A Comparative Study of Multilateration Methods for Single-Source Localization in Distributed Audio

Srđan Kitić, Clément Gaultier, Grégory Pallone
Orange Labs
Cesson-Sévigné, France
srdan.kitic, clement.gaultier, gregory.pallone@orange.com

***Abstract*—In this article we analyze the state-of-the-art in multilateration - the family of localization methods enabled by the range difference observations. These methods are computationally efficient, signal-independent, and flexible with regards to the number of sensing nodes and their spatial arrangement. However, the multilateration problem does not admit a closed-form solution in the general case, and the localization performance is conditioned on the accuracy of range difference estimates. For that reason, we consider a simplified use case where multiple distributed microphones capture the signal coming from a near field sound source, and discuss their robustness to the estimation errors. In addition to surveying the relevant bibliography, we present the results of a small-scale benchmark of few "mainstream" multilateration algorithms, based on an in-house Room Impulse Response dataset.**

I. Introduction

As the audio technologies incorporating distributed acoustic sensing – like Internet Of Audio Things [1] – gain momentum, the questions regarding efficient exploitation of such acquired data naturally arise. A valuable information that could be provided by these networks is the location of the sound source, which can be a daunting task in the adverse acoustic conditions, namely in the presence of noise and reverberation. On the flipside, localization is usually only a pre-processing block of a larger processing chain (*e.g.* in the case of location-guided separation and enhancement [2]), thus its computational efficiency is of uttermost importance.

In this work we consider a specific *distributed audio context*, where we assume a sensor network composed of distributed single-channel microphones with potentially different gains (Fig.1). Such a network could be seen as one large scale microphone array - note that this is markedly different from a network whose nodes are (compact) microphone arrays themselves, as detailed in the following paragraph. The array geometry is assumed known in advance, and the microphones are already synchronized/syntonized [3]. Lastly, we assume the presence of a single sound source and a direct path (line-of-sight) between the source and microphones. The latter assumption is essential for most of the traditional sound source localization methods to work, in order to avoid an extremely challenging "hearing behind walls" problem [4].

Albeit deceivingly simple, the considered scenario imposes several technical constraints. First, the absence of compact arrays prevents the node-level Direction-of-Arrival (DOA) estimation. Second, no knowledge of the source emission time prohibits the Time-of-Flight (TOF) estimation. Third, large array size implies significant spatial aliasing, which, along with the relatively small number of microphones, seriously degrades performance of beamforming-based techniques, at least in narrowband [5]. The approaches based on *distributed* beamforming, *e.g.* [6], [7], [8], could still be appealing if they operate in the wideband regime: unfortunately, the literature on wideband beamforming by distributed mono microphones is somewhat scarce. Another major downside of beamforming-based localization is generally high computational cost, although various attempts have been made in order to reduce its complexity, *e.g.* [9]. Finally, the fact that the number of sensors and their spatial arrangement can vary precludes the use of contemporary learning-based localization methods (*e.g.* [10], [11], [12]), which have shown remarkable performance in more restricted use cases.

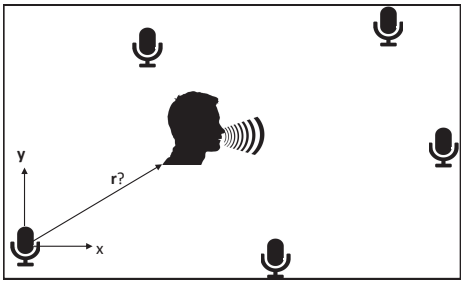


Fig. 1. Source localization with 5 single-channel microphones

Under these constraints, the pairwise Time Difference Of Arrival (TDOA) features emerge as a viable choice for sound source localization (rectifying the need for adequate synchronization, since the TDOA estimation is known to be sensitive to clock offsets and internal delays [13]). The corresponding sound source localization pipeline is presented in Fig. 2. First, TDOAs are estimated for each microphone pair, followed by their conversion to (pseudo) Range Differences (RD). These estimates are then processed by a *multilateration* algorithm [14], which finally yields the source position. In this article we focus on the last part of the pipeline, *i.e.* we discuss exclusively and thoroughly the localization algorithms belonging to the multilateration class, as opposed to related review papers [14], [15], [16] that study localization methods in a broader sense.

Simultaneous localization of multiple sound sources is of a great practical interest. From the perspective of a multilateration algorithm, as long as multiple *sets* of RDs (corresponding to each sound source) are available, the localization of each source can be done independently of the rest. Therefore, discussing the single-source case only does not endure a loss

ISSN 2305-7254

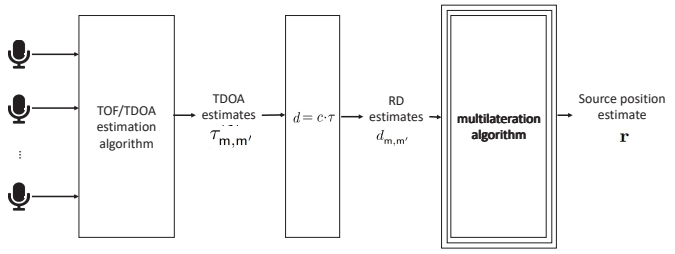


Fig. 2. Standard multilateration processing chain (notations are defined in section II)

of generality with regards to multilateration-based localization. However, we underline that the performance of multilateration methods is conditioned on the accuracy of TDOA/RD estimation, which is a difficult problem in its own right [17], [18], [19]. The problem becomes even more challenging in the presence of multiple overlapping sources [20], but it falls out of the scope of the present article. Indeed, as practitioners we are particularly interested in performance of multilateration algorithms using the pseudo RDs obtained from off-the-shelf TDOA estimators, such as Generalized Cross Correlation - PHAse Transform (GCC-PHAT) [21].

II. Problem statement

Distances between microphones are considered to be of the same order as the distances between the microphones and the source, hence – with regards to the audible frequency range of speech – we discuss the near field scenario. The general formulation of the time-domain signal $y_{\mathsf{m}}(t)$, recorded at the m^{th} microphone is given by the time-variant convolution:

$$y_{\mathsf{m}}(t) = \int_0^t a_m(t,\tau) x_{\mathsf{s}}(t-\tau) d\tau + n_{\mathsf{m}}(t), \tag{1}$$

where $a_m(t,\tau)$ is the time-variant Room Impulse Response (RIR) filter, relating the m^{th} microphone position $\mathbf{r}_{\mathsf{m}}$ with the source position $\mathbf{r}_{\mathsf{s}}$, $x_{\mathsf{s}}(t)$ is the source signal, and $n_{\mathsf{m}}(t)$ is the additive noise of the considered microphone. In (1), the microphone gains are absorbed by RIRs. In practice, various simplifications are commonly used instead of the general expression (1). Commonly, a free-field, time-invariant approximation is adopted, as follows [22]:

$$y_{\mathsf{m}}(t) = a_m x_{\mathsf{s}}(t-\tau_{\mathsf{m}}) + n_{\mathsf{m}}(t), \tag{2}$$

where the offset τ_{m} denotes the TOF value, which is proportional to the source-microphone distance.

