Acoustic Source Localization With Distributed Asynchronous Microphone Networks

A. Canclini, F. Antonacci, A. Sarti, *Member, IEEE*, and S. Tubaro, *Member, IEEE*

***Abstract*—We propose a method for localizing an acoustic source with distributed microphone networks. Time Differences of Arrival (TDOAs) of signals pertaining the same sensor are estimated through Generalized Cross-Correlation. After a TDOA filtering stage that discards measurements that are potentially unreliable, source localization is performed by minimizing a fourth-order polynomial that combines hyperbolic constraints from multiple sensors. The algorithm turns to exhibit a significantly lower computational cost compared with state-of-the-art techniques, while retaining an excellent localization accuracy in fairly reverberant conditions.**

***Index Terms*—Source localization, distributed microphone arrays, hyperbolas intersection.**

I. Introduction

The problem of acoustic source localization has significantly evolved with technological needs. The literature, in fact, is rich with localization solutions that adapt to various operating conditions. Particularly interesting are those that add the range to the vector of unknowns, with the result of linearizing the source localization problem and improving both accuracy and computational efficiency [1]–[3]. In the past few years, however, there has been a growing interest for spatial distributions of independent (unsynchronized) acoustic sensors, each made of two or more synchronized microphones. Examples of solutions in the literature first compute the Global Coherence Field (GCF) [4] or Steered Response Power (SRP) [5] maps associated to all the microphone pairs over a spatial grid and then localize the source as the peak of the cumulative global map, with overall computational costs that are often too demanding for the application at hand. Better computational efficiency is achieved in [6] where the SRP algorithm accommodates a different computation over a coarser grid. Alternate approaches based on Least Squares (LS) were proposed in [7], which proved efficient for compact arrays but with a certain sensitivity to environmental noise; and in [8] where a Stochastic Region Contraction of the grid was proposed, adopting a multi-resolution approach.

In this article we propose a novel solution that is suitable for spatial distributions of sensors and requires a modest computational load without giving up on localization accuracy and robustness. We consider spatially distributed sensors, each with $L \geq 2$ synchronized microphones. Different sensors are assumed as independently clocked and placed in space in an unconstrained but known fashion. Their location, in fact, can be accurately estimated using any of the self-calibration methods that are available in the literature [9]–[12]. The reduction in the computational cost of localization, which is important for balancing the distribution of computational load among sensors of a network, is achieved by using Time Differences Of Arrival (TDOAs) between microphones of the same sensor. A preliminary removal of TDOA outliers is performed by computing a reliability index based on the ratio between direct-path and reverberant components of the GCC. Surviving TDOAs are then turned into geometric constraints on the source location (hyperbolas, whose foci are on the microphone locations) at a node demanded with the localization (central node). A novel global cost function is then defined, which combines such constraints at best. Localization is then performed by minimizing the corresponding fourth-order polynomial.

In order to test the robustness of the system against measurement errors, in this paper we use the error propagation analysis introduced in [13] for theoretically characterizing the performance that can be achieved with a given configuration of sensors. Robustness against this uncertainty is tested using Monte Carlo simulations as well, after preliminary self-calibration based on [12].

The paper is structured as follows: Section II introduces the problem and the related notation. Section III describes how TDOAs are converted into hyperbolic constraints. Section IV illustrates the localization technique and assesses its computational cost. Section V describes the simulation setup and the results obtained.

II. Notation and Computation of TDOAs

For the sake of simplicity, in this paper we describe a method that is suitable for 2D geometries. A generalization to the 3D case, however, would be rather straightforward and would not result in a significant growth of computational cost. In the case of grid-based localization techniques such as GCF and LS, a generalization to the 3D case would result in a significantly higher computational cost.

Let us consider a spatial distribution of M sensors, each accommodating $L \geq 2$ microphones. Let $\boldsymbol{x}_l^m = (x_l^m, y_l^m)^T$, $l = 1, \dots L$, $m = 1, \dots M$, be the coordinates of each microphone, and $s_l^m(n)$ be the corresponding acquired signal. For the sake of notational simplicity, and with no loss of generality, let us assume that each sensor has the same number L of microphones. We also assume that the signals acquired by the microphones of the same sensor are synchronized, while no such assumption is made between different sensors. The acoustic source is located at $\boldsymbol{x}_s = (x_s, y_s)^T$.

