A Robust and Low-Complexity Source Localization Algorithm for Asynchronous Distributed Microphone Networks

A. Canclini, P. Bestagini, F. Antonacci, *Member, IEEE*, M. Compagnoni, A. Sarti, *Member, IEEE*, and S. Tubaro, *Senior Member, IEEE*

Abstract—In this paper, we propose a robust and low-complexity acoustic source localization technique based on time differences of arrival (TDOA), which addresses the scenario of distributed sensor networks in 3D environments. Network nodes are assumed to be unsynchronized, i.e., TDOAs between microphones belonging to different nodes are not available. We begin with showing how to select feasible TDOAs for each sensor node, exploiting both geometrical considerations and a characterization of the overall generalized cross correlation (GCC) shape. We then show how to localize sources in the space-range reference frame, where TDOA measurements have a clear geometrical interpretation that can be fruitfully used in the scenario of unsynchronized sensors. In this framework, in fact, the source corresponds to the apex of a hypercone passing through points described by the sole microphone positions and TDOA measurements. The localization problem is therefore approached as a hypercone fitting problem. Finally, in order to improve the robustness of the estimate, we include an outlier detection procedure based on the evaluation of the hypercone fitting residuals. A refinement of source location estimate is then performed ignoring the contributions coming from outlier measurements. A set of simulations shows the performance of individual blocks of the system, with particular focus on the effect of TDOA selection on source localization and refinement steps. Experiments on real data validate the localization algorithm in an everyday scenario, proving that good accuracy can be obtained while saving computational cost in comparison with state-of-the-art techniques.

Index Terms—Acoustic source localization, distributed acoustic sensor networks, space-range reference frame, time difference of arrival (TDOA) selection.

I. INTRODUCTION

THANKS to the increasing availability of inexpensive small-scale hardware devices with wireless communication capabilities, distributed sensor networks are steadily gaining popularity in a wide range of application scenarios, including security, environmental monitoring and elderly care. Acoustic source localization is a feature that is often required

in these scenarios [1], as it is used, for example, for automatically pointing a video camera towards the active speaker in a tele-conference [2], [3], [4], [5]; to detect and localize intrusions in areas where vision-based systems area likely to fail due to poor illumination or occlusions [6], [7], [8]; to monitor the environment [9], [10].

In this manuscript we propose a technique for the 3D localization of acoustic sources using a distributed network of compact microphone arrays. Numerous algorithms have been developed for the localization of acoustic sources. Among the most popular of them are those based on the local (at node level) computation of the Generalized Cross Correlation (GCC) [11] (or modified versions thereof [12], [13]). These techniques are roughly split into two categories: i) those based on Steered Response Power (SRP) [14], [15]; and ii) those based Time Difference Of Arrival (TDOA) [16], [17], [18], [19]. In its original formulation, SRP localizes sources through an inverse mapping of the GCC samples onto candidate source locations. Building the GCC map and searching for its maximum, however, is a computationally intensive task if approached in a brute-force fashion. This is why the following proposed improvements focused on the reduction of the computational cost [20], [21], [22]. In a distributed sensor network scenario, however, the fact that each microphone array has to transmit the whole GCC to the central processing node, is an inherent disadvantage because it has an evident impact on bandwidth requirements.

Methods based on TDOAs aim at localizing the source from the sole knowledge of the propagation delay of the acoustic wavefront at different microphone pairs. The TDOA is estimated at each microphone pair as the lag corresponding to the maximum of the GCC computed on signals acquired by the considered pair. In a typical scenario, TDOAs are computed between a reference microphone and all the remaining ones. Different methodologies of this sort have been proposed, based on both Maximum Likelihood Estimation [23], [24], [25], [26], [27] and Least Squares method [16]–[18], [28]–[31]. The main advantage of these techniques is their modest computational cost, as they do not require either building an inverse map or performing an exhaustive search. Instead, they typically localize the source through the minimization of a cost function, and the most computationally efficient approaches are those based on a closed-form optimization (e.g., [29], [30]). Furthermore, bandwidth requirements turn out to be rather modest, because what is being transmitted over the network are TDOA values only. It is well-known in the literature, however, that TDOA-based

methods tend to be sensitive to reverberation and background noise [32]. Moreover, most of the referenced techniques assume that all the microphones are synchronized, whereas in a distributed sensor network scenario only microphones in the same node are synchronized.

In this manuscript we propose a TDOA-based methodology for the localization of acoustic sources in moderately reverberant environments, which is suitable for distributed sensor networks. Two are the main challenges that we tackle, while keeping computational cost at bay: i) robust TDOA estimation in reverberant environments; ii) source localization using TDOAs coming from unsynchronized microphone arrays. As far as the first challenge is concerned, reverberation severely impacts on the GCC generating multiple spurious peaks, among which the correct TDOA must be selected. In this manuscript, robustness against reverberation is obtained through a method inspired by the DATEMM (DisAmbiguation of TDOA measurement in Multisource MultiPath) algorithm [33], [34], [35], [36], [37]. In its original formulation, DATEMM selects the TDOAs to be used for localization from a number of candidate peaks in the GCC exploiting geometric constraints. In this manuscript, we rethink the DATEMM algorithm in order to render it suitable for the scenario of distributed networks of small microphone arrays, where the reduction of computational complexity is of utmost importance. In particular, from DATEMM we inherit the zero-sum condition (TDOAs over closed-loop microphone paths must add to zero) and propose a two-step procedure: the zero-sum condition on cyclic paths is first used for narrowing down the set of GCC peaks; we then define and use criteria based on the overall shape of the GCC and on the magnitude of the peaks of the candidate TDOAs to select the final TDOA estimates. This second step is partially inspired by previous works [38], [39], [40], [41]. Notice that one key feature of this approach is that TDOA selection can be performed locally (at node level), and many operations can be implemented in the form of simple sums and lookups on precomputed tables. The impact on the computational cost is therefore very modest compared to the original DATEMM algorithm.

In order to perform source localization, we start from the space-range method introduced in [19] for 2D geometries and further investigated in [42], and re-formulate it to render it suitable for 3D localization. We will see that, in the 3D geometric space, TDOA measurements and microphone positions define the surface of a hyper-cone, whose apex corresponds to the source location. Therefore, localizing the source means finding the hyper-cone that best fits the measurements through the minimization of a cost function. As done in [19], we show that this representation can be fruitfully used to accommodate the case of unsynchronized sensor nodes. TDOAs coming from unsynchronized sensors, in fact, suffer from a shift along the range axis (which is only one of the dimensions of the space-range frame). The cost function can therefore be defined to accommodate this shift, and enable simultaneous source localization and realignment of measurements. In this work, however, the cost function is modified to account for all the possible TDOAs within each sensor node. Therefore, with respect to [19], we relax the constraint of having to work with a reference microphone for each node.

In order to further improve the localization accuracy, we include an *a-posteriori* refinement step, which is based on the evaluation of the residual of the hyper-cone fitting cost function. If the residual exceeds a prescribed threshold, we search for the sensor nodes that are responsible for corrupted TDOAs, and remove their contribution to the cost function. Localization is then recomputed without such contribution. Simulations and experimental results show that this refinement procedure greatly improves the localization accuracy.

This manuscript is organized as follows. Section II provides a brief overview of some state-of-the-art techniques, formulates the problem and introduces the notation used throughout the manuscript. Section III explains the overall localization procedure detailing each step. An analysis of the computational complexity is given in Section IV. Section V describes some experiments and compares localization accuracy and computational cost with a well-known state-of-the-art technique. Finally, Section VI draws some conclusions.

II. PROBLEM FORMULATION AND BACKGROUND

In this manuscript we discuss the problem of acoustic source localization using Distributed Sensor Networks (DSN) in the presence of reverberations. Our goal is to achieve state-of-the-art localization accuracy and robustness with a significant reduction of the computational complexity.

With reference to Fig. 1, let us consider a spatial distribution of L independent and unsynchronized nodes. Each node is equipped with a compact microphone array that accommodates $N^{(l)}$ microphones, $l = 1, \ldots, L$. Let $\mathbf{m}_i^{(l)} = [x_i^{(l)}, y_i^{(l)}, z_i^{(l)}]^T$, $i = 1, \ldots, N^{(l)}$ be the 3D coordinates of the ith microphone of the lth array. We assume that only the microphones that belong to the same array (node) are synchronized. This means that we can only estimate TDOAs of microphone pairs that belong to the same array, i.e., only the TDOAs $\tilde{\tau}_{ij}^{(l)}$ computed between the ith and jth microphones of the same node l are available. The source is assumed to be at the unknown location $\mathbf{x}_S = [x_S, y_S, z_S]^T$. With no loss of generality we center the reference system at the first microphone of the first array, i.e., $\mathbf{m}_1^{(1)} = [0, 0, 0]^T$. With reference to this scenario, in this Section we briefly outline some state-of-the-art methods for TDOA extraction in challenging situations. We then introduce the framework for source localization from TDOAs, pointing out the problems that need to be addressed for distributed sensor networks.

A. TDOA measurements in adverse environments

The TDOA measures the difference between the times of arrival of an acoustic wavefront at two given microphones. In ideal conditions (no reverberation or any other forms of interferences) TDOAs are quite straightforwardly estimated using the Generalized Cross Correlation (GCC). Given the time-discrete signal $s_i^{(l)}(n)$ acquired by the ith microphone in the lth array, the GCC between the ith and jth microphone of the same array is computed as

$$R_{ij}^{(l)}(n) = \mathcal{F}^{-1} \left\{ \frac{\mathcal{F}[s_i^{(l)}(n)]\mathcal{F}[s_j^{(l)}(n)]^*}{\left| \mathcal{F}[s_i^{(l)}(n)]\mathcal{F}[s_j^{(l)}(n)]^* \right|} \right\}, \quad (1)$$
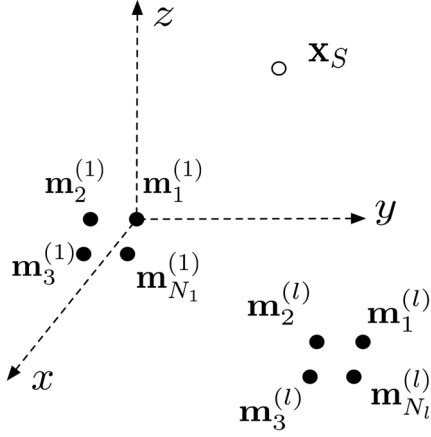
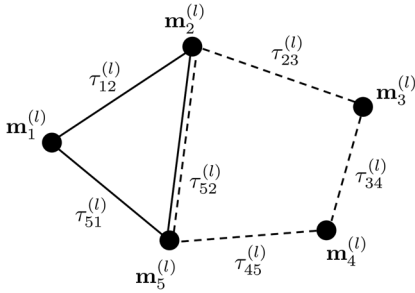Fig. 1. Setup of the distributed acoustic sensor system.