The TDOA, corresponding to the difference in propagation delay between the microphones m and m', is defined as $\tau_{\mathsf{m},\mathsf{m}'} = \tau_{\mathsf{m}'} - \tau_{\mathsf{m}}$. In homogeneous propagation media, the TDOA values $\tau_{\mathsf{m},\mathsf{m}'}$, directly translate into Range Differences (RD) $d_{\mathsf{m},\mathsf{m}'}$, given the sound speed c:

$$d_{\mathsf{m},\mathsf{m}'} := D_{\mathsf{m}'} - D_{\mathsf{m}} = \|\mathbf{r}_{\mathsf{m}'} - \mathbf{r}_{\mathsf{s}}\| - \|\mathbf{r}_{\mathsf{m}} - \mathbf{r}_{\mathsf{s}}\| = c \cdot \tau_{\mathsf{m},\mathsf{m}'}, \tag{3}$$

where D_{m} denotes the source-microphone distance. The observation model (3) defines the two-sheet hyperboloid with respect to $\mathbf{r}_{\mathsf{s}}$, with foci in $\mathbf{r}_{\mathsf{m}}$ and $\mathbf{r}_{\mathsf{m}'}$ [23], [24]. Note that the RDs could be easily determined from the TOF measurements as well (since $D_{\mathsf{m}} = c \cdot \tau_{\mathsf{m}}$), provided that the emission time is known. Surprisingly, despite theoretical evidence, a simulation study in [25] has revealed that the TOF and TDOA features seem to perform similarly in terms of localization accuracy.

Since the microphone signals are often corrupted by noise and reverberation, the TDOA measurements – thereby RDs – could be erroneous, which negatively affects the performance of localization algorithms. In the RD domain, such degradations are usually modeled by an additive noise term. Another cause of localization errors is the inexact knowledge of microphone positions. As shown in [26], the Cramér-Rao lower Bound (CRB) [27] of the source location estimate increases rather quickly with the increase in the microphone position "noise" (fortunately, somewhat less fast in the near field, than in the far field setting). Finally, the localization accuracy also depends on the array geometry [28], which is assumed arbitrary in our case.

In noiseless conditions, the number of linearly independent RD observations is equal to $\mathsf{M}-1$, and all the remaining RDs could be calculated from such a set. However, in the presence of noise, considering the full set of observations (of size $\mathsf{M}(\mathsf{M}-1)/2$) may be useful for alleviating the noise-related degradation [25], [29], [30], [31]. If only independent RDs are to be used, one microphone is chosen as a reference (the choice of which, under the same noise level, does not affect the theoretical localization accuracy [25]). The same microphone could be conveniently put at the coordinate origin, *e.g.* $\mathbf{r}_1 = \mathbf{0}$, where $\mathbf{0}$ is the null vector. By denoting $\mathbf{r} := \mathbf{r}_{\mathsf{s}}$, from (3), the non-redundant RDs are compactly given as

$$d_{\mathsf{m}'} := d_{1,\mathsf{m}'} = \|\mathbf{r}_{\mathsf{m}'} - \mathbf{r}\| - \|\mathbf{r}\|. \tag{4}$$

Finally, given a minimal $\{d_{\mathsf{m}'} |\ \mathsf{m}' \in [2,\mathsf{M}]\}$, or an extended $\{d_{\mathsf{m},\mathsf{m}'} |\ (\mathsf{m},\mathsf{m}') \in [1,\mathsf{M}] \times [1,\mathsf{M}], \mathsf{m} \neq \mathsf{m}'\}$ set of observations, along with microphone position $\{\mathbf{r}_{\mathsf{m}}\}$, the goal now is to estimate source position $\mathbf{r}$. In the following sections, we discuss two major classes of such localization algorithms: maximum likelihood and least squares estimators.

III. Maximum likelihood estimation

Since the observations (4) are non-linear, a statistically efficient estimate (*i.e.* the one that attains CRB) may not be available. The common approach is to seek the maximum likelihood (ML) estimator instead.

Let $\hat{\mathbf{r}}$ and $\hat{d}_{\mathsf{m}'}(\hat{\mathbf{r}})$ denote the estimated source position, and the corresponding RD, respectively:

$$\hat{d}_{\mathsf{m}'}(\hat{\mathbf{r}}) = \|\mathbf{r}_{\mathsf{m}'} - \hat{\mathbf{r}}\| - \|\hat{\mathbf{r}}\|.$$

Under the hypothesis that the observation noise is Gaussian, the ML estimator is given as the minimizer of the negative log-likelihood [32], [33]

$$\mathcal{L}(\mathbf{r}) = \left(\mathbf{d} - \hat{\mathbf{d}}(\hat{\mathbf{r}})\right)^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} \left(\mathbf{d} - \hat{\mathbf{d}}(\hat{\mathbf{r}})\right), \tag{5}$$

where $\mathbf{d} = [d_2\ d_3 \dots d_{\mathsf{M}}]^{\mathsf{T}}$, $\hat{\mathbf{d}}(\hat{\mathbf{r}}) = \left[\hat{d}_2(\hat{\mathbf{r}}),\ \hat{d}_3(\hat{\mathbf{r}}) \dots \hat{d}_{\mathsf{M}}(\hat{\mathbf{r}})\right]^{\mathsf{T}}$, and $\boldsymbol{\Sigma}$ is the covariance matrix of the measurement noise.

Note, however that the Gaussian noise assumption for the RD measurements may not hold. For instance, the digital quantization alone can induce RD errors on the order of 2 cm [34]. Moreover, the ML estimators are proven to attain the CRB in the asymptotic regime, while the number of microphones (*i.e.* the number of RDs) is often small. Therefore, non-statistical estimators, such as least squares, are often used in practice instead. Anyhow, in this section we discuss two families of methods proposed for the RD maximum likelihood estimation: the ones that aim at solving the non-convex problem (5) directly, and the ones based on convex relaxations. The former should not be confused for "direct" localization methods based on grid search, such as steered response power beamformer.

A. Direct methods

The problem (5) is difficult to solve directly, due to nonlinear dependence of the RDs $\{\hat{d}_{\mathsf{m}'}(\hat{\mathbf{r}})\}$ on the position variable $\hat{\mathbf{r}}$. Early approaches, based on iterative schemes, such as linearized gradient descent and Levenberg-Marquardt algorithm [35], [36], suffer from sensitivity to initialization, increased computational complexity and ill-conditioning (though the latter could be improved using regularization techniques [37]). The method proposed in [38] exploits correlation among noises within different RD measurements, and defines a constrained ML cost function tackled by a Newton-like algorithm. According to simulation results, it is more robust to adverse localization geometries [39], [28] than [35], or the least squares methods [40], [30], [41], [42]. Another advantage of this method is the straightforward way to provide the initial estimate (however, as usual, global convergence cannot be guaranteed).