The localization is performed over chunks of data extracted with a length-W rectangular window. Over the duration of this window, we assume that the source will not move significantly. The GCC-PHATs [4], [6], [8] $\Gamma_{lj}^m(n)$ between all possible pairs of microphones are computed on the same chunk of data. In order to keep the computational cost down, data processing is restricted to only when the source is found to be active, using the method in [14]. Given $\Gamma_{lj}^m(n)$, the discrete Time Difference Of Arrival is estimated as [4]

$$\hat{n}_{lj}^m = \arg\max_n \Gamma_{lj}^m(n) \quad \text{s.t.} \quad n \in [-n_{\max}, n_{\max}]$$

where the maximum TDOA $n_{\max}$ is determined by the distance between the two microphones. Notice that, in principle, for each sensor we do not need to compute more than $L(L-1)/2$ GCCs. According to the outlined application scenario, however, the computational power available at each sensor might be limited. The localization algorithm proposed in this paper is therefore flexible enough to accommodate a variable number of TDOAs.

In the presence of reverberations, some of the peaks in $\Gamma_{lj}^m(n)$ related to reflective paths might have a magnitude that turns out to be comparable to or larger than that of the direct path. In this situation we expect the estimation of the direct-path discrete TDOA to be unreliable. For this reason, we propose a reliability index that computes the

Manuscript received April 13, 2012; revised June 22, 2012; accepted August 11, 2012. Date of publication August 28, 2012; date of current version nulldate. This work was supported by the EU FET-Open SCENIC project under GA 226007. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Boaz Rafaely.

The authors are with the Dipartimento di Elettronica ed Informazione, Politecnico di Milano, 20133, Milan, Italy (e-mail: canclini@elet.polimi.it; antonacci@elet.polimi.it; sarti@elet.polimi.ittubaro@elet.polimi.it).

Digital Object Identifier 10.1109/TASL.2012.2215601

ratio between the energy in a window containing the highest peak and the energy of the remaining samples in $\Gamma_{lj}^m(n)$. In particular,

$$\mathrm{R}_{lj}^m = \frac{\sum_{n\in\mathcal{D}_{lj}^m}\Gamma_{lj}^m(n)^2}{\sum_{n\notin\mathcal{D}_{lj}^m}\Gamma_{lj}^m(n)^2}, \quad n\in[-n_{\max}, n_{\max}],$$

where $\mathcal{D}_{lj}^m = [\hat{n}_{lj}^m - n_{\mathrm{D}}, \hat{n}_{lj}^m + n_{\mathrm{D}}]$ is an interval around the highest peak in the GCC of extension $2n_{\mathrm{D}}+1$, in which the energy of the direct path is expected to be comprised. For each microphone pair at sensor m, the values R_{lj}^m are transferred to the central node, along with the corresponding discrete TDOA $\hat{n}_{lj}^m$. The measurement $\hat{n}_{lj}^m$ is considered reliable if $\mathrm{R}_{lj}^m \geq \kappa\bar{\mathrm{R}}$, where $\bar{\mathrm{R}}$ is the average R for all the TDOAs and κ is a threshold value. The optimal value of κ is a trade-off between i) retaining only the most reliable measures (which happens for high values of κ) and ii) exploiting all the TDOAs, even if inaccurate (which is the case for $\kappa = 0$). In our experiments we used $\kappa = 0.9$.

Discrete TDOAs are then converted into TDOAs through $\hat{\tau}_{lj}^m = (\hat{n}_{lj}^m)/(f_s)$, f_s being the sample frequency.

The central node is also preliminarily informed about the position $(x_l^m, y_l^m)^T$ of the individual microphones. Alternatively, their locations could be estimated during the initialization of the system using algorithms such as [12], [11].

In order to localize the source the central node first converts measurements into constraints in which the source position appears as an unknown (Section III), and then localizes the source through the combination of multiple constraints (Section IV).

III. From Measurements to Geometric Constraints

Let us consider two microphones of sensor m, which are placed in $\boldsymbol{x}_l^m$ and $\boldsymbol{x}_j^m$; and a single source in the unknown location $\boldsymbol{x}_s$. The TDOA $\hat{\tau}_{lj}^m$ measured from signals acquired by the microphone pair constrains the source to lie on a hyperbola whose foci are $\boldsymbol{x}_l^m$ and $\boldsymbol{x}_j^m$ and whose vertices are $T_{lj}^m = c\hat{\tau}_{lj}^m$ apart from each other, c being the sound speed. The generic expression for the hyperbola that corresponds to such measurements is

$$a_{lj}^m x^2 + b_{lj}^m xy + c_{lj}^m y^2 + d_{lj}^m x + e_{lj}^m y + f_{lj}^m = 0, \tag{1}$$

which depends on six parameters, and it is satisfied by all the points $\boldsymbol{x} = (x, y)^T$ lying on the hyperbola. This hyperbola can be rewritten in terms of the positions of the microphones; and of the TDOA $\hat{\tau}_{lj}^m$:

$$\sqrt{(x_l^m - x)^2 + (y_l^m - y)^2} - \sqrt{(x_j^m - x)^2 + (y_j^m - y)^2} = T_{lj}^m. \tag{2}$$