Fig. 2. Graph representation of a microphone array. Each graph node is a microphone, while each branch is a TDOA. Dashed and solid lines represent two different closed circuits. The sum of signed TDOAs over closed circuits must be zero.

where $\mathcal{F}$ is the Discrete Fourier Transform (DFT), $\mathcal{F}^{-1}$ is its inverse and the superscript $*$ denotes complex conjugation. The discrete-time TDOA estimate (expressed in samples) corresponds to the time lag $\widetilde{n}_{ij}^{(l)}$ that maximizes the GCC. The time-domain TDOA is given by $\widetilde{\tau}_{ij}^{(l)} = \widetilde{n}_{ij}^{(l)}/F_s$, where F_s is the sampling frequency.

In the presence of reverberations, the GCC presents multiple secondary peaks, the magnitude of which could be even larger than that of the direct path of the target source. Even in such adverse conditions, however, the GCC is still informative, as peaks related to reverberation can be removed using some geometrical and mathematical considerations based on signal propagation. A popular algorithm developed for this purpose is DATEMM [33], [34], which exploits two criteria: the raster condition and the zero-sum condition. The former compares the autocorrelation with the GCCs for discarding the GCC peaks generated by reflections. The latter aims at matching peaks in the GCC to each source that is present in the environment. The zero-sum condition exploits the fact that the sum of TDOAs over closed paths of three or more microphones (i.e., paths that begin and end at the same microphone) is bound to be zero (see Fig. 2). After enabling the selection of feasible paths, DATEMM aggregates further compatible microphone paths in order to extend the set of available measurements.

Implementing the raster and the zero-sum conditions could be too demanding for a DSN. In order to reduce the computational cost, in [35] the authors show that the use of the sole zero-sum
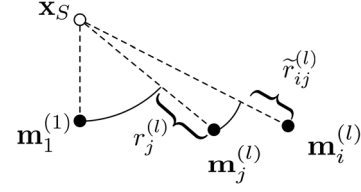
Fig. 3. Relation of the range differences measured between microphones $\mathbf{m}_i^{(l)}$ and $\mathbf{m}_j^{(l)}$ in the lth array and the global reference microphone $\mathbf{m}_1^{(1)}$.

condition is, in fact, sufficient for discarding wrong TDOAs. The goal of [34], [35], however, is to collect the largest number of correct TDOAs without worrying about outliers. This, in a scenario like ours, would cause the TDOA selection process to become overwhelming. Moreover, as shown in Section V-A, just a few correct TDOAs are sufficient to accurately localize the source but the presence of even a single outlier in the TDOA set (e.g., a TDOA related to a reflected path) would severely affect the localization accuracy. This is why the selection of accurate TDOAs is very important to us, and we base that upon criteria related to the quality of selected measurements, which are inspired by [39], [41] and are based on the overall shape of the GCC from which the TDOAs have been extracted.

B. Localization from unsynchronized TDOAs

In [19], we proposed an algorithm for source localization that envisions the collaboration between unsynchronized arrays. In that setup, however, a reference microphone for each array was defined. In this manuscript we remove that condition, and define a single global reference microphone.

With no loss of generality, we set the origin of the time axis on the global reference microphone $\mathbf{m}_1^{(1)}$. Moreover, TDOA measurements $\widetilde{\tau}_{ij}^{(l)}$ are converted to range differences $\widetilde{r}_{ij}^{(l)} = \eta \cdot \widetilde{\tau}_{ij}^{(l)}$, by simply scaling each value according to the propagation speed of sound η. If we define with r the range difference between a microphone located in a generic point $\mathbf{x} = [x, y, z]^T$ and the reference one, we can write that

$$\|\mathbf{x}_S - \mathbf{x}\| - \|\mathbf{x}_S - \mathbf{m}_1^{(1)}\| = r. \tag{2}$$

By inserting $r_S = -\|\mathbf{x}_S - \mathbf{m}_1^{(1)}\|$ into (2), after expanding and rearranging the terms we obtain

$$(x - x_S)^2 + (y - y_S)^2 + (z - z_S)^2 = (r - r_S)^2. \tag{3}$$

Eq. (3) describes a hyper-cone in the space-range reference domain (x, y, z, r), with apex in the source location (x_S, y_S, z_S, r_S). For the ith microphone in the lth array we can write that

$$(x_i^{(l)} - x_S)^2 + (y_i^{(l)} - y_S)^2 + (z_i^{(l)} - z_S)^2 = (r_i^{(l)} - r_S)^2. \tag{4}$$

The term $r_i^{(l)}$ represents the range difference between the ith microphone in the lth array and the global reference microphone $\mathbf{m}_1^{(1)}$, i.e., $r_i^{(l)} = \|\mathbf{x}_S - \mathbf{m}_i^{(l)}\| - \|\mathbf{x}_S - \mathbf{m}_1^{(1)}\|$. Due to the lack of synchronization, $r_i^{(l)}$ cannot be directly measured. However, as shown in Fig. 3, it is possible to relate $r_i^{(l)}$ to the range difference
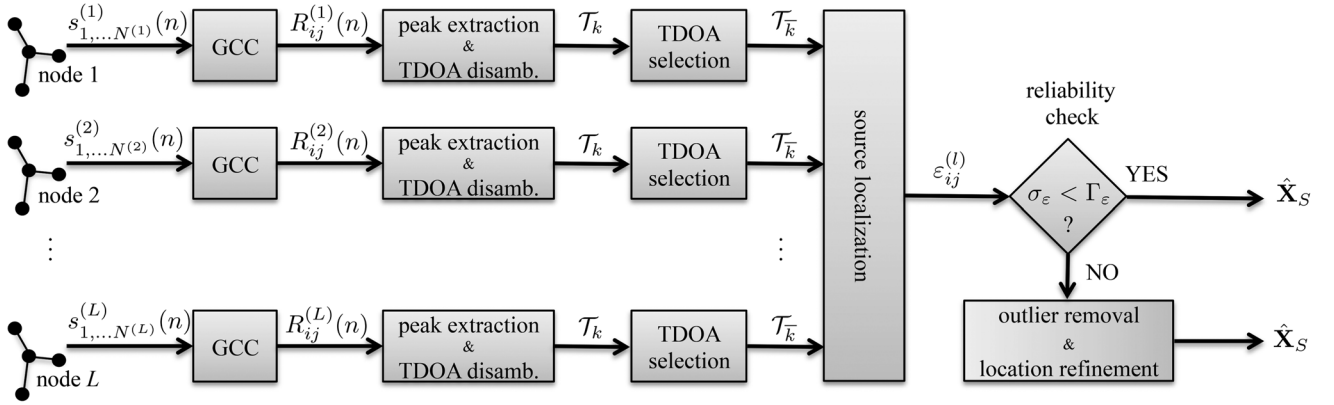
Fig. 4. Overall system block diagram.

$\widetilde{r}_{ij}^{(l)} = \|\mathbf{x}_S - \mathbf{m}_i^{(l)}\| - \|\mathbf{x}_S - \mathbf{m}_j^{(l)}\|$ measured between the pair of microphones $\mathbf{m}_i^{(l)}$ and $\mathbf{m}_j^{(l)}$. In particular, we can write

$$
\begin{aligned}
r_i^{(l)} &= \|\mathbf{x}_S - \mathbf{m}_i^{(l)}\| - \|\mathbf{x}_S - \mathbf{m}_j^{(l)}\| + \\
&\quad + \|\mathbf{x}_S - \mathbf{m}_j^{(l)}\| - \|\mathbf{x}_S - \mathbf{m}_1^{(1)}\| = \\
&= \widetilde{r}_{ij}^{(l)} + r_j^{(l)},
\end{aligned}
\tag{5}
$$

where

$$
r_j^{(l)} = \|\mathbf{x}_S - \mathbf{m}_j^{(l)}\| - \|\mathbf{x}_S - \mathbf{m}_1^{(1)}\|.
\tag{6}
$$

By using (5) in (4) we finally obtain

$$
\begin{aligned}
(x_i^{(l)} - x_S)^2 + (y_i^{(l)} - y_S)^2 + (z_i^{(l)} - z_S)^2 &= \\
= (\widetilde{r}_{ij}^{(l)} + \|\mathbf{x}_S - \mathbf{m}_j^{(l)}\| - \|\mathbf{x}_S - \mathbf{m}_1^{(1)}\| - r_S)^2.
\end{aligned}
\tag{7}
$$

In doing so, we do not need to explicitly estimate the range offsets (as done, e.g., in [43]), since (6) constrains $r_j^{(l)}$ to $\mathbf{x}_S$. Therefore, in (7) the only unknowns are x_S, y_S, z_S and r_S, i.e., the coordinates of the source in the space-range reference frame. Notice also that (5) assumes the knowledge of the microphone locations in a common reference frame. Several algorithms are available in the literature for the self-localization of arrays [44]–[47]. This problem, however, goes beyond the scope of this manuscript. We, therefore, assume the arrays to be calibrated in advance.

As we can see, the 3D acoustic source localization problem has now become that of finding the hyper-cone that best fits the data $[x_i^{(l)}, y_i^{(l)}, z_i^{(l)}, \widetilde{r}_{ij}^{(l)}]^T$, $i = 1, \ldots, N^{(l)}$, $l = 1, \ldots, L$, where $\widetilde{r}_{ij}^{(l)}$ denote the measured range differences. The first three coordinates of the hyper-cone apex $\hat{\mathbf{x}}_S = [\hat{x}_S, \hat{y}_S, \hat{z}_S]^T$ yield the estimate of the source position $\mathbf{x}_S$.

III. SOURCE LOCALIZATION SYSTEM

The acoustic source localization system that we propose is based on the scheme of Fig. 4. As a first step, TDOAs are estimated from the acquired signals. In order to do so, GCCs are computed for all the microphone pairs of the same node. Multiple TDOA candidates $\mathcal{T}_k$ are estimated from the peaks of the GCCs. Then, a TDOA disambiguation system is applied to select physically acceptable TDOA combinations $\mathcal{T}_{\overline{k}}$ that are referred to the same source. At this point, 3D source location

estimate $\hat{\mathbf{X}}_S$ are computed from the selected TDOAs through the minimization of a suitable cost function. The residuals of the cost function $\varepsilon_{ij}^{(l)}$ are used for computing a reliability index related to the estimated source location. If this index exceeds a prescribed threshold Γ_ε, the TDOAs that pertain the sensor node that introduces the highest residuals are marked as outliers. This check is repeated iteratively until the reliability index goes below the threshold or a minimum number of sensor is reached. Eventually, the outliers are removed from the set of TDOAs, and source localization is repeated in order to obtain a refined estimate. In the following we provide a detailed description of each algorithm's step and analyze the system computational complexity.

A. Computation of the TDOAs

The first step is the computation of the GCC for all the "intra-node" microphone pairs, for a total of $\frac{N^{(l)}(N^{(l)}-1)}{2}$ GCCs per array. In the DSN scenario each node is usually equipped with just a few microphones (in our experiments $N^{(l)} = 4$, $\forall l$). As TDOAs are independent from one node to another, for notational compactness, in this subsection we will drop the index l and focus on a single array.

In real environments GCCs exhibit many peaks (as mentioned in Section II), which complicates the selection of the correct peak. In order to solve this problem, we apply a disambiguation method that aims at selecting a small trusted set of TDOAs belonging to a single active source for each array. This is done by first selecting feasible TDOA sets, then ranking them and selecting the best one.

