the fact that the estimate is obtained by fusing the observations of the two arrays. Note that the matrix multiplication $\hat{\mathbf{R}}\mathbf{\Sigma}''\hat{\mathbf{R}}^T$ is used to express the covariance matrix $\mathbf{\Sigma}''$ in the coordinate system of the first array. Similarly, $\hat{\mathbf{R}}\hat{\mathbf{X}}'' + \hat{\mathbf{t}}$ represents the position of the source as estimated by the second array mapped to the coordinate system of the first array.

Step 2: In the second step, we assume that the source locations are known and equal to $\hat{\mathbf{X}}^F_s$, $s = 1, \ldots, S$, together with the corresponding covariance matrices $\mathbf{\Sigma}'_s$ and $\mathbf{\Sigma}''_s$. The goal is to obtain estimates $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$ of the rotation matrix and the translation vector. To this end, we maximize the following likelihood function

$$L(\mathbf{R}, \mathbf{t}) = \prod_{l=1}^{L} p(\hat{\mathbf{X}}'_s; \hat{\mathbf{X}}^F_s, \mathbf{\Sigma}'_s)p(\hat{\mathbf{X}}''_s; \hat{\mathbf{X}}^F_s, \mathbf{\Sigma}''_s, \mathbf{R}, \mathbf{t}) \quad (16)$$

When the distribution of the localization noise is Gaussian, as expressed in (10) and (11), the parameters can be obtained solving the following optimization problem, which is equivalent to maximizing the logarithm of the likelihood function in (16)

$$(\hat{\mathbf{R}},\hat{\mathbf{t}}) = \arg.\ \min_{\mathbf{R},\mathbf{t}} \sum_{l=1}^{L} (\hat{\mathbf{X}}'_s - \hat{\mathbf{X}}^F_s)^T(\mathbf{\Sigma}'_s)^{-1}(\hat{\mathbf{X}}'_s - \hat{\mathbf{X}}^F_s) + \\ + (\mathbf{R}\hat{\mathbf{X}}''_s + \mathbf{t} - \hat{\mathbf{X}}^F_s)^T(\mathbf{R}\mathbf{\Sigma}''_s\mathbf{R}^T)^{-1}(\mathbf{R}\hat{\mathbf{X}}''_s + \mathbf{t} - \hat{\mathbf{X}}^F_s) \quad (17)$$

The objective function of (17) depends on $\mathbf{R}$ and $\mathbf{t}$ and it is characterized by several local minima. We solve the problem adopting nonlinear least squares. In order to initialize the solver, we use the values of $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$ estimated at the previous iteration. As for the first iteration, we leverage the estimate provided by the algorithm in [6].

Observations: Unlike the method in [6], problem (17) explicitly takes the anisotropic characteristic of localization errors into account, that arise especially when the probing sources are located in the far-field of the microphone arrays. This is illustrated in Figure III-B for an exemplary configuration and considering only one of the S probing sources. First, we show the case in which the proposed algorithm has reached convergence, and the estimated values of $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$ approximate the true values $\mathbf{R}$ and $\mathbf{t}$. In Figure III-B, both source locations and covariance matrices are expressed in the coordinate system of the first array. Each microphone array provides an estimate of the source location, i.e. $\hat{\mathbf{X}}'$ and $\hat{\mathbf{R}}\hat{\mathbf{X}}'' + \hat{\mathbf{t}}$, which are used to estimate $\hat{\mathbf{X}}^F$, according to (15). Then, for current estimated values $\hat{\mathbf{R}}$ and $\hat{\mathbf{t}}$, the likelihood of the estimated source locations is computed as the sum of the Mahalanobis distances between the location as estimated by each array and the estimated location $\hat{\mathbf{X}}^F$. The covariance matrices $\mathbf{\Sigma}'$ and $\hat{\mathbf{R}}\mathbf{\Sigma}''\hat{\mathbf{R}}^T$ are used to weigh the such distance measures, penalizing estimation errors related to the direction of arrival more than those related to the range. In Figure III-B we observe that, when the algorithm has not reached convergence, the Mahalanobis distance is significantly higher.

IV. Experiments

We carried out simulations in order to evaluate the accuracy of the proposed calibration method. We compared our method

(MLE-based) with other approaches previously presented in the literature, hereafter denoted DOA-based [4] and RBM-based [6] (Rigid Body Motion).

The system setup consisted of two arrays composed by 18 microphones disposed as two parallel planar arrays of 9 microphones each. The inter-microphone spacing in each plane was $0.35m$, while the offset between the two planes was $0.2m$, as shown in Figure 4.

The probing source signals were realizations of white noise in the frequency band $[0 - 10kHz]$. The sampling frequency was set equal to $F_s = 96kHz$, which is supported by most of commercial audio-cards. Each simulation was repeated $M = 40$ times in order to average over several realizations.