If we move the second square root to the right-hand side of (2), we square twice both members and we compare the resulting equation with (1), we finally obtain

$$\begin{aligned}
a_{lj}^m &= -4\left[\left(T_{lj}^m\right)^2 - \left(x_j^m - x_l^m\right)^2\right] \\
b_{lj}^m &= 8\left(x_j^m - x_l^m\right)\left(y_j^m - y_l^m\right) \\
c_{lj}^m &= -4\left[\left(T_{lj}^m\right)^2 - \left(y_j^m - y_l^m\right)^2\right] \\
d_{lj}^m &= 4\left\{\left(T_{lj}^m\right)^2\left(x_j^m + x_l^m\right) - \left(x_j^m - x_l^m\right)\right. \\
&\quad \left.\times\left[\left(x_j^m\right)^2 + \left(y_j^m\right)^2 - \left(x_l^m\right)^2 - \left(y_l^m\right)^2\right]\right\} \\
e_{lj}^m &= 4\left\{\left(T_{lj}^m\right)^2\left(y_j^m + y_l^m\right) - \left(y_j^m - y_l^m\right)\right. \\
&\quad \left.\times\left[\left(x_j^m\right)^2 + \left(y_j^m\right)^2 - \left(x_l^m\right)^2 - \left(y_l^m\right)^2\right]\right\} \\
f_{lj}^m &= \left(T_{lj}^m\right)^4 - 2\left(T_{lj}^m\right)^2\left[\left(x_j^m\right)^2 + \left(y_j^m\right)^2 + \left(x_l^m\right)^2 + \left(y_l^m\right)^2\right] \\
&\quad + \left[\left(x_j^m\right)^2 + \left(y_j^m\right)^2 - \left(x_l^m\right)^2 - \left(y_l^m\right)^2\right]^2.
\end{aligned}$$

IV. Source Localization

In this Section we discuss the problem of localizing the acoustic source from the combination of multiple constraints. Individual hyperbolas are combined in a global cost function, whose minimum corresponds to the searched source location. The cost function, however, is non-linear. An iterative minimization is therefore in order. An evaluation of the computational cost is also provided for reasons of comparison with other localization techniques.

A. Combination of Multiple Constraints

We assume that the source moves slowly enough that TDOAs coming from different sensors will be referred to the same position of the source (within the window of observation). If measurements are not affected by noise, the source position is determined as the solution of an overconstrained system of equations. In a real scenario, however, noise causes independent pairs of hyperbolas to intersect at multiple points. We estimate the source location $\boldsymbol{x}_s$ as the global minimum of a cost function $J(\boldsymbol{x}), \boldsymbol{x} = (x, y)^T$, obtained by combining the individual constraints. We define

$$\varepsilon_{lj}^m(\boldsymbol{x}) = v_{lj}^m\left[a_{lj}^m x^2 + b_{lj}^m xy + c_{lj}^m y^2 + d_{lj}^m x + e_{lj}^m y + f_{lj}^m\right],$$

where v_{lj}^m assumes the value of 1 for all the reliable TDOAs and 0 for the unreliable ones. We stack all the elements $\varepsilon_{lj}^m(\boldsymbol{x})$ in the column vector $\boldsymbol{\varepsilon}(\boldsymbol{x})$. The source position is then estimated as

$$\hat{\boldsymbol{x}}_s = \arg\min_x J(\boldsymbol{x}), \quad J(\boldsymbol{x}) = \boldsymbol{\varepsilon}(\boldsymbol{x})^T\boldsymbol{\varepsilon}(\boldsymbol{x}). \tag{3}$$

The resulting cost function is a fourth order polynomial, to minimize which we need to adopt an iterative technique. In order to efficiently minimize the cost function we adopt a Taylor series expansion of $\boldsymbol{\varepsilon}(\boldsymbol{x})$ about the initial point $\boldsymbol{x}_{s,0}$ as

$$\boldsymbol{\varepsilon}(\boldsymbol{x}) \simeq \boldsymbol{\varepsilon}(\boldsymbol{x}_{s,0}) + \nabla\boldsymbol{\varepsilon}\big|_{\boldsymbol{x}_{s,0}} \cdot (\boldsymbol{x} - \boldsymbol{x}_{s,0}), \tag{4}$$

where $\boldsymbol{x}_{s,0} = (x_{s,0}, y_{s,0})^T$ is the initial guess of the source location, and

$$\nabla\boldsymbol{\varepsilon} = \begin{bmatrix} \frac{\partial\varepsilon_{1,2}^1}{\partial x} & \frac{\partial\varepsilon_{1,2}^1}{\partial y} \\ \vdots & \vdots \\ \frac{\partial\varepsilon_{L-1,L}^M}{\partial x} & \frac{\partial\varepsilon_{L-1,L}^M}{\partial y} \end{bmatrix}. \tag{5}$$