TDOAs disambiguation. Disambiguation begins with the selection of the TDOAs associated with the three largest peaks of each GCC, as proposed in [35]. More formally, considering the GCC $R_{ij}(n)$ between microphones i and j, we select the TDOA candidates $\widetilde{\tau}_{ij}^o = \widetilde{n}_{ij}^o / F_s$, $o = 1, 2, 3$, where $\widetilde{n}_{ij}^o$ is the lag associated to the oth peak in $R_{ij}(n)$. Considering all the closed paths of three or four microphones, we build all the K possible sets of three or four TDOA candidates. Denoting with $\mathcal{T}_k$ the kth candidate set, we check the zero-sum condition for each of them. The choice of checking the zero-sum condition only on paths of three or four microphone has a twofold reason: the former being that sensors are typically equipped with a small number of microphones; the latter is that, as shown in Section V, an accurate localization is possible with just four microphones

per node. Theoretically, for each feasible set $\mathcal{T}_k$, the zero-sum condition

$$z_k = \sum_{\widetilde{\tau}_{ij}^o \in \mathcal{T}_k} \widetilde{\tau}_{ij}^o = 0$$

should be exactly verified. In practice, due to noisy TDOAs, z_k generally differs from zero. For this reason, in order to preliminarily purge unfeasible measurements, we discard all the sets $\mathcal{T}_k$ such that $|z_k| > \Gamma_z$, Γ_z being a tolerance threshold ($\Gamma_z = 3$ samples in our experiments). After this step, we are left with just a few sets of physically feasible TDOAs $\mathcal{T}_k$.

Selection of TDOAs. The zero-sum condition is necessary but not sufficient for determining the TDOAs that are associated to direct path from the acoustic source [34]. Consequently, in order to select the set $\mathcal{T}_k$ that most likely contains TDOAs belonging to the real source, we cannot rely only on TDOA disambiguation. Instead, we compute three additional quality measures associated to each $\mathcal{T}_k$. The rationale behind the selected quality metrics is that a TDOA corresponding to a direct path usually corresponds to the largest peak of the GCC [34]. Moreover, as shown in [40], TDOAs coming from large GCC peaks with are more reliable than others. The first quality metric is the average of normalized GCC peaks associated to TDOAs in set $\mathcal{T}_k$:

$$q_k^{(1)} = \frac{1}{|\mathcal{T}_k|} \sum_{\substack{i,j,o: \\ \exists \widetilde{\tau}_{ij}^o \in \mathcal{T}_k}} \frac{R_{ij}(\widetilde{n}_{ij}^o)}{\max[R_{ij}(n)]}, \qquad (8)$$

where $|\mathcal{T}_k|$ is the cardinality of the set $\mathcal{T}_k$ (i.e., $|\mathcal{T}_k| \in \{3, 4\}$ depending on the considered closed-path length) and $\widetilde{n}_{ij}^o$ is the lag associated to TDOA $\widetilde{\tau}_{ij}^o$. What $q_k^{(1)} \in [0, 1]$ describes is the magnitude of peaks of TDOAs in the set $\mathcal{T}_k$. The second quality metric is the average between GCC maxima

$$q_k^{(2)} = \frac{1}{|\mathcal{T}_k|} \sum_{\substack{i,j,o: \\ \exists \widetilde{\tau}_{ij}^o \in \mathcal{T}_k}} \max[R_{ij}(n)], \qquad (9)$$

which is closely related to $q_k^{(1)}$. This quantity rates the GCCs that originate TDOAs, instead of rating TDOAs. The third quality measure is defined as the product of all ratios between first and second peak, over all the GCCs used for generating the TDOAs in the set $\mathcal{T}_k$:

$$q_k^{(3)} = \prod_{\substack{i,j,o: \\ \exists \widetilde{\tau}_{ij}^o \in \mathcal{T}_k}} \frac{\max[R_{ij}(n)]}{\max_2[R_{ij}(n)]}, \qquad (10)$$

where $\max_2[\cdot]$ extracts the second maximum value. The first-to-second peak ratio is proved to be an effective measure of the quality of the GCC [40]. Notice that the product is extended to three or four values, depending on the considered closed-circuit length. This results in a different weighting for circuits of different length.

In order to associate a global quality measure to each $\mathcal{T}_k$, we use a linear combination of the above quality values:

$$Q_k = \sum_{p=1}^{3} \lambda_p q_k^{(p)}, \qquad (11)$$

where λ_p is the weight of the quality measure $q_k^{(p)}$, which are to be experimentally determined and tuned to fit a specific environment. Eventually, we find $\overline{k} = \arg\max_k Q_k$, which denotes the closed TDOA path $\mathcal{T}_{\overline{k}}$ with the highest quality score.

B. Source Localization

Consider a source located at $\mathbf{x}_S = [x_S, y_S, z_S]^T$. If all range differences were noiseless, the hyper-cone equation (4) would be exactly satisfied. However, finite sampling frequency, finite bandwidth of the source and noise in the acquisition circuitry produce inaccuracies in the measurement chain. For this reason we define the Hyper-Cone Equation (HCE) metric as

$$\varepsilon_{ij}^{(l)}(\mathbf{X}) = \left(x_i^{(l)} - x\right)^2 + \left(y_i^{(l)} - y\right)^2 + \left(z_i^{(l)} - z\right)^2 + \\ -\left(\widetilde{r}_{ij}^{(l)} + r_j^{(l)} - r\right)^2, \qquad (12)$$

which measures how well (4) is satisfied by the point $\mathbf{X} = [\mathbf{x}^T, r]^T = [x, y, z, r]^T$ in the space-range domain. Here, according to (6), $r_j^{(l)} = \|\mathbf{x} - \mathbf{m}_j^{(l)}\| - \|\mathbf{x} - \mathbf{m}_1^{(1)}\|$. The corresponding cost function is defined as

$$J(\mathbf{X}) = \sum_{l=1}^{L} \sum_{i=1}^{N^{(l)}} \sum_{j=1}^{N^{(l)}} P_{ij}^{(l)} \left[\varepsilon_{ij}^{(l)}(\mathbf{X})\right]^2, \qquad (13)$$

where $P_{ij}^{(l)}$ is an indicator function defined as

$$P_{ij}^{(l)} = \begin{cases} 1 & \text{if } \tau_{ij}^{(l)} \in \overline{\mathcal{T}}^{(l)} \\ 0 & \text{otherwise} \end{cases}. \qquad (14)$$

The set $\overline{\mathcal{T}}^{(l)}$ is obtained from the set $\mathcal{T}_{\overline{k}}^{(l)}$, and is defined to contain the TDOAs relative to the best closed path according to the metric (11) for the lth sensor node. In particular, $\overline{\mathcal{T}}^{(l)}$ contains: the TDOAs $\tau_{ij}^{(l)} \in \mathcal{T}_{\overline{k}}^{(l)}$; the TDOAs $\tau_{ji}^{(l)} = -\tau_{ij}^{(l)}$ such that $\tau_{ij}^{(l)} \in \mathcal{T}_{\overline{k}}^{(l)}$; and the zero-valued TDOAs $\tau_{ii} = 0$ for which a TDOA with subscript i (in any of the two positions) exists in $\mathcal{T}_{\overline{k}}^{(l)}$. More formally,

$$\overline{\mathcal{T}}^{(l)} = \left\{\tau_{ij}^{(l)}, \tau_{ji}^{(l)} = -\tau_{ij}^{(l)} : \tau_{ij}^{(l)} \in \mathcal{T}_{\overline{k}}^{(l)}\right\} \cup \\ \cup \left\{\tau_{ii}^{(l)} = 0, \tau_{jj}^{(l)} = 0 : \exists \tau_{ij}^{(l)} \in \mathcal{T}_{\overline{k}}^{(l)}\right\} \qquad (15)$$

As an example, if $\mathcal{T}_{\overline{k}}^{(l)} = \{\tau_{12}^{(l)} \tau_{23}^{(l)} \tau_{31}^{(l)}\}$, then $\overline{\mathcal{T}}^{(l)} = \{\tau_{12}^{(l)} \tau_{23}^{(l)} \tau_{31}^{(l)}\} \cup \{\tau_{21}^{(l)} \tau_{32}^{(l)} \tau_{13}^{(l)}\} \cup \{\tau_{11}^{(l)} \tau_{22}^{(l)} \tau_{33}^{(l)}\}$.

It is important to observe that the dependency of the cost function in (13) from the global reference point $\mathbf{m}_1^{(1)}$ is purely geometrical. Thus, the cost function can be correctly evaluated even when no TDOAs referred to the microphone at $\mathbf{m}_1^{(1)}$ are considered (i.e., when $\tau_{i1}^{(1)} \notin \overline{\mathcal{T}}^{(1)}, i = 2, \dots N^{(1)}$).

The source location $\mathbf{x}_S$, along with the term r_S, is estimated as

$$\hat{\mathbf{X}}_S = [\hat{\mathbf{x}}_S^T, \hat{r}_S]^T = \arg\min_{\mathbf{X}} J(\mathbf{X}). \qquad (16)$$

However, as $J(\mathbf{X})$ is non-linear, we cannot exploit a trivial closed-form solution to find its minimum. For this reason, the minimization of $J(\mathbf{X})$ is accomplished as in [19], where an iterative optimization is formulated starting from a Taylor expansion of the cost function. The rationale behind the minimiza-

tion approach is to approximate the cost function through linearization, find the minimum of the approximated cost-function in closed-form, update the linear approximation and iterate the process until convergence.