We adopted the same metrics as in [6] and [4], originally defined in [13] and repeated here for convenience:

$$\epsilon_{\mathbf{R}} = \arccos\left(\frac{\mathrm{tr}(\hat{\mathbf{R}}(r,p,y)^T\mathbf{R}) - 1}{2}\right), \tag{18}$$

$$\epsilon_{\mathbf{t}} = \arccos(\hat{\mathbf{t}}^T\mathbf{t}) \ . \tag{19}$$

In particular, $\epsilon_{\mathbf{R}}$ measures the rotation angle of the matrix $\hat{\mathbf{R}}(r,p,y)^T\mathbf{R}$ in exponential coordinates, while $\epsilon_{\mathbf{t}}$ is the angle between $\mathbf{t}$ and $\hat{\mathbf{t}}$.

We conducted three experiments in order to compare the performance of the tested methods. In each experiment, we let only one parameter vary, while keeping the others fixed.

1) **Distance between the two arrays (d)**: The second array was rotated by an angle $\pi/2$ around the vertical axis (x_3) with respect to the first array. In order to span both the near- and the far-field conditions, we let d vary in the range $1 - 9m$, according to the configuration shown in Figure 5. Thus, the translation vector was set to $\mathbf{t} = [d, d, 0]^T$, while the rotation matrix $\mathbf{R}$ was expressed in terms of the following roll, pitch and yaw angles $(r, p, y) = (0°, 0°, 90°)$. The probing sources were randomly located around the average position $[d, 0, 0]^T$, in order to be in front of both arrays. The number of sources for each repetition of the experiment was $S = 40$. We observe that when the sources were in the near-field the RBM- and MLE-based methods achieve similar accuracy, while the DOA-based method does not enable array calibration. On the other hand, when the distance d increases, the RBM-based method suffers from significant errors, while the accuracy of the proposed MLE-based method remains almost constant for different values of d.
2) **Number of sources (S)**: The geometry of this experiment was analogous to the previous one, and illustrated in Figure 5. In this case, we fixed the distance $d = 6m$ and let the number of sources S vary between 8 and 50. Even if at $6m$ we are in the far-field, at this distance both RBM-based and MLE-based methods exhibit an acceptable accuracy. Note that the minimum number of sources for the DOA-based method is 8, while for both RBM-based and MLE-based is 5. The results of the simulations are shown in Figure 7. We notice that the MLE-based method exhibited errors smaller than 20 degrees starting from 10 sources, while the DOA-based attained large errors especially when only few sources are available.

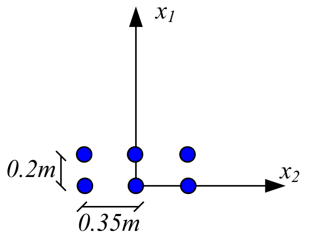


Fig. 4. Top view of the geometry of the arrays

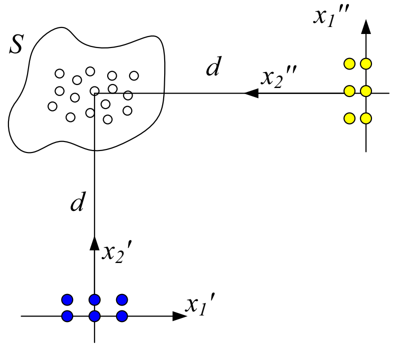


Fig. 5. Top view of the geometry for the first and second experiment: the arrays are rotated by a yaw angle of $\pi/2$ and are at a distance d from the probing sources.

3) **Yaw angle (y)**: The geometry of this experiment is shown in Figure 8. Both arrays were at distance $d = 6m$ from the probing sources and the yaw angle between the first and the second array is varied in the range $[-180°, -20°]$. For each angle, 40 realizations of the experiment were executed, each with $S = 20$ sources. Rotation and translation errors are shown in Figure 9. These results confirm that the accuracy of the MLE-based algorithm overcame that of RBM- and DOA-based methods. In particular we notice that the MLE-based calibration achieved good accuracy for a wider range of angles.

V. Conclusions

In this paper we proposed a method based on maximum likelihood estimation for the geometric calibration of microphone arrays. A set of probing sources are activated in front of the microphone arrays, in order to find correspondences between them. Unlike previous methods, the anisotropic characteristic of localization errors is explicitly taken into account. Simulation results confirm that the proposed methods improves state-of-the-art techniques as it is able to achieve good calibration accuracy in both near- and far-field conditions. As future work, we are extending the algorithm in such a way to exploit reverberations that arise in indoor environments.

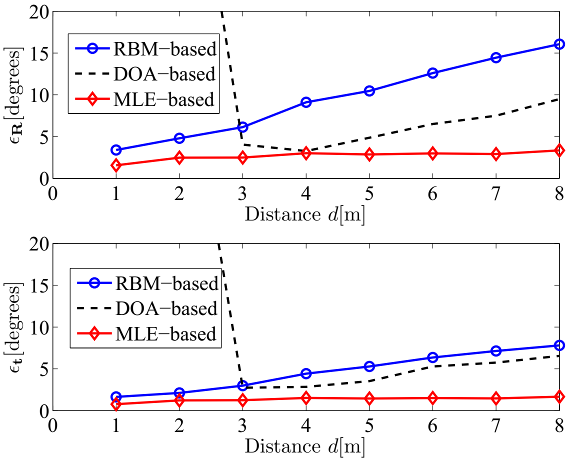


Fig. 6. Rotation and translation calibration errors as a function of the distance d, for the geometric configuration in Figure 5.