If we use (4) in (3) we obtain the update equation of the iterative minimization procedure

$$\hat{\boldsymbol{x}}_{s,i+1} = \hat{\boldsymbol{x}}_{s,i} - \nabla\boldsymbol{\varepsilon}^\dagger\Big|_{\hat{\boldsymbol{x}}_{s,i}}\boldsymbol{\varepsilon}(\boldsymbol{x}_{s,i}), \tag{6}$$

where $\nabla\boldsymbol{\varepsilon}^\dagger|_{\hat{\boldsymbol{x}}_{s,i}}$ is the Moore-Penrose pseudoinverse of $\nabla\boldsymbol{\varepsilon}|_{\hat{\boldsymbol{x}}_{s,i}}$; and the symbol i is the iteration number. We assume $\hat{\boldsymbol{x}}_s = \hat{\boldsymbol{x}}_{s,i+1}$ when $\|\hat{\boldsymbol{x}}_{s,i+1} - \hat{\boldsymbol{x}}_{s,i}\|$ is smaller or equal to a given threshold. In the area enclosed by sensors the cost function rarely exhibits multiple minima, according to the experiments conducted for the paper. The choice of the initial point, therefore, is not critical. In our simulations, we have set the initial point to be in the center of the area of interest.

B. Computational Cost

While algorithms such as SRP and LS are based on a grid-search optimization, the proposed algorithm is based on an iterative minimization, with a clear advantage on computational cost. In this subsection we aim at providing an upper-bound estimation of the cost, in terms of multiplication and accesses to the memory of the proposed technique and of an improved version of SRP. Computations are split between sensors and central node. As far as sensors are concerned, there is no

difference between the proposed technique and GCF/SRP, as both require the computation of the GCC. The additional operations entailed by the computation of the reliability index are marginal. More interesting is, instead, the comparison of the computational costs at the central node. We identify three steps in the proposed localization procedure: filtering of the TDOAs, conversion from measurements to hyperbolas and, finally, source localization. For each of these three steps we estimate the number of multiplications with accumulation (MACs) and of the accesses to the memory (AMs). In particular, we assume that no measurement is discarded in the TDOA filtering process. Let $F = ML(L-1)/2$ be the number of TDOAs and k the number of iterations of the optimization algorithm.

- **Filtering of the TDOAs:** $F + \log F$ MACs and $F + \log F$ AMs;
- **Conversion from TDOAs to hyperbolas**: $36F$ MACs and $23F$ AMs;
- **Source localization**: $53Fk$ MACs and $35Fk$ AMs.

According to the above scheme, the proposed methodology entails $C_{\mathrm{h}} = \log F + 37F + 53Fk$ MACs.

On the other side SRP techniques are based on the evaluation on a grid of a kernel function and on a following maximization. Each evaluation of the kernel function involves the following operations

- three multiplications and a square root to assess the theoretical range difference of the point on the grid. These operations involve approximately $4F$ MACs and $2F$AMs;
- conversion from range difference to TDOA, which involves F MACs;
- accumulation on the grid, which involves 0 MACs (mo multiplications are needed but only an accumulation) and F AMs.

The above operations are repeated for all the points of the grid. The computational cost of the method in [8], in terms of MACs is $C_{\mathrm{s}} = 4FG_{\mathrm{s}}$ MACs, where $G_{\mathrm{s}} = \sum_i J_i$ is the total number of evaluations of the kernel function for all the steps of the multi-resolution process. As for the technique in [6], the number of MACs required for the localization is $C_{\mathrm{m}} = 4FG_{\mathrm{m}}$, where G_{m} is the number of points of the search grid.

V. Evaluation

In this section we prove the effectiveness of the proposed localization algorithm, first showing the accuracy in the localization when the TDOAs are assumed to be corrupted by a zero-mean Gaussian error, independently distributed at each microphone pair. The results were predicted in a theoretical fashion by means of the Error Propagation Toolbox [15], which implements the error propagation analysis presented in [13]. After that, we report the results of an extensive simulation campaign in a reverberant room, organized in two acquisitions. First, the sensor positions were assumed to be known in advance. In the second case, the position of the sensors has been estimated through the self-calibration technique described in [12], thus introducing an uncertainty on the location of the microphones. The results are compared with those of the SRP-SRC technique [8] and of the SRP-Modified technique [6], obtained under the same operating conditions.