To provide the reader a better insight into the minimization procedure, let us consider the first order Taylor expansion of the HCE metric around a given initial guess point $\mathbf{X}_{S,0}$

$$\varepsilon_{ij}^{(l)}(\mathbf{X}) \simeq \varepsilon_{ij}^{(l)}(\mathbf{X}_{S,0}) + \nabla\varepsilon_{ij}^{(l)}(\mathbf{X})\big|_{\mathbf{X}_{S,0}} \cdot (\mathbf{X} - \mathbf{X}_{S,0}), \quad (17)$$

where

$$\nabla\varepsilon_{ij}^{(l)}(\mathbf{X}) = \begin{bmatrix} \frac{\partial\varepsilon_{ij}^{(l)}(\mathbf{x})}{\partial x} & \frac{\partial\varepsilon_{ij}^{(l)}(\mathbf{x})}{\partial y} & \frac{\partial\varepsilon_{ij}^{(l)}(\mathbf{x})}{\partial z} & \frac{\partial\varepsilon_{ij}^{(l)}(\mathbf{x})}{\partial r} \end{bmatrix} \quad (18).$$

We can rewrite (17) using vector notation as

$$\boldsymbol{\varepsilon}(\mathbf{X}) \simeq \boldsymbol{\varepsilon}(\mathbf{X}_{S,0}) + \nabla\boldsymbol{\varepsilon}(\mathbf{X})\big|_{\mathbf{X}_{S,0}} \cdot (\mathbf{X} - \mathbf{X}_{S,0}), \quad (19)$$

where we stack all the terms $\varepsilon_{ij}^{(l)}(\mathbf{X})$ with $P_{ij}^{(l)} \neq 0$ into the column vector $\boldsymbol{\varepsilon}(\mathbf{X})$ and we build the Jacobian matrix $\nabla\boldsymbol{\varepsilon}(\mathbf{X})$, whose rows contain the gradient of the elements in $\boldsymbol{\varepsilon}(\mathbf{X})$ computed using (18). Evaluating (19) in $\mathbf{X}_S$ leads to

$$\boldsymbol{\varepsilon}(\mathbf{X}_{S,0}) + \nabla\boldsymbol{\varepsilon}(\mathbf{X})\big|_{\mathbf{X}_{S,0}} \cdot (\mathbf{X}_S - \mathbf{X}_{S,0}) \simeq \boldsymbol{\varepsilon}(\mathbf{X}_S) = 0, \quad (20)$$

where $\boldsymbol{\varepsilon}(\mathbf{X}_S) = 0$ holds by definition in case of noiseless measurements. Rearranging the terms of (20) we obtain

$$\nabla\boldsymbol{\varepsilon}(\mathbf{X})\big|_{\mathbf{X}_{S,0}} \cdot (\mathbf{X}_S - \mathbf{X}_{S,0}) \simeq -\boldsymbol{\varepsilon}(\mathbf{X}_{S,0}),$$

which leads to

$$\mathbf{X}_S \simeq \mathbf{X}_{S,0} - \nabla\boldsymbol{\varepsilon}^\dagger(\mathbf{X})\big|_{\mathbf{X}_{S,0}} \cdot \boldsymbol{\varepsilon}(\mathbf{X}_{S,0}), \quad (21)$$

where

$$\nabla\boldsymbol{\varepsilon}^\dagger(\mathbf{X})\big|_{\mathbf{X}_{S,0}} = \left[\nabla\boldsymbol{\varepsilon}^T(\mathbf{X})\big|_{\mathbf{x}_{S,0}} \cdot \nabla\boldsymbol{\varepsilon}(\mathbf{X})\big|_{\mathbf{x}_{S,0}} \right]^{-1} \cdot \nabla\boldsymbol{\varepsilon}^T(\mathbf{X})\big|_{\mathbf{x}_{S,0}}. \quad (22)$$

Notice that (21) provides us with a closed form solution for $\mathbf{X}_S$ if the HCE cost function were linear. Since it is non-linear we need to resort to an iterative procedure. We estimate the source position at iteration v with the linear model (21), we linearize the HCE metric around the new source position estimate according to (19), and iterate the process until convergence. More formally, we can describe this iterative method by means of a single update equation. The update equation at the iteration $v + 1$, starting from the estimate $\hat{\mathbf{X}}_{S,v}$ at the iteration v, is defined as

$$\hat{\mathbf{X}}_{S,v+1} = \hat{\mathbf{X}}_{S,v} - \nabla\boldsymbol{\varepsilon}^\dagger(\mathbf{X})\big|_{\hat{\mathbf{X}}_{S,v}} \cdot \boldsymbol{\varepsilon}(\hat{\mathbf{X}}_{S,v}), \quad (23)$$

The iterative process stops when $\|\hat{\mathbf{X}}_{S,v+1} - \hat{\mathbf{X}}_{S,v}\|$ goes below a prescribed threshold. The source position is estimated considering the first three components of the vector $\hat{\mathbf{X}}_S = \hat{\mathbf{X}}_{S,v+1}$ at the final iteration.

It is worth noticing that, according to (23), the source position estimate depends on the initial guess $\mathbf{X}_{S,0}$ (i.e., the initialization point of our algorithm). In our experiments, $\mathbf{X}_{S,0}$ was chosen as the center of the volume of potential source locations, still

enabling the iterative algorithm to always converge to the global minimum in a few steps. This proves that the iterative method does not strongly depend on the initialization point. Therefore, we can conclude that, even though the proposed cost function is non-linear, it is actually fairly smooth, thus easy to minimize.

C. Location Estimate Refinement

The estimate $\hat{\mathbf{x}}_S$ of the source position can be inaccurate due to the presence of outlier measurements. For example, the disambiguation procedure could fail for some sensors and lead to TDOAs associated to an image source that satisfies the zero-sum condition. In this paragraph we exploit the structure of the cost function $J(\mathbf{X})$ to detect potential outliers. This process is based on the following considerations, confirmed by the experiments of Section V:

- a small set of correct TDOAs is preferable to a larger set that includes potential outliers. In fact, a small number of nodes and TDOAs are sufficient to guarantee good localization accuracy, while the presence of just a few outliers is sufficient cause of relevant localization errors. This is quite common to most regression problems [48];

- if most of the measurements are associated to the correct source position, an analysis of the distribution of the residuals $\varepsilon_{ij}^{(l)}$ of the cost function in (13) can be used for detecting potential outliers (which are expected to produce higher residuals [48]);

- the last condition is often verified if: i) the sensors are sufficiently far from one another; ii) the spatial node distribution enables the observation of the scene from multiple points of view. Under these assumptions, we trust that some sensors will be in good visibility conditions and produce correct TDOAs. The remaining sensors are more likely to be affected by reverberation and to produce outliers. Anyway, as different nodes are unlikely to be in the line of sight of the same image source, most of the coherent measurements are expected to come from the real source to be localized.

These considerations suggest us how to define a suitable reliability index. As already explained, the source localization problem can be seen as a hyper-cone least-square fitting problem. At the end of the localization iterative procedure, the set $\mathcal{E} = \{\varepsilon_{ij}^{(l)}(\hat{\mathbf{X}}_S) : \tau_{ij}^{(l)} \in \overline{\mathcal{T}}^{(l)}, l = 1, \ldots L\}$ represents how well each TDOA measurement fits the hyper-cone in terms of least-square residuals. It is thus possible to check the value of each component of $\mathcal{E}$ to detect whether some measurements are not compatible with the others. As suggested in [48] for generic fitting problems, we compute the standard deviation of the set $\mathcal{E}$ as

$$\sigma_{\mathcal{E}} = \sqrt{\frac{1}{|\mathcal{E}|} \sum_{\varepsilon \in \mathcal{E}} (\varepsilon - \bar{\varepsilon})^2}, \quad (24)$$

where $|\mathcal{E}|$ is the cardinality of the set, ε is an element of the set, and $\bar{\varepsilon}$ is the average value of the set. The standard deviation is then compared to a threshold $\Gamma_{\mathcal{E}}$. If $\sigma_{\mathcal{E}} > \Gamma_{\mathcal{E}}$ the estimate requires to be refined. In order to do so, we search for the TDOA associated to the maximum value in $\mathcal{E}$. This points to the sensor node that worst fits the hyper-cone. We then remove all the measurements associated to this sensor, as they are likely

to identify a reflected source. We do so by deleting the corresponding $\varepsilon_{ij}^{(l)}$ values from $\mathcal{E}$ and re-computing $\sigma_{\mathcal{E}}$. We iterate this procedure until $\sigma_{\mathcal{E}} \leq \Gamma_{\mathcal{E}}$ or a minimum number of sensors has been reached. At the end, the source location is re-estimated using the algorithm in par. 3.B. As noticed before, the choice of the global reference point is independent of the set of TDOAs used for localization. Therefore, the global reference microphone $\mathbf{m}_1^{(1)}$ can be selected as reference even when the first sensor node is an outlier.

Notice that the TDOA outlier identification procedure is computationally efficient (details are provided in Section IV). The values of $\mathcal{E}$, in fact, are already available after the initial localization. Moreover, the re-estimation only happens once, i.e., when all the outliers have been purged. The proposed technique is therefore suitable for the considered distributed sensor network scenario, as all the computations take place in the node used for the localization, without requiring any further data exchange between nodes. To summarize, the pseudo-code of the algorithm is shown in Alg. 1.

Algorithm 1: Pseudo-code of the proposed localization system

Data: Microphone positions $\mathbf{m}_{1,\dots,N^{(l)}}^{(l)}$ and acquired signals $s_{1,\dots,N^{(l)}}^{(l)}(n)$

Result: Source location estimate $\hat{\mathbf{X}}_S$

For $l = 1, \dots, L$ **do**

    Compute the GCCs $R_{ij}^{(l)}(n)$

    Perform peak-picking and TDOA disambiguation to obtain the sets $\mathcal{T}_k$ of candidate TDOAs

    Perform TDOA selection on $\mathcal{T}_k$ to obtain the set of TDOAs $\mathcal{T}_{\overline{k}}$.

end

Compute the source location estimate $\hat{\mathbf{X}}_S$ and the residuals $\varepsilon_{ij}^{(l)}(\hat{\mathbf{X}}_S)$

Compute the variance σ_ε of the residuals $\varepsilon_{ij}^{(l)}(\hat{\mathbf{X}}_S)$

if $\sigma_\varepsilon > \Gamma_\varepsilon$ **then**

    while $\sigma_\varepsilon > \Gamma_\varepsilon$ **and** Number of used arrays ≥ 2

    do

        Remove the array with the highest contribution to $J(\mathbf{X})$

        Recompute σ_ε

    end

    Compute the refined source location estimate $\hat{\mathbf{X}}_S$

end

IV. COMPUTATIONAL COMPLEXITY ANALYSIS

In this section we analyze the computational cost of the proposed system. We first provide details about all the functional blocks composing the algorithm, then we estimate the complexity of the whole system.

A. Complexity of the individual components

For simplicity, we assume that the number of microphones in the sensor nodes is constant, i.e., $N^{(l)} = N$ for $l = 1, \dots L$.