In the pioneering article [32], the authors proposed a closed-form, two-stage approach, that approximates the solution of (5). Firstly, the (weighted) unconstrained least-squares solution (to be explained in the next section) is computed, which is then improved by exploiting the relation between the estimates of the position vector and its magnitude. The minimal number of microphones, due to the unconstrained LS estimation is 5 in three dimensions. It has been shown [32] that the method attains the CRB at high to moderate Signal-to-Noise-Ratios (SNRs). Unfortunately, it suffers from a nonlinear "threshold effect" - its performance quickly deteriorates at low SNRs. Instead, an approximate, but more stable version of this ML method has been proposed in [43]. In addition, the estimator [32] comes with a large bias [37], which cannot be reduced by increasing the amount of measurements. This bias has been theoretically evaluated and reduced in [44].

The method proposed in [45] uses Monte Carlo importance sampling techniques [27] to approximate the solution of the problem (5). As an initial point, it uses the estimate computed by a convex relaxation method. According to simulation experiments, its localization performance is on par with the convex method [46], but the computational complexity is much lower.

A very recent article [47] proposes the linearization approach that casts the original problem into an eigenvalue one, which can be solved optimally in closed form. Additionally, the authors propose an Iterative Reweighted Least Squares scheme that approximates the ML estimate for different noise distributions.

B. Convex relaxations

Another important line of work are the methods based on convex relaxations of ML estimation problems. In other words, the original problem is approximated by a convex one [48], which is usually far easier to solve. Two families of approaches dominate this field: methods based on semidefinite programming (SDP), and the ones relaxing the original task into a second-order cone optimization problem (SOCP). In the former, the non-convex quadratic problem (5) is first *lifted* such that the non-convexity appears as a rank 1 constraint, which is then substituted by a positive semidefinite one [49]. Lifting refers to a problem reformulation (by a suitable variable substitution), such that the original problem is redefined in a higher-dimensional space. The rationale behind lifting is that the new problem becomes easier to solve, despite being high dimensional (particularly, it leads to a SDP problem). On the other hand, solving the SDP optimization problems can be computationally expensive, and the SOCP framework has been proposed as a compromise between the approximation quality and computational complexity (*cf.* [50] and the references therein for technical details).

One of the first convex relaxation approaches for the RD localization is [51], based on SDP. The algorithm requires the knowledge of the microphone closest to the source, in order to ensure that all RDs (with that microphone as a reference) are positive. The article [46] discusses three convex relaxation methods. The first one, based on SOCP relaxation is computationally efficient, but restricts the solution to the convex hull [52], [48] of microphone positions. The other two SDP-based remove this restriction, but are somewhat more computationally demanding. In addition, one of these is the *robust* version - it minimizes the worst-case error due to imprecise microphone locations. The latter requires tuning of several hyperparameters, among which is the variance of the microphone positioning error. All three versions are based on the white Gaussian noise model for the RD measurements, however, whithening could be applied in order to support the correlated noise case. However, the SDP solutions are not the final output of the algorithms, but are used to initialize nonlinear iterative scheme, such as [35].

Interestingly, a recent article [53] has shown that the ideas of the direct approach [32] and the constrained least-squares approach could be mixed together. Moreover, the cost function can be cast to a convex problem, for which an interior-point method has been proposed. However, in practice, it is a compound algorithm which iteratively solves a sequence of convex problems in order to re-calculate a weighting matrix dependant on the estimated source position. The accuracy depends on the number of iterations, which, in turn, increases computational complexity. As for [32], it requires 5 microphones for the 3D localization.

IV. Least-squares estimation

Largely due to computational convenience, the least-squares (LS) estimation is often a preferred parameter estimation approach. It is noteworthy that all LS approaches optimize a somewhat "artificial" estimation objective, which can induce large errors in very low SNR conditions, when the measurement noise is not white, and/or for some adverse array geometries [38], [54], [44].

Three types of cost functions are discussed: hyperbolic, spherical and conic LS.

A. Hyperbolic Least Squares

The goal is to minimize the sum of squared distances ϵ_{h} between the true and estimated RDs:

$$\epsilon_{\mathrm{h}}(\hat{\mathbf{r}}) := \sum_{\mathsf{m}'=2}^{\mathsf{M}} \left(d_{\mathsf{m}'} - \|\mathbf{r}_{\mathsf{m}'} - \hat{\mathbf{r}}\| + \|\hat{\mathbf{r}}\|\right)^2 = \left(\mathbf{d} - \hat{\mathbf{d}}(\hat{\mathbf{r}})\right)^{\mathsf{T}} \left(\mathbf{d} - \hat{\mathbf{d}}(\hat{\mathbf{r}})\right), \quad (6)$$

which is analogous to the ML estimation problem (5) for $\mathbf{\Sigma} = \mathbf{I}$, with $\mathbf{I}$ being the identity matrix. Thus, in the case of *white* Gaussian noise, the hyperbolic LS solution coincides with the ML solution. Otherwise, solving (6) comes down to finding the point $\hat{\mathbf{r}}$ whose cumulative distance $\sum d_{\mathsf{m}'}$ to all hyperboloids, defined in (4), is minimal.

However, the hyperbolic LS problem is also non-convex, and its global solution cannot be guaranteed. Instead, local minimizers are found by iterative procedures, such as (nonlinear) gradient descent or particle filtering [55], [22]. Obviously, the quality of the output result of such algorithms depends on their initial estimates, the choice of which is usually not mathematical, but rather application-based.

B. Spherical Least Squares

By squaring the idealized RD measurement expression (4), followed by some simple algebraic manipulations, we have

$$d_{\mathsf{m}'}\|\mathbf{r}\| + \mathbf{r}_{\mathsf{m}'}^{\mathsf{T}}\mathbf{r} - \underbrace{\frac{1}{2}\left(\|\mathbf{r}_{\mathsf{m}'}\|^2 - d_{\mathsf{m}'}^2\right)}_{b_{\mathsf{m}'}} = 0.$$

The interest of this operation lies in decoupling of the position vector and its magnitude, which are to be replaced by their estimates $\hat{\mathbf{r}}$ and $\hat{D} := \|\hat{\mathbf{r}}\|$, respectively.