A. Simulation Setup

We localized the source using $M = 4$ identical sensors distributed on each side of a 4 m × 5 m rectangular room. Each sensor had $L = 3$ microphones. Using the simulation setup of Fig. 1, we tested 165 source locations, distributed on a rectangular grid. The source produced 10 s of a phonetically rich speech signal, sampled at 44.1 kHz. Wave propagation was simulated by estimating the room impulse responses between each source-microphone pair using a beam tracing method [16], using the specified wall reflection coefficient β. Microphone signals have been obtained as the convolution of the source signal with the relative impulse responses. Moreover, in order to reproduce more realistic conditions, microphone signals have been corrupted with additive zero-mean Gaussian noise, whose standard deviation determines the Signal-To-Noise Ratio (SNR). Source localization is performed on audio segments of 100 ms length. The parameter n_{D} used for the computation of the reliability index was set to 3.

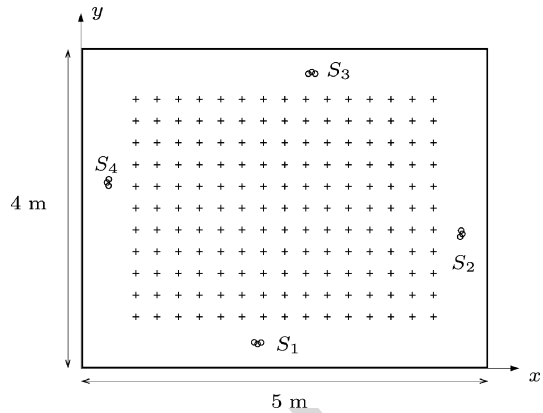

Fig. 1. Simulation setup: four sensors $(S_1, \ldots S_4)$ are used for localizing an acoustic source in a rectangular room; the source occupies 165 test locations, denoted by the symbol $(+)$, disposed on a regular grid.

The internal geometry of the sensors is given by their nominal coordinates in the sensor reference system, i.e., $\bar{\boldsymbol{x}}_1 = (-0.05\,\mathrm{m}, 0\,\mathrm{m})^T$, $\bar{\boldsymbol{x}}_2 = (0\,\mathrm{m}, 0.025\,\mathrm{m})^T$ and $\bar{\boldsymbol{x}}_3 = (0.05\,\mathrm{m}, 0\,\mathrm{m})^T$. The location of the microphones in the mth sensor (in the room reference system), are obtained through rotation and translation of the nominal coordinates, specified by the angle θ_m and by the vector $\boldsymbol{t}_m$, respectively. In particular $\theta_1 = -2.22°, \boldsymbol{t}_1 = (2.6\,\mathrm{m}, 0.4\,\mathrm{m})^T$; $\theta_2 = 81.55°, \boldsymbol{t}_2 = (5.6\,\mathrm{m}, 2.1\,\mathrm{m})^T$; $\theta_3 = 180.15°, \boldsymbol{t}_3 = (3.4\,\mathrm{m}, 4.6\,\mathrm{m})^T$; and $\theta_4 = 267.58°, \boldsymbol{t}_4 = (0.4\,\mathrm{m}, 2.9\,\mathrm{m})^T$.

As far as the SRP-SRC method is concerned, we have implemented the SRP-SRC-III search technique [8], which envisions a fixed number of points to be evaluated at each step of the multi-resolution optimization procedure. In particular, we used $\mathrm{S} = 5$ multi-resolution steps and at each step a grid of $N_{\mathrm{pts}} = 750$ points was defined. With these numbers at hand, at the initial step the grid has a resolution of 20 cm × 20 cm, while the average resolution at the last step of the optimization procedure is approximately of 1 cm × 1 cm. As far as the SRP-Modified method is concerned, the resolution of the grid was set to 4 cm × 4 cm.

B. Theoretical Analysis

In this paragraph we show the theoretical accuracy of the proposed localization algorithm, as predicted using the error propagation analysis [13]. The analysis has been conducted in the scenario of Fig. 1, setting the standard deviation of the error on TDOAs to $\sigma_\tau = 4.4\ \mu$s. This value has been estimated from preliminary tests considering a reflection coefficient of the walls equal to $\beta = 0.5$, corresponding to a moderately reverberant room; and a SNR of 10 dB. The results are shown in Fig. 2(a), which reports the Root Mean Square Error (RMSE) of the localization on the y coordinate, at all the tested source positions. For the sake of comparison, Fig. 2(b) shows the RMSE on the coordinate y relative to the results obtained from simulations. The theoretical analysis reveals that the best accuracy is achieved when the source is placed on the two black spots that are visible in Fig. 2(a) (close to the center of the region of interest). Conversely, the worst performance is reached when the source is located near the corners of the area. We notice that, although a Gaussian distribution of the error on TDOAs is far from being realistic in practice, the theoretical results give a preliminary idea on the accuracy that can be achieved by the presented method. In fact, this theoretical prediction is confirmed by simulative