Signals at the microphones are analyzed on a frame-to-frame basis, and the length of the frames is $M/2$ samples. For each frame, the following operations are required:

GCC computation. The GCC (1) is computed for all the microphone pairs, whose number is $F = N(N-1)/2$ for each sensor node. We distinguish among the following steps:

- **computing FFTs** (Modified Split-Radix alg. [49]): $N \times L \times \frac{34}{9} M \log_2(M)$ multiplications/additions;
- **computing the cross-spectra**: $4 \times F \times L \times M$ multiplications and $3 \times F \times L \times M$ additions;
- **spectral processing**: $4 \times F \times L \times M$ multiplications, $3 \times F \times L \times M$ additions, $F \times L \times M$ square roots, and $F \times L \times M$ divisions;
- **computing IFFTs**: $F \times L \times \frac{34}{9} M \log_2(M)$ multiplications/additions;

TDOAs disambiguation and selection. In each node, there exist $C_3 = \binom{N}{3}\frac{(3-1)!}{2} = \binom{N}{3}$ and $C_4 = \binom{N}{4}\frac{(4-1)!}{2} = 3\binom{N}{4}$ independent closed paths (loops) of 3 and 4 microphones, respectively. Therefore, starting from 3 peaks extracted from each TDOA, the number of candidate loops of length 3 is $Q_3 = 3^3 C_3$; and that of length 4 is $Q_4 = 3^4 C_4$. The resulting $L \times (Q_3 + Q_4)$ combinations can be stored in a lookup table, which can be computed offline. The computational cost of disambiguation and selection, at run-time, can be summarized as follows:

- **extracting the highest peaks from each GCC**: $F \times L \times (M - 1)$ additions (to compute numerical derivatives), $F \times L \times (M - 2)$ comparisons (to compute the sign of derivatives), M comparisons (to find the highest peaks);
- **selecting loops that satisfy** $z_k \leq \Gamma_z$: $(3 \times Q_3 + 4 \times Q_4) \times L$ comparisons;
- **computing** $\max[R_{ij}(n)]$ for all the sets $\mathcal{T}_k$: $(3 \times Q_3 + 4 \times Q_4) \times L$ comparisons;
- **verifying the zero-sum condition**: $(3 \times Q_3 + 4 \times Q_4) \times L$ additions, and $(Q_3 + Q_4) \times L$ comparisons;
- **computing** $q_k^{(1)}$ **(8)**: $(3 \times Q_3 + 4 \times Q_4) \times L$ additions, comparisons, and divisions;
- **computing** $q_k^{(2)}$ **(9)**: $(3 \times Q_3 + 4 \times Q_4) \times L$ sums, $(Q_3 + Q_4) \times L$ divisions;
- **computing** $q_k^{(3)}$ **(10)**: $(Q_3 + Q_4) \times L$ divisions, $(3 \times Q_3 + 4 \times Q_4) \times L$ multiplications;
- **computing the quality values** Q_k **(11)**: $3 \times (Q_3 + Q_4) \times L$ multiplications, $2 \times (Q_3 + Q_4) \times L$ additions;
- **extracting the best** Q_k **score**: $(Q_3 + Q_4) \times L$ comparisons.
- Source localization. The minimization of the cost function (13) follows an iterative procedure, based on the computation of the first-order Taylor expansion of $J(\mathbf{X})$. Here we give an upper bound of the complexity, corresponding to the case in which all available TDOA measurements are used for each node (i.e., $P_{ij}^{(l)} = 1, \forall i, \forall j, \forall l$). Each iteration of the minimization procedure involves the following operations:
- **computing the residuals (12)**: $8 \times L \times N^2$ additions, $4 \times L \times N^2$ multiplications;
- **evaluating (13)**: $L \times N^2$ multiplications and additions;
- **computing the Jacobian matrix** $\nabla\varepsilon$: $108 \times L \times N^2$ additions; $64 \times L \times N^2$ multiplications; $6 \times L \times N^2$ divisions; $16 \times L \times N^2$ square roots;

- **computing the pseudo-inverse (22)**: $16 \times (L \times N^2)$ multiplications and $16 \times (L \times N^2 - 1)$ additions for the term $[\nabla\varepsilon^T \nabla\varepsilon]$; 376 multiplications and 103 additions for its inversion; $16 \times L \times N^2$ multiplications and $12 \times L \times N^2$ additions for determining $\nabla\varepsilon^\dagger$;
- **applying the update equation (23)**: $4 \times L \times N^2$ multiplications, $3 \times L \times N^2 + 4$ additions;

<u>Reliability check</u>. The computation of the reliability consists of evaluating (24), which, in the worst case, involves: $L \times N^2 + 1$ additions, $L \times N^2$ multiplications, 1 division and 1 square root. Finally, $L \times N^2$ comparisons are needed for detecting the potential outliers.

B. Overall computational cost

Assuming that all the basic operations (additions, multiplications, divisions, comparisons, square roots) have the same cost, we can estimate the computational cost in terms of the total number of operations. We denote with I_c the average number of iterations needed for (23) to converge to the optimum solution. Moreover, we set to $L - 2$ the maximum number of nodes that can be discarded after repeating the reliability check. Thus, in the worst case, the check is repeated exactly $L - 2$ times, and the localization step is performed twice. Under these assumptions, the number of operations required by each functional block are

$$4LMN \left[2(N-1) + \frac{17}{9} \log_2(M) \right] \tag{25}$$

for computing the GCCs;

$$M + L \left[810(C_3 + 3.8C_4 + 5.4C_3C_4) \right] + \\ + L \left[N(N-1)(M-1.5) \right] \tag{26}$$

for the TDOA disambiguation and selection;

$$2 \times I_c \times (275LN^2 + 467) \tag{27}$$

to execute 2 source localizations; and

$$(L-2) \times 3(LN^2 + 1). \tag{28}$$

to perform $L - 2$ reliability checks.

The total number of operations required for executing the complete system is thus given by the sum of the terms (25)-(28).

V. Experiments

In this section we evaluate the proposed localization system. After presenting some preliminary simulations aimed at validating specific steps of the system, we test the localization method in a real scenario. Results of the experiments are compared to those obtained through the state-of-the-art Stochastic Region Contraction (SRC) algorithm [20].

A. Simulative Validation

In this paragraph we are interested in assessing the validity of the iterative method used for the minimization of the cost function proposed in Section III-B. Although the cost function has already been validated in [19] for 2D geometries, here we aim at testing the minimization algorithm under several operation conditions. Moreover, we aim at validating the reliability criterion introduced in Section III-C.
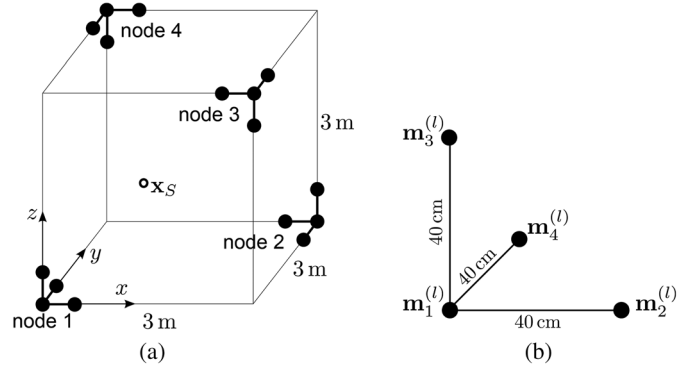

Fig. 5. Simulation setup. (a) geometry of the setup (b) lth sensor node.

For this set of tests, we considered the simulation setup in Fig. 5-(a), where $L = 4$ sensor nodes were employed to localize a source $\mathbf{x}_S$ in a $3 \times 3 \times 3$ m^3 cube. Each node consists of $N = 4$ microphones, disposed as in Fig. 5(b). The first microphone of the lth node (i.e., $\mathbf{m}_1^{(l)}$) is placed at one vertex of the cube, as in Fig. 5(a). We tested 500 source positions, uniformly distributed within the cube. We simulated the measurement of the TDOAs by corrupting the nominal values $\tau_{ij}^{(l)}$ with I.I.D. additive zero-mean Gaussian noise. We set the standard deviation to $\sigma_\tau = \sigma_r/\eta$, $\sigma_r = 0.7$ cm, which corresponds to an error of approximately one sample at $F_s = 48$ kHz. Throughout this paragraph, we will specify the error on the TDOAs by means of the term σ_r, which refers to the standard deviation of the error on the nominal range differences $r_{ij}^{(l)} = \eta \cdot \tau_{ij}^{(l)}$. For each test source position, we obtained a set of 6 TDOAs for each sensor node, for a total of 24 TDOA values. With the aim of simulating the TDOA selection procedure, for each test position the source was localized using different subsets of TDOAs extracted from the 24 available. More specifically, we extracted subsets of $V_t = 3, 4, 5, 6$ TDOAs chosen from $V_n = 2, 3, 4$ sensor nodes. For each pair (V_t, V_n) we simulated $Z = 5000$ Monte-Carlo trials, where both the active sensor nodes and the subset of TDOAs were randomly selected at each repetition. Notice that the case $(V_t = 6, V_n = 4)$ coincides with that of using all the 24 measurements; while $V_t = 3$ and $V_t = 4$ represent the number of TDOAs that define the loops considered in the TDOA disambiguation procedure. For each source position we computed the Root Mean Square Error (RMSE)

$$E(\mathbf{x}_S) = \sqrt{\frac{1}{Z} \sum_{i=1}^{Z} \|\hat{\mathbf{x}}_S(i) - \mathbf{x}_S\|^2},$$

where $\hat{\mathbf{x}}_S(i)$ denotes the estimate of the source at the ith Monte-Carlo trial. Moreover, we computed the Cramer-Rao Lower Bound (CRLB) [17], given by the terms $\sigma_x^2(\mathbf{x}_S), \sigma_y^2(\mathbf{x}_S)$ and $\sigma_z^2(\mathbf{x}_S)$. They represent the lower bound on the variance of an unbiased estimator of the source position, for the x, y, and z coordinates, respectively. In all the considered tests, the bias of the estimation turned out to be negligible (< 0.15 mm for all the spatial dimensions). It follows that the proposed estimator can be considered unbiased. Moreover, the RMSE corresponds to the variance of the estimator, thus the term $\Sigma(\mathbf{x}_S) = \sqrt{\sigma_x^2(\mathbf{x}_S) + \sigma_y^2(\mathbf{x}_S) + \sigma_z^2(\mathbf{x}_S)}$ represents the lower bound for the RMSE, i.e. $E(\mathbf{x}_S) \geq \Sigma(\mathbf{x}_S)$.
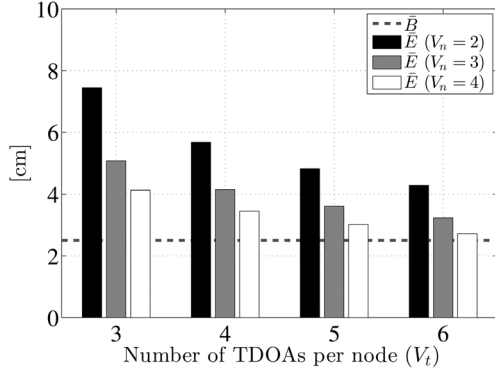
Fig. 6. RMS localization error as a function of the number of TDOAs per node (V_t) and number of nodes (V_n), compared with the lower bound $\bar{B}$.
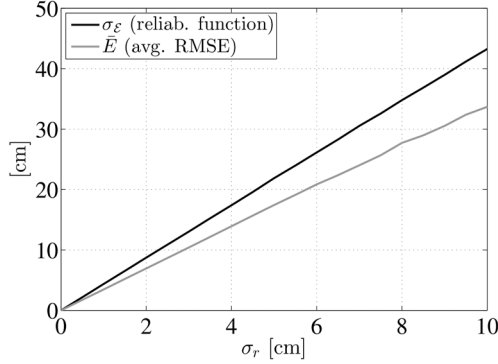

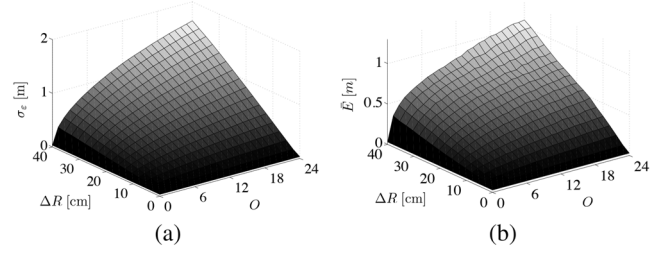Fig. 7. $\sigma_{\mathcal{E}}$ and $\bar{E}$ as a function of the range difference error.