The goal now becomes driving the sum of left hand sides (for all microphones) to zero:

$$\epsilon_{\mathrm{sp}} = \sum_{\mathsf{m}'=2}^{\mathsf{M}} \left(d_{\mathsf{m}'}\hat{D} + \mathbf{r}_{\mathsf{m}'}^{\mathsf{T}}\hat{\mathbf{r}} - b_{\mathsf{m}'}\right)^2, \ \hat{D}^2 = \|\hat{\mathbf{r}}\|^2, \quad (7)$$

which leads to the following (compactly written) constrained optimization problem [56]:

$$\underset{\hat{\mathbf{c}}}{\text{minimize}} \ \|\mathbf{\Phi}\hat{\mathbf{c}} - \mathbf{b}\|^2 \quad (8)$$
$$\text{subject to } \hat{\mathbf{c}}^{\mathsf{T}} \begin{bmatrix} 1 & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} \end{bmatrix} \hat{\mathbf{c}} = 0 \text{ and } \hat{c}_{(1)} \geq 0,$$

where $\mathbf{\Phi} = \begin{bmatrix} d_2 & \mathbf{r}_2^{\mathsf{T}} \\ d_3 & \mathbf{r}_3^{\mathsf{T}} \\ \cdots & \cdots \\ d_{\mathsf{M}} & \mathbf{r}_{\mathsf{M}}^{\mathsf{T}} \end{bmatrix}$, $\hat{\mathbf{c}} = \begin{bmatrix} \hat{D} \\ \hat{\mathbf{r}} \end{bmatrix}$, $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \cdots \\ b_{\mathsf{M}} \end{bmatrix}$, and $\hat{c}_{(1)}$ denotes the first entry of the column vector $\hat{\mathbf{c}}$.

In the literature, the problem above is tackled as:

a) Unconstrained LS: by ignoring the constraints relating the position estimate $\hat{\mathbf{r}}$ and its magnitude $\hat{D}$, the problem (7) admits a closed-form solution $\hat{\mathbf{c}}^* = \left(\mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^{\mathsf{T}}\mathbf{b}$. As pointed in [57], [58], several well-known estimation algorithms [40], [41], [42] actually yield the unconstrained LS estimate. The minimum of $\mathsf{M} = 5$ microphones (*i.e.* four RD measurements), in three dimensions, are required in order for $\left(\mathbf{\Phi}^{\mathsf{T}}\mathbf{\Phi}\right)^{-1}$ to be an invertible matrix.

b) Constrained LS: While the unconstrained LS is simple and computationally efficient, its estimate is known to have a large variance compared to the CRB [58], hence the interest for solving the constrained problem. Unfortunately, (8) is non-convex due to quadratic constraints. To directly incorporate the constraint(s), a Lagrangian-based iterative method has been proposed in [33], albeit without any performance guarantees.

Later, in their seminal paper [56], Beck and Stoica provided a closed-form *global* solution of the problem, and demonstrated that it gives orders of magnitude more accurate solution (at an increased computational cost) than the unconstrained LS estimator. Moreover, the results in [45] indicate that it is generally more accurate than the two-stage ML solution [32].

C. Conic Least Squares

In [30], Schmidt has shown that (in two dimensions) the RDs of *three* known microphones define the major axis of a general conic (a hyperbola, an elipse or a parabola), on which the corresponding microphones lie. In addition, the source is positioned on its focus. In three dimensions, this axis becomes a plane containing the source. The fourth (non-coplanar) microphone is needed to infer the source position $\mathbf{r}$, by calculating the intersection coordinates of three such planes (hence the name *plane intersection method* in the literature [40]). Thus, the method attains the theoretical minimum for the required number of microphones for RD localization. Nevertheless, given the minimal number of measurements, multilateration often yields an ill-posed problem [59]. Thereby, in practice, more sensors are needed for obtaining a meaningful result.

To illustrate the approach, let one such triplet of microphones be described by $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$, and (D_1, D_2, D_3) – their position vectors, and the distances to the source, respectively. For a pair (i, j) of these microphones, we have the following expression for the product of the *range sum* $\Sigma_{\mathsf{i},\mathsf{j}} = D_{\mathsf{i}} + D_{\mathsf{j}}$ and the *range difference* $d_{\mathsf{i},\mathsf{j}} = D_{\mathsf{j}} - D_{\mathsf{i}}$:

$$\Sigma_{\mathsf{i},\mathsf{j}} d_{\mathsf{i},\mathsf{j}} = \|\mathbf{r}_{\mathsf{j}}\|^2 - \|\mathbf{r}_{\mathsf{i}}\|^2 - 2(\mathbf{r}_{\mathsf{j}} - \mathbf{r}_{\mathsf{i}})^{\mathsf{T}}\mathbf{r}. \quad (9)$$

Note here that the conic method uses a full set of RD observations, as opposed to the spherical least squares approach.

By rearranging the terms in (9), and having $d_{\mathsf{k},\mathsf{i}} = \Sigma_{\mathsf{i},\mathsf{j}} - \Sigma_{\mathsf{j},\mathsf{k}}$, the range sums can be eliminated. Eventually, this gives the aforementioned plane equation

$$\left(d_{2,3}\mathbf{r}_1 + d_{3,1}\mathbf{r}_2 + d_{1,2}\mathbf{r}_3\right)^{\mathsf{T}}\mathbf{r} = \frac{1}{2}\left(d_{1,2}d_{2,3}d_{3,1} + d_{2,3}\|\mathbf{r}_1\|^2 + d_{3,1}\|\mathbf{r}_2\|^2 + d_{1,2}\|\mathbf{r}_3\|^2\right). \quad (10)$$

This is a linear equation of three unknowns, thus the exact solution is obtained when three triplets (*i.e.* four non-coplanar microphones) are available. Browsing the literature, we found that exactly the same closed-form approach has been recently reinvented in the highly cited article [60].

For M microphones, one ends up with $\binom{\mathsf{M}}{3}$ such equations (in 3D) - the classical LS solution is to stack them into a matrix form, and calculate the position $\mathbf{r}$ by applying the Moore-Penrose pseudoinverse. Let A_{pqr}, B_{pqr}, C_{pqr} and F_{pqr} denote the coefficients and the right hand side of the expression (10), for the microphone triplet $\mathsf{m} \in \{\mathsf{p}, \mathsf{q}, \mathsf{r}\}$, respectively. For all such triplets, we have

$$\underbrace{\begin{bmatrix} A_{123} & B_{123} & C_{123} \\ \vdots & \vdots & \vdots \\ A_{\mathsf{pqr}} & B_{\mathsf{pqr}} & C_{\mathsf{pqr}} \\ \vdots & \vdots & \vdots \end{bmatrix}}_{\boldsymbol{\Psi}} \mathbf{r} = \underbrace{\begin{bmatrix} F_{123} \\ \vdots \\ F_{\mathsf{pqr}} \\ \vdots \end{bmatrix}}_{\boldsymbol{\psi}}, \tag{11}$$