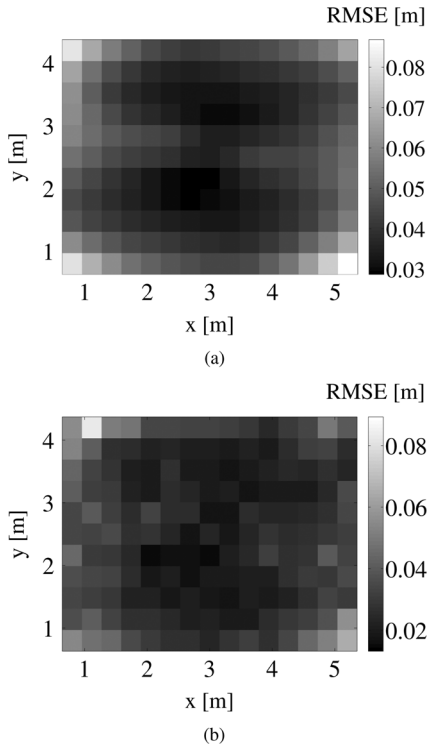

Fig. 2. Comparison between theoretical error propagation and simulative results. Results are relative to the y coordinate. (a) Theoretical analysis, (b) Simulation results.

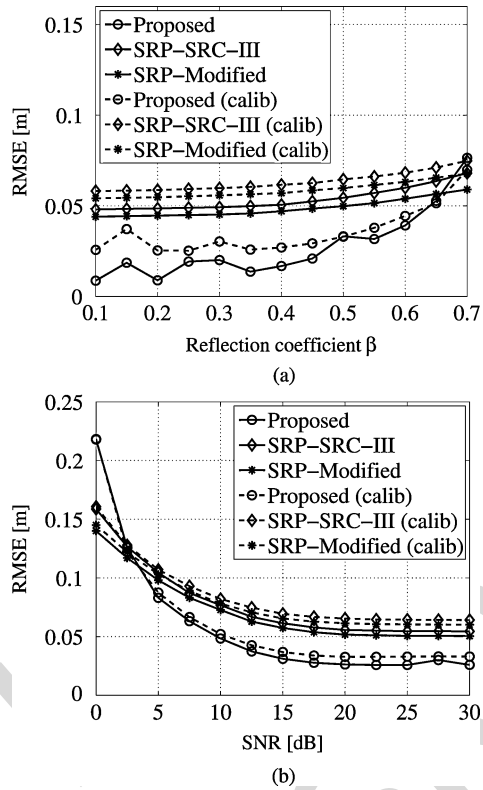

Fig. 3. Localization error of the proposed algorithm and of the SRP-SRC-III method. (a) RMSE vs refl. coeff. (b) RMSE vs SNR.

results in Fig. 2(b), which exhibit the same two black spots, as well as the same bright regions near the corners.

C. Simulation Results

The three algorithms are evaluated in terms of the Euclidean distance (localization error) between the estimated and the reference positions of the source. In particular, we tested the accuracy of the algorithm when sensors locations are known or have been preliminarily estimated through the self-calibration technique described in [12].

As a first test, we evaluate the robustness of the localization against reverberation. In particular, we measure the localization error as a function of the reflection coefficient β of the walls. The SNR is kept fixed to 25 dB on all the microphones. Fig. 3(a) shows the RMSE, obtained averaging the localization errors of all the audio segments, at every source position. Continuous lines refer to the case of known sensor locations, dashed lines to estimated ones. We observe that in both scenarios the proposed algorithm exhibits higher accuracy with respect to the SRP-SRC-III and SRP-Modified methods when the reverberation is moderate ($\beta < 0.5$), presenting an RMSE around 2 cm. The accuracy tends to degrade for $\beta > 0.5$, approaching that of the SRP-based techniques when $\beta = 0.7$ (medium-high reverberation), with an RMSE about 7 cm. The performance reduction suffered by the proposed technique and by the SRP-based methods when sensor locations are not known is similar.

As a second test, we evaluate the localization algorithm for different SNRs at the microphone signals. The reflection coefficient is fixed to $\beta = 0.5$. The resulting RMSEs for the three algorithms are shown in Fig. 3(b). Again, continuous lines refer to the case of known sensor locations, dashed lines to estimates ones. We notice that for SNRs higher than 10 dB the RMSE is almost constant around 3 cm and 6 cm for the proposed and for the SRP-based methods, respectively. Low SNR values cause a noticeable degradation of all the algorithms. In the extreme case of a 0 dB SNR, the RMSE is of 16 cm for SRP-SRC-III; 14 cm for SRP-Modified; and 22 cm for the proposed technique. As for the previous experiment, the error introduced by the preliminary self-localization of sensors does not affect the overall trend.