Fig. 8. $\sigma_{\mathcal{E}}$ (a) and RMSE localization error $\bar{E}$ (b) as a function of the range difference error ΔR and number of corrupted TDOAs O (a) reliability (b) RMSE.

The results of the simulation are shown in Fig. 6, which reports the average RMSE $\bar{E}$ for the considered configurations, obtained by averaging $E(\mathbf{x}_S)$ over all source positions. The dashed line in Fig. 6 represents the average lower bound $\bar{B}$, whose value was computed averaging $\Sigma(\mathbf{x}_S)$ for all the tested positions. We observe that, when all the 24 TDOAs are adopted for the localization, $\bar{E}$ approaches the bound $\bar{B}$, thus confirming that the minimization procedure converges to the optimum solution. Moreover, we notice that reducing the number of measurements used for the localization has a limited impact on the accuracy. In fact, even in the most challenging case with $V_t = 3$ and $V_n = 2$, corresponding to localizing with only 6 TDOAs, $\bar{E}$ maintains below 8 cm.

We now focus on the reliability function defined in (24). Considering again the setup described before, we conducted several tests. First, we investigated the behavior of the function $\sigma_{\mathcal{E}}$ when the TDOAs are affected by I.I.D. additive zero-mean Gaussian noise. The results are shown in Fig. 7, which shows $\sigma_{\mathcal{E}}$ and $\bar{E}$ as a function of the standard deviation σ_r of the noise applied to the nominal TDOAs. The results were obtained considering $V_t = 6$ and $V_n = 4$, averaging the outcomes of 5000 Monte-Carlo trials. Notice that both $\sigma_{\mathcal{E}}$ and $\bar{E}$ are directly proportional to σ_r. Consequently, the reliability function can be interpreted as an indicator of the accuracy of the localization. In fact, in the absence of outlier measurements, we expect small values of $\sigma_{\mathcal{E}}$ (e.g., $\sigma_{\mathcal{E}} < 20$ cm when $\sigma_r < 4$ cm). In the realistic case of noise of standard deviation $\sigma_r = 0.7$ cm, we obtained $\sigma_{\mathcal{E}} \approx 3$ cm and $\bar{E} \approx 2.7$ cm.

In order to gain more insight into the function $\sigma_{\mathcal{E}}$, we studied its behavior also in presence of outlier measurements. In par-

ticular, we computed the value of $\sigma_{\mathcal{E}}$ when a number O of TDOAs (considered as outliers), randomly selected from the full set of measurements, were corrupted with a bias equal to $\Delta T = \Delta R / \eta$. As before, we specify the bias on the range differences, by means of the term ΔR. Similarly to the previous tests, for each $(O, \Delta R)$ pair we run 5000 Monte-Carlo trials. Results are shown in Figs. 8(a) and 08(b), which show the reliability $\sigma_{\mathcal{E}}$ and the average RMSE $\bar{E}$, respectively, as a function of O and ΔR. We first notice that the trend of $\bar{E}$ is highly correlated to that of $\sigma_{\mathcal{E}}$, thus confirming the validity of the latter as an indicator of the reliability of the estimation. We can also observe that $\sigma_{\mathcal{E}}$ rapidly increases in the presence of outliers. As an example, when $O = 6$, which equals the number of TDOAs measured in one sensor node, we obtained $\sigma_{\mathcal{E}} \approx 22$ cm for $\Delta R = 10$ cm; $\sigma_{\mathcal{E}} \approx 44$ cm for $\Delta R = 20$ cm; $\sigma_{\mathcal{E}} \approx 85$ cm for $\Delta R = 40$ cm. In the light of these considerations, it is reasonable to mark as unreliable an estimation if $\sigma_{\mathcal{E}}$ is above a certain threshold $\Gamma_{\mathcal{E}}$. The analysis of Fig. 8 confirms that the presence of biased TDOAs introduces relevant localization errors, which is one of the assumptions which the localization refinement step is based on (see par. III-C).

Finally, we show an example of detection of potential outliers from the TDOA measurements, focusing on two realizations of the previous simulation. In particular, we randomly selected one source position $\mathbf{x}_S$, and we considered the cases of introducing 3 outliers within 1 and 2 sensor nodes, adding a bias $\Delta R = 40$ cm to the corresponding TDOAs. The resulting RMSEs are $E(\mathbf{x}_S) = 52$ cm for the first case, and $E(\mathbf{x}_S) = 74$ cm for the second case. The reliability function is $\sigma_{\mathcal{E}} = 47$ cm and $\sigma_{\mathcal{E}} = 75$ cm, respectively. These values are considerably higher than the threshold $\Gamma_{\mathcal{E}} = 10$ cm, selected as the value of $\sigma_{\mathcal{E}}$ which corresponds to TDOAs affected by I.I.D. additive Gaussian noise with $\sigma_r = 2$ cm. The localization results are thus marked as unreliable, and we detect the outliers by observing the individual contributions of the cost function, i.e., the values in the set $\mathcal{E}$. Fig. 9 shows the terms $\varepsilon^2, \varepsilon \in \mathcal{E}$, grouped by sensor nodes, in the case of one (a) and two (b) corrupted nodes. The height of the peaks associated to the outliers makes possible to detect them. In these two particular cases, outliers are present in the second node (case in Fig. 9-(a)); and in the second and fourth nodes (Fig. 9-(b)). Fig. 9 confirms therefore that it is possible to identify the nodes containing the outliers, as assumed in par. III-C.

B. Experiments in a Real Setup

We now consider the implementation of the localization system in a real setup. $L = 4$ sensor nodes with $N^{(l)} = 4$
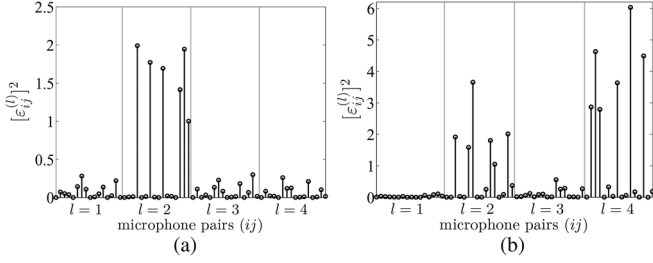
Fig. 9. Distribution of ε for the two cases of one (a) or two (b) arrays affected by corrupted TDOAs.

microphones, with the same internal geometry of the nodes in Fig. 5, were installed in a medium-sized reverberant office room. The room measures $7 \times 7 \times 3 m^3$ and we estimated a reverberation time $RT_{60} = 0.73$ s. Each sensor accommodated 4 identical Beyerdynamic MM1 measurement microphones, with a capsule of 9 mm of diameter. Similarly to the setup in Fig. 5, the nodes were placed at 4 vertices of a volume of size $3.8 \times 2.5 \times 1.8$ m^3 delimiting the potential positions of the source. We defined 36 test source positions within the volume, disposed on 3 horizontal rectangular grids at different heights from the bottom side, namely at $z = 0.4$ m, $z = 0.7$ m and $z = 1$ m. Each grid arranged 3×4 sources with a spacing of 0.75 m on the x coordinate and 0.9 m on the y coordinate. A small loudspeaker (whose cone has a diameter of 4 cm) was moved in all the 36 positions, and emitted 30 s of white noise and 30 s of a speech signal at the sampling frequency $F_s = 48$ kHz. Acquired signals were segmented in frames of length 4096 samples, corresponding to a duration of about 170 ms. As far as the speech signal was concerned, we discarded the silent frames using the technique described in [50]. For all the tested source locations, the estimate was accomplished with the proposed methodology and with the Stochastic Region Contraction (SRC) method, presented in [20]. SRC is a relaxation of the SRP technique, which revealed to be particularly suitable for real-time applications. For the tests we adopted the SRC Matlab implementation released by the authors[1] selecting 3000 initial random points (i.e., using the default parameters of the implementation). The metric used for the evaluation is the RMSE. In order to quantify the computational complexity of the algorithms under analysis, for each frame analyzed we kept trace of: the average number of evaluations I_s of the SRC functional; and the average number of iterations I_c required for minimizing (13).

Fig. 10 shows the RMSE obtained with the proposed localization system (a,b,c) and with SRC (d,e,f). Each group of three diagrams refer to the three grids at different heights from the bottom of the volume. For each source position, the radius of the circle represents the RMS localization error. Notice that the proposed algorithm achieves performances comparable to that of SRC for all the considered source positions. The localization of speech sources, shown in Fig. 11, where the diagrams are organized as in Fig. 10, confirms this fact. Due to the nature of speech signals, which yield noisier GCC functions, the RMSE is slightly higher than the one obtained with the random noise signal.
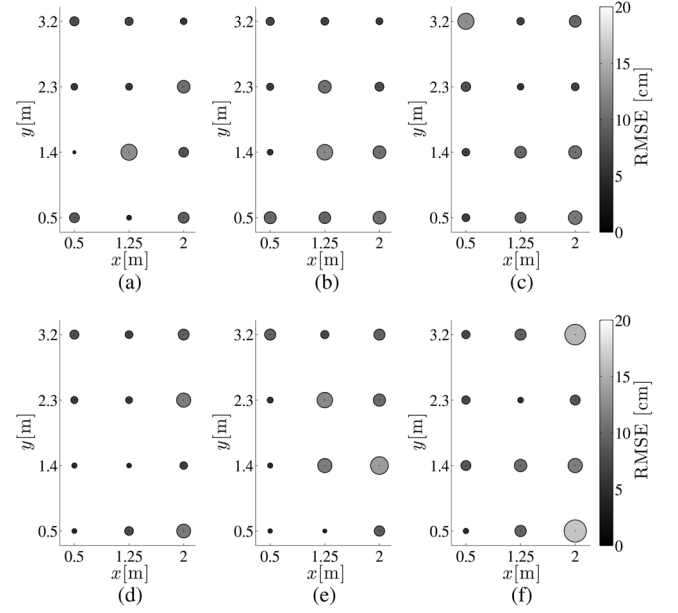
[1]http://www.lems.brown.edu/array/download.html

Fig. 10. The radius of each circle is equal to the RMSE for the proposed system (top) and SRP (bottom) in a real reverberant environment using a random Gaussian noise audio signal. (a) proposed($z = 0.4$ m) (b) proposed($z = 0.7$ m) (c) proposed($z = 1$ m) (d) SRC($z = 0.4$ m) (e) SRC($z = 0.7$ m) (f) SRC($z = 1$ m).
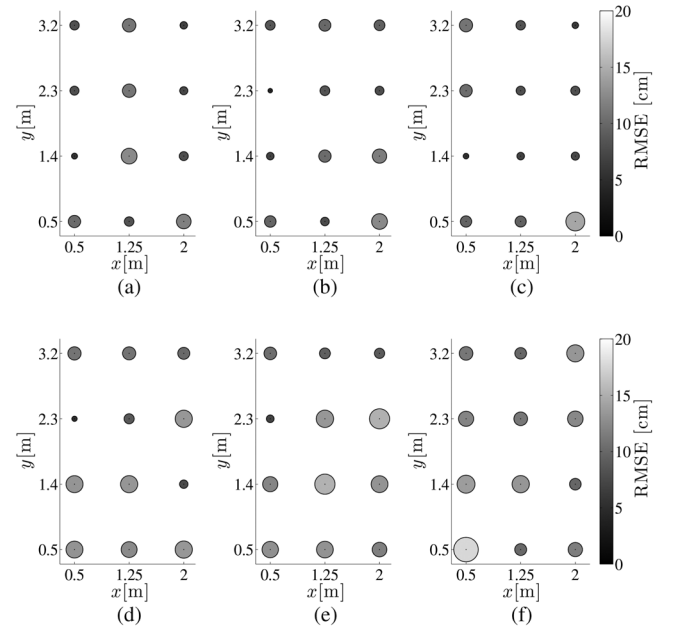

Fig. 11. The radius of each circle is equal to the RMSE for the proposed system (top) and SRC (bottom) in a real reverberant environment using a male speech audio signal. (a) proposed($z = 0.4$ m) (b) proposed($z = 0.7$ m) (c) proposed($z = 1$ m) (d) SRC($z = 0.4$ m) (e) SRC($z = 0.7$ m) (f) SRC($z = 1$ m).