where

$$\begin{aligned} A_{\mathsf{pqr}} &= d_{\mathsf{q,r}} r_{\mathsf{p}(1)} + d_{\mathsf{r,p}} r_{\mathsf{q}(1)} + d_{\mathsf{p,q}} r_{\mathsf{r}(1)}, \\ B_{\mathsf{pqr}} &= d_{\mathsf{q,r}} r_{\mathsf{p}(2)} + d_{\mathsf{r,p}} r_{\mathsf{q}(2)} + d_{\mathsf{p,q}} r_{\mathsf{r}(2)}, \\ C_{\mathsf{pqr}} &= d_{\mathsf{q,r}} r_{\mathsf{p}(3)} + d_{\mathsf{r,p}} r_{\mathsf{q}(3)} + d_{\mathsf{p,q}} r_{\mathsf{r}(3)} \text{ and} \\ F_{\mathsf{pqr}} &= \frac{1}{2}\left(d_{\mathsf{p,q}} d_{\mathsf{q,r}} d_{\mathsf{r,p}} + d_{\mathsf{q,r}} \|\mathbf{r}_{\mathsf{p}}\|^2 + d_{\mathsf{r,p}} \|\mathbf{r}_{\mathsf{q}}\|^2 + d_{\mathsf{p,q}} \|\mathbf{r}_{\mathsf{r}}\|^2\right), \end{aligned}$$

as in (10). However, such LS solution is strongly influenced by the triplets having large $A_{\cdot}$, $B_{\cdot}$, $C_{\cdot}$ or $F_{\cdot}$ values. Instead, as proposed in [30], the matrix $\boldsymbol{\Psi}$ needs to be preprocessed prior to computing the pseudoinverse - its rows should be scaled by $1/\sqrt{A_{\cdot}^2 + B_{\cdot}^2 + C_{\cdot}^2}$, as well as the corresponding entry of the vector $\boldsymbol{\psi}$.

Likewise, the presence of noise in the RD measurements $d_{\mathsf{i,j}}$ could seriously degrade the localization accuracy. In that case, the observation model (3) contains an additive noise term, which varies accross different measurements, rendering them *inconsistent*. This means that the intrinsic redundancy within RDs does not hold, *e.g* $d_{\mathsf{i,k}} \neq d_{\mathsf{i,j}} + d_{\mathsf{j,k}}$. In the noiseless case, the vector $\mathbf{d}$ of concatenated RD measurements, lies in the range space of a simple first-order difference matrix [61], specified by (3) and the ordering of distances D_{m}. Thus, the measurements could be preconditioned, by replacing them with the closest feasible RDs, in the LS sense. This is done by projecting the measured $\mathbf{d}$ onto the range space of a finite difference matrix, or, equivalently by the technique called "TDOA averaging" [61]. While computationally efficient, the proposed preconditioning technique assumes Gaussian distribution of TDOA (or RD) estimation errors [31], which may produce suboptimal results.

V. Speaker localization experiments

To the best of our knowledge, to date, there is no comprehensive benchmark of multilateration algorithms in the context of distributed single-channel audio sensing. The aim of this section is to contribute by providing an empirical analysis of three algorithms representative of the least squares class. In particular, we conduct a small-scale benchmark of the conic least squares [30], unconstrained [40] and constrained [56] spherical least squares (referred to as conic, usrd-ls and srd-ls, respectively). We first present the setup, data and performance measures before proceeding to the benchmark results.

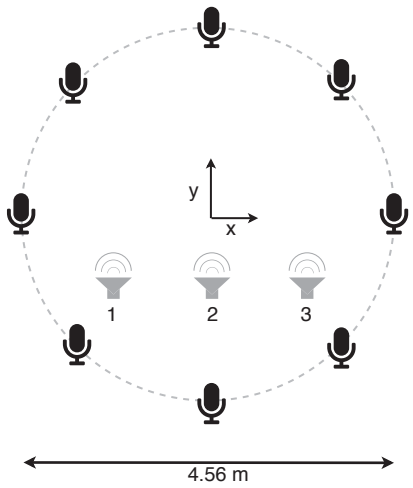


Fig. 3. Experimental setup

TABLE I. Source coordinates & orientations

	Position 1	Position 2	Position 3
x [cm]	-80	0	80
y [cm]	-80	-80	-80
z [cm]	119	119	119
Azimuth [°]	{0, ±90, 180}	{0, ±90, 180}	{0, ±90, 180}

The microphone signals are generated by convolving dry source signals with real RIRs measured in our audio lab whose $RT60 = 350\ ms$. This approach allows to spare time and use different excitation signals afterwards. Eight omnidirectionnal DPA® 4060 microphones, arranged in a circle of radius 2.28m, surround the Genelec® 1031A loudspeaker which serves as a source, as shown in 3. The microphones and the loudspeaker are approximately in the same horizontal plane (the difference in their z-coordinates is about 15cm). The RIRs are retrieved using exponential sine sweep excitation signals ranging from 20 Hz to 20 kHz [62]. During each recording, the loudspeaker is static, placed at one of the 3 different positions within the circle (Table I details the source positions and azimuths). To account for the loudspeaker directivity, at each position, the loudspeaker is oriented in 4 different directions, by rotating it around the z-axis in steps of 90°. The original sampling rate of the microphone recordings was 48kHz (because that database will have other uses) and has been subsequently down-sampled to 16kHz to match the signals described later, and all devices share the same clock.

For computing the TDOAs, we opted for the widely used GCC-PHAT method. The signals are first segmented in 50%-overlapping frames of duration 0.064s, and modified by Hanning window. The frame duration is chosen as the limit of the local (quasi-) stationarity of speech signals [63], to maximize the number of samples used for estimating correlations. Then, the TDOA estimation is performed for each frame pair, and a single output value is produced by median aggregation. For the aggregation, we either consider all frame-wise TDOAs (the "no-VAD", *i.e.* no Voice Activity Detector setting), or we choose a subset of these using a simple energy-based criterion. For the latter, we evaluate the energy of each frame of the pair,

and if neither has the energy that exceeds the half median energy of the sum of the two windowed representations, it is discarded ("VAD" variant). At the end, the obtained TDOA matrix is either directly provided to a multilateration algorithm, or is further postprocessed by the TDOA averaging method [61] (the "denoised" TDOA).

Concerning the algorithm-specific settings, we tested the conic LS with and without row-wise normalization, and spherical LS with two choices of the reference microphone. We refer to the choices as the "max" and "min", designating the microphones that produce the recordings of highest and lowest energy, respectively. This is motivated by the fact that, as the optimal choice of the reference microphone is not always straightforward, the corresponding non-redundant RD subset may contain more or less inaccuracies.