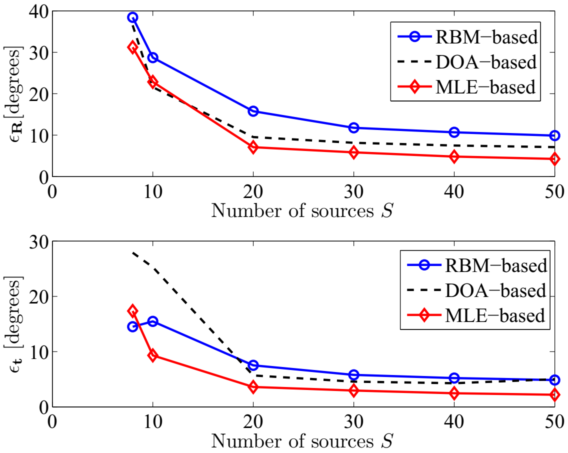


Fig. 7. Rotation and translation calibration errors as a function of the number S of sources, for the configuration in Figure 5, when $d = 6m$.

Indeed, we argue that it is possible to reduce the number of probing sources necessary to achieve a target calibration accuracy.

References

[1] A. Weiss and B. Friedlander, "Array shape calibration using sources in unknown locations - a maximum likelihood approach," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 1958–1966, Dec. 1989.

[2] S. T. Birchfield and A. Subramanya, "Microphone array position calibration by basis-point classical multidimensional scaling," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1025–1034, Sep. 2005.

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

[4] A. Redondi, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Geometric calibration of distributed microphone arrays," in *Proc. of IEEE International Workshop on Multimedia Signal Processing*, 2009, pp. 1–5.

[5] M. Hennecke, T. Plotz, G. Fink, J. Schmalenstroer, and R. Hab-Umbach, "A hierarchical approach to unsupervised shape calibration of microphone array networks," in *IEEE/SP 15th Workshop on Statistical Signal Processing, 2009, SSP '09*, Sept. 2009, pp. 257–260.

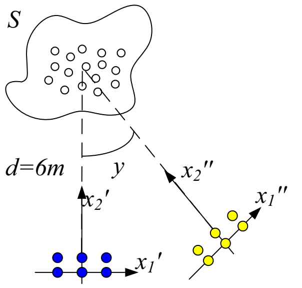


Fig. 8. Top view of the geometry of the third experiment: the sources are located at an average distance of $6m$ from the arrays and the yaw angle α between the two arrays is varied from 20° to 180°.

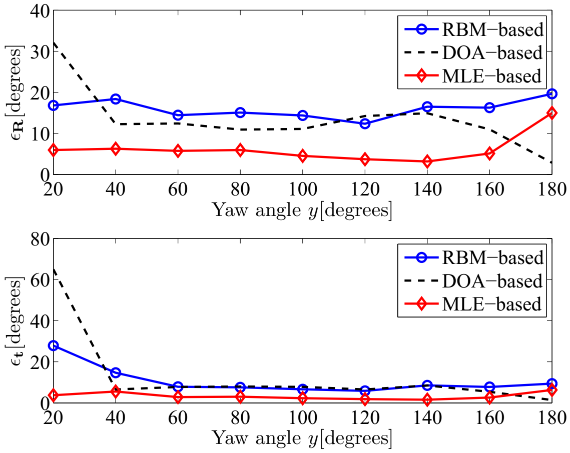


Fig. 9. Rotation and translation calibration errors as a function of the yaw angle y, for the configuration in Figure 8. The number of probing sources is $S = 20$.

[6] S. Valente, F. Antonacci, M. Tagliasacchi, A. Sarti, and S. Tubaro, "Selfcalibration of two microphone arrays from volumetric acoustic maps in nonreverberant rooms," in *proc. of International Symposium on Communications Control and Signal Processing (ISCCSP 2010)*, Mar. 2010.

[7] M. Omologo and P. Svaizer, "Use of the crosspower-spectrum phase in acoustic event location," *IEEE Transactions on Speech and Audio Processing*, vol. 5, pp. 288–292, 1993.

[8] Y.Huang, J.Benesty, and G.Elko, *Audio Signal Processing for Next Multimedia Communication Systems*, Y.Huang and J.Benesty, Eds. Kluwer Academic Publishers, 2004.

[9] R. Hartley and A. Zisserman, *Multiple View Geometry in computer vision*, 2nd ed. Cambridge Univ. Press, 2003.

[10] D. Eggert, A. Lorusso, and R. Fisher, "Estimating 3-d rigid body transformations: a comparison of four major algorithms," *Machine Vision and Applications*, vol. 9, no. 5-6, pp. 272–290, 1997.

[11] 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, pp. 381–395, June 1981.

[12] G. Prandi, G. Valenzise, M. Tagliasacchi, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization by fusing distributed microphone arrays measurements," in *proc. of 16th European Signal Processing Conference, EUSIPCO*, 2010.

[13] G. Chesi, "Camera displacement via constrained minimization of the algebraic error," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 370–375, Feb. 2009.