As far as the computational cost is concerned, SRC required on average $I_s = 32327$ evaluations of the functional; while an average of $I_c = 10$ iterations guaranteed the convergence of the gradient method used for minimizing the proposed cost function. Each evaluation of the SRC functional requires 5 operations for each microphone pair [20], thus, in the considered setup, the complexity of SRC is equal to $5LN(N-1) \times I_s$ operations. Both proposed technique and SRC rely on the computation of $LN(N-1)$ GGCs, which, according to (25), re-

TABLE I
AVERAGE COMPUTATIONAL COMPLEXITY OF THE
PROPOSED AND THE SRC APPROACHES

Technique	Steps	N. of operations
PROPOSED	TDOA disamb. and selection	$661 \cdot 10^3$
	First localization	$9 \cdot 10^3$
	Reliability check	195
	Final localization	$9 \cdot 10^3$
	total	$\mathbf{679 \cdot 10^3}$
SRC	**total**	$\mathbf{7758 \cdot 10^3}$

TABLE II
RMSE ON x, y AND z COORDINATES CONSIDERING BOTH RANDOM NOISE
AND SPEECH. RESULTS ARE SHOWN USING SRC AND THE PROPOSED
SYSTEM BEFORE AND AFTER THE LOCALIZATION REFINEMENT STEP

		$RMSE_x$ [cm]	$RMSE_y$ [cm]	$RMSE_z$ [cm]
Random Noise	proposed (before ref.)	18.85	13.28	14.61
	proposed (after ref.)	4.02	3.45	4.62
	SRC	3.94	3.96	5.01
Speech	proposed (before ref.)	32.60	22.72	23.63
	proposed (after ref.)	5.04	4.59	4.83
	SRC	5.60	5.52	6.12

TABLE III
PERCENTAGE OF LOCALIZATIONS USING A DIFFERENT AMOUNT OF
ARRAYS FOR BOTH RANDOM NOISE AND SPEECH SIGNAL

	4 Arrays	**3 Arrays**	**2 Arrays**
Random Noise	83.95%	5.13%	10.92%
Speech	66.05%	3.88%	30.07%

TABLE IV
PERCENTAGE OF REMOVED OUTLIERS

ΔR [cm]	**TPR**	**TNR**	**FPR**	**FNR**
5	91.1%	100%	0%	8.9%
10	96.9%	100%	0%	3.1%

is positive and no refinements are necessary. When the speech source is considered, this percentage drops to about 66%, confirming that the TDOAs extracted from the GCCs are less reliable in this case. When the reliability check fails, the localization is more likely to be performed through 2 nodes instead of 3. This is not unexpected, as 2 is set to be the minimum number of nodes used for localizing a source. In fact, a single node may lead to very inaccurate estimates, particularly when the source is in far field with respect to that node. Consequently, when the reliability check fails twice, the 2 remaining nodes are always used for the localization, even if a further reliability check would possibly fail again. Nevertheless, in almost all the cases, we observed that the TDOAs coming from the 2 remaining nodes are free of outliers.

Finally, Table IV shows some statistics about the ability of the proposed algorithm to detect the presence of outlier measurements in the real case scenario. To this purpose, we considered a node as affected by outliers if at least one of the range differences measured by the sensor node were affected by an error greater than ΔR with respect to the ground truth. We then checked whether our algorithm was able to detect the outlier exploiting the reliability index (24). More specifically, considering all the source positions and realizations, Table IV reports: i) the true positive rate (TPR), which is the percentage of nodes, averaged over all the realizations, affected by outliers that were correctly detected; ii) the true negative rate (TNR), which is the percentage of nodes, averaged over all the realizations, not affected by outliers correctly classified as outlier-free; iii) the false positive rate (FPR), which is the fraction of nodes, averaged over all the realizations, not affected by outliers but detected as outliers; iv) the false negative rate (FNR), which is the fraction of nodes, averaged over all the realizations, affected by outliers that were detected as non-affected. Results are reported for two different values of ΔR, i.e., for both a more strict and less strict outlier detection definition. Clearly, our goal is to maximize the TPR and TNR, while minimizing the FPR and FNR. It is interesting to notice that we always detect (thus correctly exploit in our cost function) the nodes not affected by outliers (i.e., TNR = 100%). Additionally, we detect more than 91% of the outliers, thus keeping less then the 9% of them after refinement. This is the reason why the refinement step greatly improves localization accuracy.

quire $\approx 16 \cdot 10^6$ operations. The complexity of the remaining steps is summarized in Table I, which compares the number of operations required by the proposed method and by SRC (excluding the GCC step). For the proposed method, the TDOA disambiguation and selection is dominant in terms of operations. Indeed, the cost of localization and the refinement turn out to be almost negligible in evaluating the overall complexity of the system. For the considered scenario the cost of the proposed technique turns out to be over 10 times smaller that of the SRC method. For the sake of comparison, we can also estimate the number of operations required by the standard SRP method, which estimates the source position evaluating the SRP functional in all the points that sample the volume of interest. Assuming a sampling of $2\ \text{cm}^3$ of the considered volume ($3.8 \times 2.5 \times 1.8\ \text{m}^3$), SRP would require $I_s = 2137500$ evaluations of the functional, corresponding to $\approx 256 \cdot 10^6$ operations. This suggests that SRP is less suitable for real-time operation in this specific scenario.

Table II details the RMSE scores on the three spatial coordinates for both the proposed and the SRC algorithm. As for the proposed method, we included the localization results before the reliability check, i.e., before discarding potential outliers from the set of measurements. The comparison with the results obtained after the reliability check highlights the importance of this operation, which leads to a significant reduction of the RMSE. This is especially true for the localization of the speech source, where the reliability check allows to reduce the RMSE from values ≥ 22 cm to about 5 cm.

Also interesting are results in Table III, which details the percentage of localizations accomplished by using 4, 3 and 2 sensor nodes, adopting the proposed method. The percentage of localizations of the random noise source with 4 nodes is very high ($\approx 84\%$), meaning that in most of the cases the reliability check

VI. CONCLUSIONS

In this manuscript we proposed a system for the localization of acoustic sources based on TDOA measurements that works

in a distributed sensor network scenario. The localization algorithm is based on the use of the four-dimensional space-range reference frame. Here, TDOA measurements describe the surface of a hyper-cone, whose apex is the source location. Localizing the source, therefore, corresponds to finding the apex of such cone. A cost function is defined at this purpose. In order to improve the robustness against reverberation, a geometrically motivated TDOA selection step is implemented. An estimate refinement step follows the localization to further improve the accuracy. The refinement is based on the analysis of the distributions of the residuals of the hyper-cone fitting function. Simulative results prove the accuracy of the localization function, showing the effect of the reliability-based refinement. Experimental results validate the localization system in a real-world scenario, comparing the proposed methodology with Stochastic Region Contraction (SRC), an effective implementation of SRP. Results confirm that the proposed technique attains at least the same accuracy of SRP, while saving computational power. The clear subdivision of tasks between sensor nodes and central nodes, the reduced cost and the small amount of information exchanged between nodes, make the proposed algorithm suitable for the considered scenario of distributed sensor networks.

REFERENCES

[1] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," in *Proc. IEEE Symp. Commun. Veh. Technol. Benelux (SCVT)*, 2011.

[2] *Microphone Arrays Signal Processing Techniques and Applications*, M. Brandstein and D. Ward, Eds.. Berlin, Germany: Springer-Verlag, 2001.

[3] H. Wang and P. Chu, "Voice source localization for automatic camera pointing system in videoconferencing," in *Proc. IEEE Workshop Applicat. Signal Process. Audio Acoust. (WASPAA)*, 1997, pp. 187–190.

[4] F. Antonacci, D. Riva, D. Saiu, A. Sarti, M. Tagliasacchi, and S. Tubaro, "Tracking multiple acoustic sources using particle filtering," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2006.

[5] F. Antonacci, M. Matteucci, D. Migliore, D. Riva, A. Sarti, M. Tagliasacchi, and S. Tubaro, "Tracking multiple acoustic sources in reverberant environments using regularized particle filter," in *Proc. 15th Int. Conf. Digital Signal Process. (DSP)*, Jul. 2007.

[6] Y. Y. Guo and M. Hazas, "Acoustic source localization of everyday sounds using wireless sensor networks," in *Proc. ACM Int. Conf. Adjunct Papers Ubiquitous Comput. (UbiComp)*, 2010.

[7] K. Na, Y. Kim, and H. Cha, "Acoustic sensor network-based parking lot surveillance system," in *Wireless Sensor Networks*, ser. Lecture Notes in Computer Science, U. Roedig and C. Sreenan, Eds. New York, NY, USA: Springer, 2009, vol. 5432, pp. 247–262.

[8] L. Parker, B. Birch, and C. Reardon, "Indoor target intercept using an acoustic sensor network and dual wavefront path planning," in *Proc. IEEE/RSJ Int. Conf. Intell. Rob. Syst. (IROS)*, 2003, pp. 278–283.

[9] J. Alba, A. Alba, E. Escuder, J. Ramis, and R. D. Rey, "Acoustic environmental monitoring system for acoustic maps," in *Proc. Int. Conf. Mobile Ubiquitous Comput., Syst., Services, Technol. (UBICOMM)*, 2008.

[10] H. Wang, C. Chen, A. Ali, S. Asgari, R. Hudson, K. Yao, D. Estrin, and C. Taylor, "Acoustic sensor networks for woodpecker localization," *SPIE Adv. Signal Process. Algorithms, Architect., Implement.*, 2005.

[11] C. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-24, no. 4, pp. 320–327, Aug. 1976.

[12] J. Ianniello, "Time delay estimation via cross-correlation in the presence of large estimation errors," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-30, no. 6, pp. 998–1003, Nov. 1982.

[13] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 549–557, Nov. 2003.

[14] M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 1994, pp. 273–276.

[15] M. Gillette and H. Silverman, "A linear closed-form algorithm for source localization from time-differences of arrival," *IEEE Signal Process. Lett.*, vol. 15, pp. 1–4, 2008.