Regarding each algorithmic setting as a separate method, and the different TDOA variants as distinct features, the benchmark comprises 6 multilateration methods, each applied to 4 types of TDOA features. For each experiment, all possible subsets of 5 out of 8 microphones are selected. Note that for some loudspeaker orientations and microphone subsets the line-of-sight assumption is essentially violated (except at very low frequency bands). As the source signals, we use 6 short audio excerpts from the TIMIT database [64], comprising 3 male and 3 female speakers (meaning that each experiment has been repeated 6 times using a different excitation signal).

The multilateration performance is quantified as the position error in meters, between the ground truth and the estimated position, *i.e.* $\|\mathbf{r} - \hat{\mathbf{r}}\|$. In addition, we track the TDOA/RD estimation errors, in order to evaluate the robustness of a localization method to various levels of RD "noise". Therefore, we also store the average absolute errors of all RDs for the considered microphone subset (the ground truth RDs are easily determined from the microphone-source geometry).

The overall results in the form of a box plot are shown on 4, which we interpret below.

First, the RD denoising (by means of TDOA averaging) does not contribute to an increased localization accuracy. Although the variance of positioning errors is reduced for the denoised RDs, the median error is generally larger than for the non-processed RD features. The denoising operation actually redistributes the error across all RDs, thereby corrupting all the "clean" RD entries as well. Contrary to the intuition, the conic LS method seem to be affected the most, probably due to the fact that it exploits all available RD observations, thereby accumulating most of the noise.

Second, our rudimentary VAD solution seems to be beneficial, despite the fact that the TIMIT source signals contain mostly uninterrupted speech. We attribute this to the selection of highly energetic frames, which are affected by noise and reverberation to a lesser degree.

Finally, according to the median multilateration performance plots in the most favorable setting (GCC-PHAT with active VAD and without denoising), the normalized conic LS seems to be the best solution. However, 2D histograms, presented in 5, which depict the position error with regards to average RD errors, indicate that the constrained spherical LS method offers the most robust performance when the reference microphone is well-chosen (e.g. by selecting the nearest microphone from the barycenter). Related to this choice, a slight drop in performance of both spherical LS methods is observed with the suboptimal ("min") choice of the reference microphone.

VI. Conclusion

Multilateration has a long history, and the methods belonging to this family of localization algorithms are theoretically well-founded. Moreover, they are generic, in the sense that they are essentially agnostic to the signal type, as long as the (pseudo) RDs are obtainable. While this article primarily considers sound source localization, multilateration could be straightforwardly applied to, *e.g.* mapping problems in sensor networks, geolocalization by positioning systems and/or base stations, or to target localization in distributed radar signal processing.

The multilateration methods of the ML class are closed to optimal in theory, however they resort to various approximations in order to combat the intrinsic hardness of the localization problem. The LS approaches instead solve easier, but artificial optimization problems. On the other hand, some of them are computationally very efficient, and seemingly work very well in practice. The small-scale benchmark of the three widely used LS methods suggests that constrained spherical LS method offers competitive performance, in terms of accuracy and robustness to TDOA estimation errors, however, at an increased computational cost compared to the unconstrained and conic LS. The choice of the method should be dictated by the use case and the a priori information that may be available: what is the type and the level of measurement noise, how important is the computational complexity, how many microphones comprise the array, what are their specifications etc. There seems to be no clear winner when different criteria are taken into account at the same time, which calls for a dedicated, more comprehensive test study in the future.

References

[1] L. Turchet, G. Fazekas, M. Lagrange, H. S. Ghadikolaei, and C. Fischione, "The Internet of Audio Things: state-of-the-art, vision, and challenges," *IEEE Internet of Things Journal*, 2020.

[2] E. Vincent, T. Virtanen, and S. Gannot, *Audio source separation and speech enhancement*. John Wiley & Sons, 2018.

[3] T. Joubaud and G. Pallone, "Electroacoustic method for the calibration of a heterogeneous distributed audio system," in *28th European Signal Processing Conference (EUSIPCO)*, Amsterdam, NL, 2020/21.

[4] S. Kitić, N. Bertin, and R. Gribonval, "Hearing behind walls: localizing sources in the room next door with cosparsity," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 3087–3091.

[5] J. Dmochowski, J. Benesty, and S. Affès, "On spatial aliasing in microphone arrays," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1383–1395, 2008.

[6] G. Valenzise, G. Prandi, M. Tagliasacchi, and A. Sarti, "Resource constrained efficient acoustic source localization and tracking using a distributed network of microphones," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*. IEEE, 2008, pp. 2581–2584.

[7] K. Yao, R. E. Hudson, C. W. Reed, D. Chen, and F. Lorenzelli, "Blind beamforming on a randomly distributed sensor array system," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1555–1567, 1998.