Finally, we analyze the spatial distribution of the error, over the grid of the test source positions. For this specific experiment, the sensor locations are known in advance. Figs. 4(a) and (b) show the RMSE for the proposed and the SRP-SRC-III techniques, respectively, having fixed the reflection coefficient to $\beta = 0.5$ and the SNR to 10 dB. From Fig. 4(a) we observe that the proposed method exhibits better accuracy for sources located at the center of the room, while the error tends to increase as the source moves away from the center position. The worst cases are represented by sources located at the corners of the grid. On the other hand, Fig. 4(b) highlights that the proposed method overcomes SRP-SRC-III in a wide area at the center of the test environment, while attaining similar accuracy at the corners.

D. Computational Cost

In this paragraph we compare the computational cost of the proposed localization algorithm, of SRP-SRC-III and of SRP-Modified. All the tests were performed on a laptop equipped with a 3 GHz dual-core processor and 4 GB of RAM. The reflection coefficient is variable between 0.3 and 0.7 and the SNR is 15 dB. Note that the computational cost of SRP-SRC-III and SRP-Modified does not vary as the reflection coefficient increases, while it could be variable for the proposed methodology, due to its iterative nature. Table I shows the details of the computational cost for the three algorithms under comparison. More specifically, for each reflection coefficient the average number of iterations $\bar{k}$ and the theoretical computational costs C_{h}, C_{s} and C_{m} are reported. Moreover, we show also the computational times $\bar{\mathrm{T}}_{\mathrm{h}}$, $\bar{\mathrm{T}}_{\mathrm{s}}$, and $\bar{\mathrm{T}}_{\mathrm{m}}$ of the three algorithms under comparison. Finally, we also show the ratio among the theoretical computational costs and the ratio among the actual computational costs of the three techniques. Notice that there

TABLE I
COMPUTATIONAL COSTS FOR THE PROPOSED METHODOLOGY AND FOR THE SRP-BASED TECHNIQUES

	Proposed			SRP-SRC III		SRP-Modified		Comparison			
β	$\bar{k}$	C_{h}	$\overline{\mathrm{T}}_{\mathrm{h}}[\mathrm{ms}]$	C_{s}	$\overline{\mathrm{T}}_{\mathrm{s}}[\mathrm{ms}]$	C_{m}	$\overline{\mathrm{T}}_{\mathrm{m}}[\mathrm{ms}]$	$\frac{C_{\mathrm{s}}}{C_{\mathrm{h}}}$	$\frac{\overline{\mathrm{T}}_{\mathrm{s}}[\mathrm{ms}]}{\overline{\mathrm{T}}_{\mathrm{h}}[\mathrm{ms}]}$	$\frac{C_{\mathrm{m}}}{C_{\mathrm{h}}}$	$\frac{\overline{\mathrm{T}}_{\mathrm{m}}[\mathrm{ms}]}{\overline{\mathrm{T}}_{\mathrm{h}}[\mathrm{ms}]}$
0,3	6,83	4379	14,8	166500	551	145000	445	37,6	37,2	30,6	30,1
0,4	6,92	4432	14,9	166500	549	145000	449	37,1	37	32,7	30,1
0,5	7,02	4489	15,0	166500	553	145000	452	36,7	36,8	32,3	30,1
0,6	7,06	4513	15,1	166500	556	145000	450	35,9	36,8	32,1	29,8
0,7	7,19	4590	15,2	166500	554	145000	455	34,7	36,3	31,6	29,9
0,8	7,45	4739	15,2	166500	580	145000	463	35,14	36,4	30,6	30,5

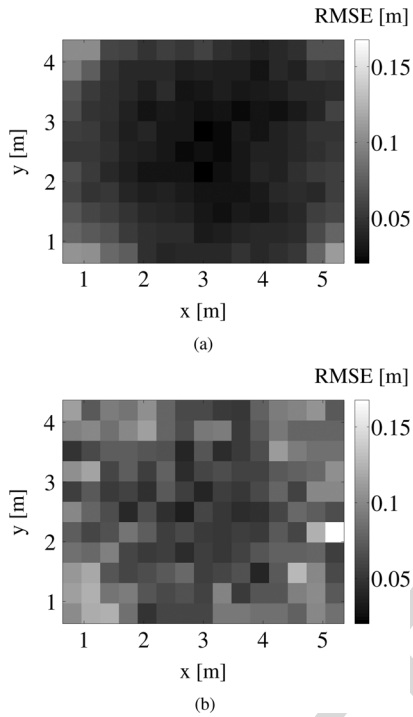

Fig. 4. Localization error as a function of the source position. Microphone positions were known in advance. (a) Proposed. (b) SRP-SRC-III.