[16] H. Schau and A. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 8, pp. 1223–1225, Aug. 1987.

[17] Y. Huang, J. Benesty, G. Elko, and R. Mersereati, "Real-time passive source localization: A practical linear-correction least-squares approach," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 8, pp. 943–956, Nov. 2001.

[18] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions of source localization problems," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1770–1778, May 2008.

[19] M. Compagnoni, P. Bestagini, F. Antonacci, A. Sarti, and S. Tubaro, "Localization of acoustic sources through the fitting of propagation cones using multiple independent arrays," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 7, pp. 1964–1975, Sep. 2012.

[20] H. Do, H. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using Stochastic Region Contraction(SRC) on a large-aperture microphone arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2007, pp. 121–124.

[21] J. Dmochowski, J. Benesty, and S. Affes, "A generalized steered response power method for computationally viable source localization," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 8, pp. 2510–2526, Nov. 2007.

[22] D. Zotkin and R. Duraiswami, "Accelerated speech source localization via a hierarchical search of steered response power," *IEEE Trans. Speech Audio Process.*, vol. 12, no. 5, pp. 499–508, Sep. 2004.

[23] W. Hahn and S. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Trans. Inf. Theory (TIT)*, vol. IT-19, no. 5, pp. 608–614, Sep. 1973.

[24] M. Wax and T. Kailath, "Optimum localization of multiple sources by passive arrays," *IEEE Trans. Acoust., Speech Signal Process.*, vol. ASSP-31, no. 5, pp. 1210–1217, Oct. 1983.

[25] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramer-Rao bound," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 1988, pp. 2296–2299.

[26] J. C. Chen, R. E. Hudson, and K. Yao, "Maximum-likelihood source localization and unknown sensor location estimation for wideband signals in the near-field," *IEEE Trans. Signal Process.*, vol. 50, no. 8, pp. 1843–1854, Aug. 2002.

[27] N. Castaneda, M. Charbit, and E. Moulines, "Source localization from quantized time of arrival measurements," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2006, pp. 933–936.

[28] R. O. Schmidt, "A new approach to geometry of range difference location," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-8, no. 6, pp. 821–835, Nov. 1972.

[29] Y. T. Chan and K. C. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. Signal Process.*, vol. 42, no. 8, pp. 1905–1915, Aug. 1994.

[30] K. Ho, "Bias reduction for an explicit solution of source localization using TDOA," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2101–2114, May 2012.

[31] L. Lin, H. So, F. K. Chan, Y. Chan, and K. Ho, "A new constrained weighted least squares algorithm for TDOA-based localization," *Signal Process.*, vol. 93, no. 11, pp. 2872–2878, 2013.

[32] R. Parisi, A. Cirillo, M. Panella, and A. Uncini, "Source localization in reverberant environments by consistent peak selection," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process. (ICASSP)*, 2007, pp. 37–40.

[33] J. Scheuing and B. Yang, "Disambiguation of TDOA estimates in multi-path multi-source environments (DATEMM)," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2006, pp. 837–840.

[34] J. Scheuing and B. Yang, "Disambiguation of TDOA estimation for multiple sources in reverberant environments," *IEEE Trans. Audio, Speech, Lang. Process. (TASLP)*, vol. 16, no. 8, pp. 1479–1489, Nov. 2008.

[35] C. M. Zannini, A. Cirillo, R. Parisi, and A. Uncini, "Improved TDOA disambiguation techniques for sound source localization in reverberant environments," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, 2010, pp. 2666–2669.

[36] M. Kreissig and B. Yang, "Fast and reliable TDOA assignment in multi-source reverberant environments," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, 2013, pp. 355–359.

[37] P. Annibale, R. Rabenstein, M. Kreissig, and B. Yang, "Joint consistent graph synthesis and speed of sound estimation for acoustic localization in multi-source reverberant environments," in *Proc. Int. Workshop Multidimensional Syst. (nDS)*, 2013.

[38] Y. Yu and H. Silverman, "An improved TDOA-based location estimation algorithm for large aperture microphone arrays," in *IEEE Int. Conf. Acoustics, Speech, Signal Process. (ICASSP)*, May 2004, pp. 77–80.

[39] A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 2, pp. 439–443, Feb. 2013.

[40] D. Bechler and K. Kroschel, "Three different reliability criteria for time delay estimates," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, 2004.

[41] D. Bechler and K. Kroschel, "Reliability criteria evaluation for TDOA estimates in a variety of real environments," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, 2005, pp. 985–988.

[42] P. Bestagini, M. Compagnoni, F. Antonacci, A. Sarti, and S. Tubaro, "TDOA-based acoustic source localization in the space range reference frame," *Multidimens. Syst. Signal Process.*, vol. 25, pp. 337–359, 2014.

[43] K. Hasegawa, N. Ono, S. Miyabe, and S. Sagayama, "Blind estimation of locations and time offsets for distributed recording devices," in *LVA/ICA*, ser. ser. Lecture Notes in Computer Science. New York, NY, USA: Springer, 2010, vol. 6365, pp. 57–64.

[44] V. Raykar, I. Kozintsev, and R. Lienhart, "Position calibration of microphones and loudspeakers in distributed computing platforms," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 70–83, Jan. 2005.

[45] M. Hennecke, T. Plotz, G. A. Fink, J. Schmalenstroer, and R. Hab-Umbach, "A hierarchical approach to unsupervised shape calibration of microphone array networks," in *Proc. IEEE/SP Workshop Statist. Signal Process. (SSP)*, 2009, pp. 257–260.

[46] A. Redondi, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Geometric calibration of distributed microphone arrays," in *Proc. IEEE Int. Workshop Multimedia Signal Process. (MMSP)*, 2009, pp. 1–5.

[47] P. Pertila, M. Mieskolainen, and M. Hamalainen, "Closed-form self-localization of asynchronous microphone arrays," in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays (HSCMA)*, 2011.

[48] R. Andersen, *Modern Methods for Robust Regression (Quantitative Applications in the Social Sciences)*. New York, NY, USA: Sage, 2008.

[49] S. G. Johnson and M. Frigo, "A modified split-radix FFT with fewer arithmetic operations," *IEEE Trans. Signal Process.*, vol. 55, no. 1, pp. 111–119, Jan. 2007.

[50] A. Brutti, M. Omologo, and P. Svaizer, "Multiple source localization based on acoustic map de-emphasis," *EURASIP J. Audio, Speech, Music Process. (JASMP)*, vol. 2010, pp. 11:1–11:17, 2010.

Antonio Canclini was born in Sondalo (Italy) on December 19, 1983. He received the M.Sc. degree (cum laude) in computer science and engineering (2008) and the Ph.D. degree (cum laude) in information technology (2012) from the Politecnico di Milano, Italy. He is currently a Post-Doc Researcher at the Image and Sound Processing Group, Politecnico di Milano. His research focuses on space-time audio signal processing (localization of acoustic sources and reflectors, spatial sound field rendering).

Paolo Bestagini was born in Novara (Italy) on February 22, 1986. He received the M.Sc. in telecommunications engineering and the Ph.D. in information technology at the Politecnico di Milano in 2010 and 2014, respectively. He is currently a Post-Doc researcher at the Image and Sound Processing Group, Politecnico di Milano. His research activity is focused on multimedia forensics and acoustic signal processing for microphone arrays.

Fabio Antonacci was born in Bari (Italy) on July 26, 1979. He received the Laurea degree in 2004 in telecommunication engineering and Ph.D. in information engineering in 2008, both at Politecnico di Milano, Italy.

He is currently Assistant Professor at Politecnico di Milano. His research focuses on space-time processing of audio signals, for both speaker and microphone arrays (source localization, acoustic scene analysis, rendering of spatial sound) and on modeling of acoustic propagation (visibility-based beam tracing). He is author of more than 50 articles in proceedings of international conferences and on peer-reviewed journals.

He published approximately 40 articles in proceedings of international conferences and on peer-reviewed journals.

Marco Compagnoni was born in Tirano (Italy) on April 25, 1980. He received the Laurea (cum laude) degree in theoretical physics (Universitá di Milano Bicocca, 2004) and the Ph.D. degree in applied mathematics (Politecnico di Milano, 2009).

He is currently A Post-Doc Researcher at the Geometry-Algebra In Application (GAIA) group in Dipartimento di Matematica at Politecnico di Milano. His research interests are applied math, especially in the application of geometry in mathematical physics and in statistics, in image processing and pattern recognition and in space-time processing of audio signals.

Augusto Sarti (M'04) received the M.S. and the Ph.D. degrees in electronic engineering, both from the University of Padua, Italy, in 1988 and 1993, respectively, with research on nonlinear communication systems. His graduate studies included a joint graduate program with the University of California, Berkeley. In 1993, he joined the Politecnico di Milano, Milan, Italy, where he is currently an Associate Professor. In 2013, he also joined the University of California, Davis, as an Adjunct Professor.

His research interests are in the area of multimedia signal processing, with particular focus on sound analysis, processing and synthesis; space-time audio processing; geometrical acoustics; music information retrieval. He also worked on problems of multidimensional signal processing, image analysis and 3D vision; motion planning and nonlinear system theory.

He coauthored well over 200 scientific publications on international journals and congresses as well as numerous patents in the multimedia signal processing area. He coordinates the activities of the Musical Acoustics Lab and of the Sound and Music Computing Lab of the Politecnico di Milano. He promoted and coordinated or contributed to numerous (20+) European projects. He is a member of the IEEE Technical Committee on Audio and Acoustics Signal Processing, and Senior Editor of IEEE SIGNAL PROCESSING LETTERS. He was co-chairman of the 2005 Edition of the IEEE International Conf. Advanced Video and Signal based Surveillance (AVSS); Chairman of 2009 edition of the Digital Audio Effects conference, (DAFx); and in the organizing committees of numerous other conferences in the area of signal processing.

Stefano Tubaro (M'01) was born in Novara, Italy, in 1957. He received his Electronic Engineering degree at the Politecnico di Milano, Milano, Italy, in 1982. He then joined the Dipartimento di Elettronica e Informazione of the Politecnico di Milano, first as a Researcher of the National Research Council, then as an Associate Professor (1991) and finally as a Full Professor (2004).

He initially worked on problems related to speech analysis; motion estimation/compensation for video analysis/coding; and vector quantization applied to hybrid video coding. In the past few years, his research interests have focused on image and video analysis for the geometric and radiometric modeling of 3-D scenes; and advanced algorithms for video coding and sound processing. He has authored over 150 scientific publications on international journals and congresses. He co-authored two books on digital processing of video sequences. He also co-authored several patents on image processing techniques.

He coordinates the research activities of the Image and Sound Processing Group (ISPG) at the Dipartimento di Elettronica, Informazione e Bioingegneria of the Politecnico di Milano; which is involved in several research programs funded by industrial partners, the Italian Government, and by the European Commission. He has been involved in the IEEE Technical Committee of Multimedia Signal Processing (2005–2009), and is currently involved in that of image video and multidimensional signal processing.