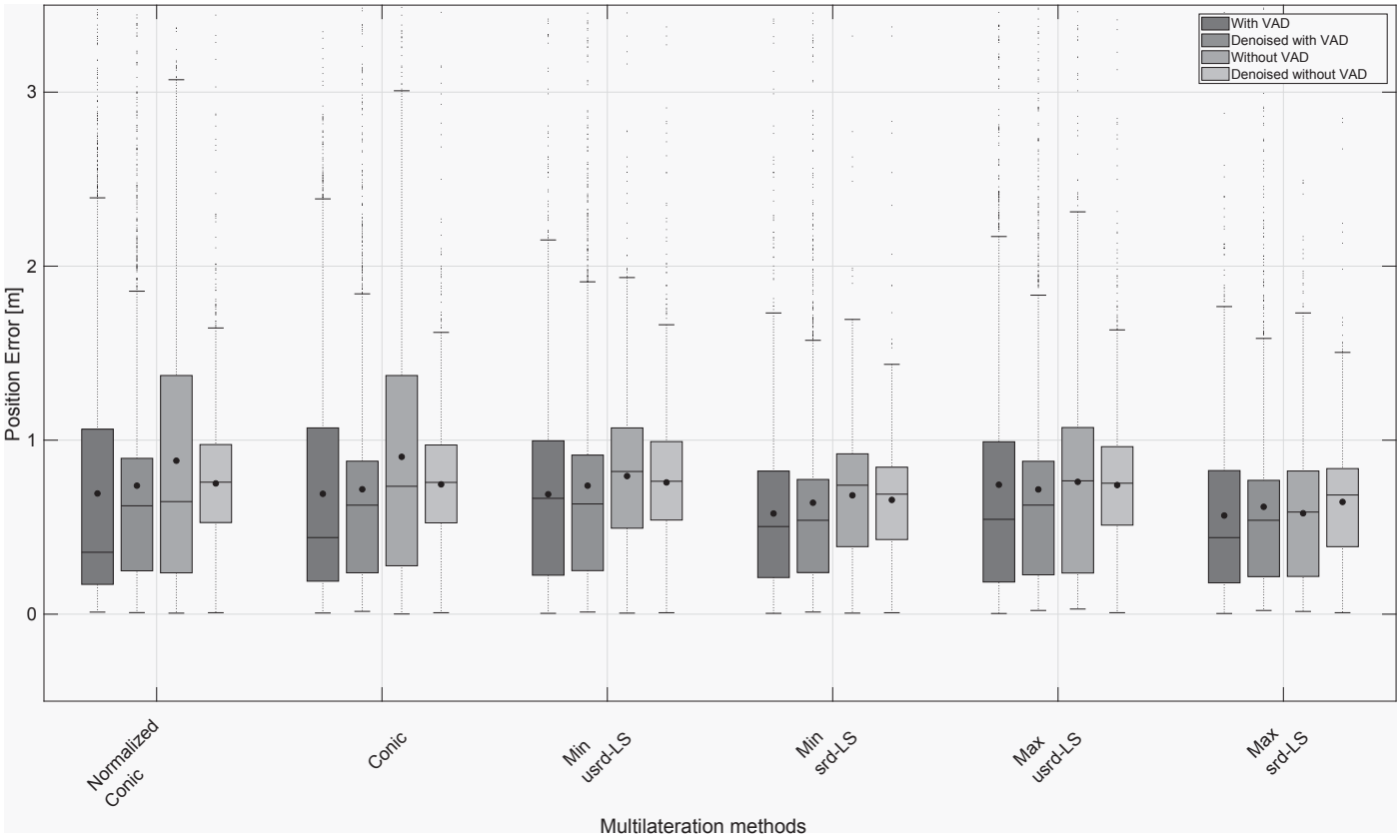
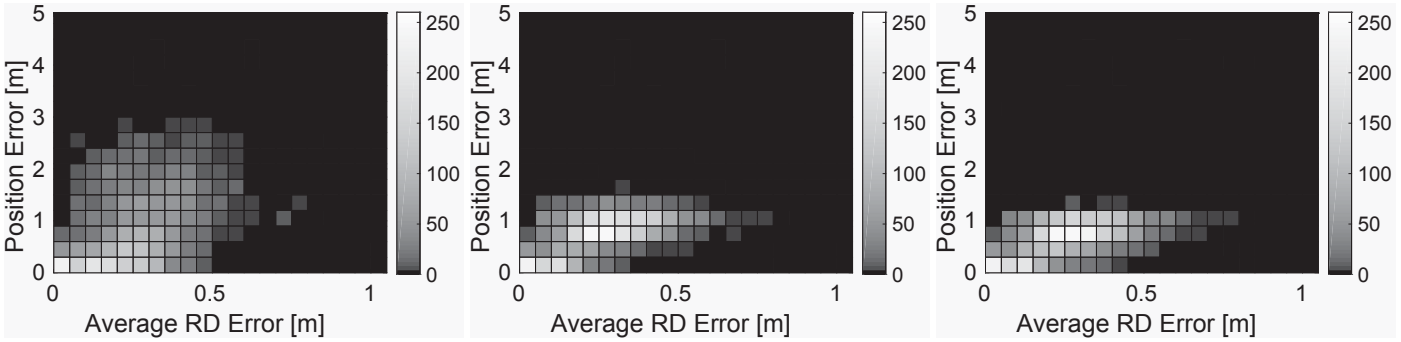


Fig. 4. Source localization performance



(a) Normalized Conic (b) usrd-ls (c) srd-ls

Fig. 5. Histogram of the average RD errors *vs* position errors

[8] F. Hummes, J. Qi, and T. Fingscheidt, "Robust acoustic speaker localization with distributed microphones," in *2011 19th European Signal Processing Conference*. IEEE, 2011, pp. 240–244.

[9] 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 Processing Letters*, vol. 18, no. 1, pp. 71–74, 2010.

[10] J. M. Vera-Diaz, D. Pizarro, and J. Macias-Guarasa, "Towards end-to-end acoustic localization using deep learning: From audio signals to source position coordinates," *Sensors*, vol. 18, no. 10, p. 3418, 2018.

[11] S. Chakrabarty and E. A. Habets, "Multi-speaker DOA estimation using deep convolutional networks trained with noise signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 8–21, 2019.

[12] L. Perotin, R. Serizel, E. Vincent, and A. Guérin, "CRNN-based multiple DoA estimation using acoustic intensity features for Ambisonics recordings," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 22–33, 2019.

[13] Y. Wang and K. Ho, "TDOA source localization in the presence of synchronization clock bias and sensor position errors," *IEEE Transactions on Signal Processing*, vol. 61, no. 18, pp. 4532–4544, 2013.

[14] J. Fresno, G. Robles, J. Martínez-Tarifa, and B. Stewart, "Survey on the performance of source localization algorithms," *Sensors*, vol. 17, no. 11, p. 2666, 2017.

[15] M. Cobos, F. Antonacci, A. Alexandridis, A. Mouchtaris, and B. Lee, "A survey of sound source localization methods in wireless acoustic sensor networks," *Wireless Communications and Mobile Computing*, vol. 2017, 2017.

[16] L. Cheng, C. Wu, Y. Zhang, H. Wu, M. Li, and C. Maple, "A survey of localization in wireless sensor network," *International Journal of Distributed Sensor Networks*, vol. 8, no. 12, p. 962523, 2012.

[17] M. Compagnoni, R. Notari, F. Antonacci, and A. Sarti, "A comprehen-

sive analysis of the geometry of TDOA maps in localization problems," *Inverse Problems*, vol. 30, no. 3, p. 035004, 2014.

[18] J. Chen, J. Benesty, and Y. A. Huang, "Time delay estimation in room acoustic environments: an overview," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, p. 026503, 2006.

[19] C. Blandin, A. Ozerov, and E. Vincent, "Multi-source TDOA estimation in reverberant audio using angular spectra and clustering," *Signal Processing*, vol. 92, no. 8, pp. 1950–1960, 2012.

[20] A. Lombard, Y. Zheng, H. Buchner, and W. Kellermann, "TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1490–1503, 2010.

[21] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE transactions on acoustics, speech, and signal processing*, vol. 24, no. 4, pp. 320–327, 1976.

[22] F. Gustafsson and F. Gunnarsson, "Positioning using time-difference of arrival measurements," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03)*, vol. 6. IEEE, 2003, pp. VI–553.

[23] R. Horaud, "Sound-source localization," MOOC: Binaural Hearing for Robots, 2011.

[24] J. C. Rodriguez Silva, "Theoretical Study of an RF Positioning System Using Phase Synchronized Anchor Nodes," Master's thesis, Universitat Politècnica de Catalunya, 2011.

[25] R. Kaune, "Accuracy studies for TDOA and TOA localization," in *15th International Conference on Information Fusion*. IEEE, 2012, pp. 408–415.

[26] K. Ho, X. Lu, and L.-o. Kovavisaruch, "Source localization using tdoa and fdoa measurements in the presence of receiver location errors: Analysis and solution," *IEEE Transactions on Signal Processing*, vol. 55, no. 2, pp. 684–696, 2007.