is a good match between $(\bar{\mathrm{T}}_{\mathrm{s}}[\mathrm{ms}])/(\bar{\mathrm{T}}_{\mathrm{h}}[\mathrm{ms}])$ and $(C_{\mathrm{s}})/(C_{\mathrm{h}})$; and between and $(\bar{\mathrm{T}}_{\mathrm{m}}[\mathrm{ms}])/(\bar{\mathrm{T}}_{\mathrm{h}}[\mathrm{ms}])$ and $(C_{\mathrm{m}})/(C_{\mathrm{h}})$. The average number of iterations $\bar{k}$ ranges from 6.82 to 7.45. However, the same increase is not observed in the computational time. In fact, when reverberations are present, the actual number of TDOAs used for the localization decreases, as some of them are discarded by the TDOA filtering module.

VI. CONCLUSION

In this paper we proposed a novel algorithm for the localization of an acoustic source, using distributed asynchronous sensors. The method is based on the hyperbolic constraints associated to the Time Differences of Arrival that are measured on microphone pairs at local nodes. Source localization is then based on the minimization of a global cost function that combines the individual constraints. We showed, both with an analytic approach and through simulation, that the proposed algorithm is very efficient both in terms of computation time and memory consumption. For this reason, it turns out to be suitable for devices with limited processing power and memory requirements (e.g., smartphones, tablets, etc.). Simulations proved the robustness of the algorithm against reverberation and noise, even when sensor positions are only approximately known. This suggests that the proposed method could be fruitfully employed in scenarios where multiple sensors are arbitrarily scattered in space, and their location is estimated through a self-calibration procedure.

REFERENCES

[1] J. Benesty, Y. Huang, and J. Chen, "Time delay estimation via minimum entropy," *IEEE Signal Process. Lett.*, vol. 14, no. 3, pp. 157–160, Mar. 2007.
[2] J. Abel and J. Smith, "The spherical interpolation method for closed-form passive source localization using range difference measurements," in *Proc. IEEE Int. Acoust., Speech, Signal Process. Conf. (ICASSP'87)*, 1987, vol. 12, pp. 471–474.
[3] 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.
[4] M. Omologo and P. Svaizer, "Acoustic event localization using a crosspower-spectrum phase based technique," in *Proc. IEEE Int. Acoust., Speech, Signal Process. Conf. (ICASSP-94)*, 1994, pp. 273–276.
[5] M. S. Brandstein, J. E. Adcock, and H. F. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 1, pp. 45–50, Jan. 1997.
[6] M. Cobos, A. Marti, and J. J. Lopez, "A modified SRP-PHAT functional for robust real-time sound source localization with scalable spatial sampling," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 71–74, 2011.
[7] D. Rabinkin, R. Renomeron, A. Dahl, J. French, and J. Flanagan, "A DSP implementation of source location using microphone arrays," in *Proc. 131st Meeting Acoust. Soc. Amer.*, 1996.
[8] H. Do, H. F. Silverman, and Y. Yu, "A real-time SRP-PHAT source location implementation using stochastic region contraction(src) on a large-aperture microphone array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '07)*, 2007, vol. 1.
[9] A. Redondi, M. Tagliasacchi, F. Antonacci, and A. Sarti, "Geometric calibration of distributed microphone arrays," in *Proc. IEEE Int. Workshop Multimedia Signal Process. (MMSP'09)*, 2009, pp. 1–5.
[10] S. D. Valente, M. Tagliasacchi, F. Antonacci, P. Bestagini, A. Sarti, and S. Tubaro, "Geometric calibration of distributed microphone arrays from acoustic source correspondences," in *Proc. IEEE Int. Multimedia Signal Process. (MMSP) Workshop*, 2010, pp. 13–18.
[11] M. H. Hennecke and G. A. Fink, "Towards acoustic self-localization of ad hoc smartphone arrays," in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays (HSCMA)*, 2011, pp. 127–132.
[12] P. Pertila, M. Mieskolainen, and M. S. Hamalainen, "Closed-form self-localization of asynchronous microphone arrays," in *Proc. Joint Workshop Hands-Free Speech Commun. Microphone Arrays (HSCMA)*, 2011, pp. 139–144.
[13] 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.
[14] A. Brutti, M. Omologo, and P. Svaizer, "Multiple source localization based on acoustic map de-emphasis," *EURASIP J. Audio Speech Music Process.* vol. 2010, pp. 11:1–11:17, Jan. 2010 [Online]. Available: http://dx.doi.org/10.1155/2010/147495
[15] Error Propagation Toolbox [Online]. Available: http://www.thescenicproject.eu/resources/software/EPT.zip
[16] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, "Fast tracing of acoustic beams and paths through visibility lookup," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 812–824, May 2008.