[27] S. Theodoridis, *Machine learning: a Bayesian and optimization perspective*. Academic Press, 2015.

[28] B. Yang and J. Scheuing, "A theoretical analysis of 2D sensor arrays for TDOA based localization," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP 2006)*, vol. 4. IEEE, 2006, pp. IV–IV.

[29] W. Hahn and S. Tretter, "Optimum processing for delay-vector estimation in passive signal arrays," *IEEE Transactions on Information Theory*, vol. 19, no. 5, pp. 608–614, 1973.

[30] R. O. Schmidt, "A new approach to geometry of range difference location," *IEEE Transactions on Aerospace and Electronic Systems*, no. 6, pp. 821–835, 1972.

[31] J. Velasco, D. Pizarro, J. Macias-Guarasa, and A. Asaei, "TDOA matrices: Algebraic properties and their application to robust denoising with missing data," *IEEE Transactions on signal processing*, vol. 64, no. 20, pp. 5242–5254, 2016.

[32] Y.-T. Chan and K. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Transactions on Signal Processing*, vol. 42, no. 8, pp. 1905–1915, 1994.

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

[34] Y. A. Huang, J. Benesty, and J. Chen, "Time delay estimation and source localization," in *Springer Handbook of Speech Processing*. Springer, 2008, pp. 1043–1063.

[35] W. H. Foy, "Position-location solutions by Taylor-series estimation," *IEEE Transactions on Aerospace and Electronic Systems*, no. 2, pp. 187–194, 1976.

[36] T. Ajdler, I. Kozintsev, R. Lienhart, and M. Vetterli, "Acoustic source localization in distributed sensor networks," in *Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004.*, vol. 2. IEEE, 2004, pp. 1328–1332.

[37] I. A. Mantilla-Gaviria, M. Leonardi, G. Galati, and J. V. Balbastre-Tejedor, "Localization algorithms for multilateration (MLAT) systems in airport surface surveillance," *Signal, Image and Video Processing*, vol. 9, no. 7, pp. 1549–1558, 2015.

[38] A. N. Bishop, B. Fidan, B. D. Anderson, K. Dogancay, and P. N. Pathirana, "Optimal range-difference-based localization considering geometrical constraints," *IEEE Journal of Oceanic Engineering*, vol. 33, no. 3, pp. 289–301, 2008.

[39] A. N. Bishop, B. Fidan, B. D. Anderson, K. Doğançay, and P. N. Pathirana, "Optimality analysis of sensor-target localization geometries," *Automatica*, vol. 46, no. 3, pp. 479–492, 2010.

[40] J. Smith and J. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 12, pp. 1661–1669, 1987.

[41] D. Li and Y. H. Hu, "Least square solutions of energy based acoustic source localization problems," in *Workshops on Mobile and Wireless Networking/High Performance Scientific, Engineering Computing/Network Design and Architecture/Optical Networks Control and Management/Ad Hoc and Sensor Networks/Compil*. IEEE, 2004, pp. 443–446.

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

[43] Y.-T. Chan, H. Y. C. Hang, and P.-c. Ching, "Exact and approximate maximum likelihood localization algorithms," *IEEE Transactions on Vehicular Technology*, vol. 55, no. 1, pp. 10–16, 2006.

[44] K. Ho, "Bias reduction for an explicit solution of source localization using TDOA," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2101–2114, 2012.

[45] G. Wang and H. Chen, "An importance sampling method for TDOA-based source localization," *IEEE Transactions on Wireless Communications*, vol. 10, no. 5, pp. 1560–1568, 2011.

[46] K. Yang, G. Wang, and Z.-Q. Luo, "Efficient convex relaxation methods for robust target localization by a sensor network using time differences of arrivals," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2775–2784, 2009.

[47] M. Larsson, V. Larsson, K. Astrom, and M. Oskarsson, "Optimal trilateration is an eigenvalue problem," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019)*. IEEE, 2019, pp. 5586–5590.

[48] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[49] W.-K. K. Ma, "Semidefinite relaxation of quadratic optimization problems and applications," *IEEE Signal Processing Magazine*, vol. 1053, no. 5888/10, 2010.

[50] S. Kim and M. Kojima, "Second order cone programming relaxation of nonconvex quadratic optimization problems," *Optimization methods and software*, vol. 15, no. 3-4, pp. 201–224, 2001.

[51] K. W. K. Lui, F. K. W. Chan, and H.-C. So, "Semidefinite programming approach for range-difference based source localization," *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1630–1633, 2008.

[52] P. Biswas and Y. Ye, "Semidefinite programming for ad hoc wireless sensor network localization," in *Proceedings of the 3rd international symposium on Information processing in sensor networks*. ACM, 2004, pp. 46–54.

[53] X. Qu and L. Xie, "An efficient convex constrained weighted least squares source localization algorithm based on TDOA measurements," *Signal Processing*, vol. 119, pp. 142–152, 2016.

[54] N. Sirola, "Closed-form algorithms in mobile positioning: Myths and misconceptions," in *2010 7th Workshop on Positioning, Navigation and Communication*. IEEE, 2010, pp. 38–44.

[55] D. J. Torrieri, "Statistical theory of passive location systems," *IEEE Transactions on Aerospace and Electronic Systems*, no. 2, pp. 183–198, 1984.

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

[57] P. Stoica and J. Li, "Lecture notes-source localization from range-difference measurements," *IEEE Signal Processing Magazine*, vol. 23, no. 6, pp. 63–66, 2006.

[58] H.-W. Wei and S.-F. Ye, "Comments on "A Linear Closed-Form Algorithm for Source Localization From Time-Differences of Arrival"," *IEEE Signal Processing Letters*, vol. 15, pp. 895–895, 2008.

[59] S. C. Herath and P. N. Pathirana, "Robust localization with minimum number of TDOA measurements," *IEEE Signal Processing Letters*, vol. 20, no. 10, pp. 949–951, 2013.

[60] R. Bucher and D. Misra, "A synthesizable VHDL model of the exact solution for three-dimensional hyperbolic positioning system," *Vlsi Design*, vol. 15, no. 2, pp. 507–520, 2002.

[61] R. Schmidt, "Least squares range difference location," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 32, no. 1, pp. 234–242, 1996.

[62] A. Farina, "Simultaneous measurement of impulse response and distorsion with a swept-sine technique," in *AES 108th Convention, Paris, France*, 2000, p. 23 pages.

[63] J. Benesty, M. M. Sondhi, and Y. Huang, *Springer handbook of speech processing*. Springer, 2007.

[64] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, "Darpa timit acoustic-phonetic continous speech corpus cd-rom. nist speech disc 1-1.1," *NASA STI/Recon technical report n*, vol. 93, 1993.