HAL
open science

Geometrically-constrained time delay estimation-based sound source localisation (gTDESSL)

Xavier Alameda-Pineda, Radu Horaud

To cite this version:

Xavier Alameda-Pineda, Radu Horaud. Geometrically-constrained time delay estimation-based sound source localisation (gTDESSL). [Research Report] RR-7988, INRIA. 2012, pp.28. hal-00704986v2

HAL Id: hal-00704986

https://inria.hal.science/hal-00704986v2

Submitted on 6 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Geometrically-constrained time delay estimation-based sound source localisation (gTDESSL)

Xavier Alameda-Pineda and Radu Horaud

N° 7988

Juin 2012

Domaine 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7988--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche GRENOBLE - RHÔNE-ALPES

Geometrically-constrained time delay estimation-based sound source localisation (gTDESSL)

Xavier Alameda-Pineda* and Radu Horaud

Domaine : Perception, cognition, interaction

Équipe-Projet PERCEPTION

Rapport de recherche n° 7988 — Juin 2012 — 28 pages

Abstract: In this paper we present a geometrically-constrained time delay estimation method for sound source localization (gTDE). An algebraic analysis reveals that the method can deal with an arbitrary number of non-coplanar microphones. We derive a constrained non-linear optimization problem that can be solved using convex programming. Unlike existing techniques, which consider pairwise TDE's, the proposed method optimally estimates a set of time delays that are consistent with the source's location. Extensive simulated experiments validate the method in the presence of noise and of reverberations.

Key-words: No keywords

This work was supported by the European project HUMAVIPS, under EU grant FP7-ICT-2009-247525.

* Corresponding author: xavier.alameda-pineda@inria.fr

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe, 38334 Montbonnot Saint Ismier
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Localisation de sources sonores en utilisant l'estimation de delais temporels contraints gometriquement.

Résumé : Dans cet article nous prsentons une mthode de temps gomtriquement contraints retard d'estimation pour la source sonore la localisation (gTDE). Une analyse algbrique rvle que la mthode peut traiter avec un nombre arbitraire de non-coplanaires microphones. Nous tirons un problme d'optimisation sous contrainte non-linaire qui peut łtre rsolu en utilisant la programmation convexe. contrairement techniques existantes, qui considrent paires TDE, la mthode propose estime de manire optimale un ensemble de retards qui sont compatibles avec l'emplacement de la source. De vastes expriences simules de valider la mthode en prsence de bruit et de rverbrations.

Mots-clés : Pas de motclef

Contents

1 Introduction **4**

2 Related work **4**

2.1 Closest work 6

3 Signal model **7**

4 Multichannel Time Delay Estimation **7**

5 Geometric model **8**

5.1 Direct model 9

5.2 Inverse model 9

6 Geometrically constrained criterion **11**

7 Implementation, Experiments & results **11**

8 Conclusions & future work **14**

A Criteria equivalence **14**

B Deriving formulaes for $\nabla\tilde{J}$ and $\mathbf{H}\tilde{J}$ **15**

C Estimating of the cross-correlation function **18**

D Deriving the $\nabla\Delta$ and $\mathbf{H}\Delta$ **21**

E Analytic cross-correlation function **21**

E.1 Exponentially decaied sinusoidal 22

E.2 The full spectrum sinusoidal 23

References **24**

1 Introduction

For the last two decades, source localization from time delay estimates (TDE's) has proven to be an extremely useful methodology with a variety of applications in such diverse fields as aeronautics, telecommunications and robotics. Also referred to as *multilateration*, this problem is highly related to the one of estimating time delays. We are particularly interested in the development of a general-purpose TDE-based method for sound-source localization in indoor environments, e.g, human-robot interaction, ad-hoc teleconferencing microphone arrays, etc. This type of consumer-oriented applications are extremely challenging for several reasons: (i) there may be several sound sources and their number varies over time, (ii) regular rooms are echoic, thus leading to reverberations, and (iii) the microphones are often embedded in devices (robot heads, smart phones, etc.) generating high-level noise.

Structure

2 Related work

The TDE problem was very well investigated: A recent review can be found in [10]. The vast majority of existing approaches deals with a microphone pair but it is not straightforward to extend most of these methods to several microphones. Methods addressing *multichannel* TDE can be roughly divided into two categories: methods estimating the acoustic impulse responses and methods exploiting the redundancy among several microphones. [13] is illustrative of the first category: a method based on generalized eigenvalue decomposition is proposed. The second category is represented by [9] where a multichannel criterion based on cross-correlation is proposed to estimate time delays using a *linear* microphone array. In both cases, experiments are performed on speech data in a simulated indoor environment.

As already mentioned, an alternative to TDE is multilateration, which makes assumptions about the time delay estimates. This provides a framework for casting the problem into maximum-likelihood estimation or into mean-squared error minimization (see [29] for a review). Two recent methods deserve to be mentioned. In [5] the authors use the acoustic maps together with the GCC-PHAT technique to localize sound sources from TDE's. The model in [30] includes the reverberations in order to enhance the localization performance while using a uniform circular array of microphones.

There are, of course, tones of work that has been done before, related to the topic. I will briefly describe several works. In the next section, some selected papers will be described in more detail.

- The method described in [7] performs 2D localization from a non-linear array of three microphones. The use an unbiased estimator for TDOA. From this paper, I derived the 3D localization from a non-coplanar array of four microphones. This is not a contribution, since the analogy with the 2D is straightforward and too obvious.

- In [34, 15] they derive a least-squares method for more than four microphones assuming independent noise in the TDOA measurements.
- In [3, 4] they approximate the hyperboloids by cones and perform linear intersection in the least-squares sense. Tested on simulated and real data respectively. TDOAs are estimated independently.
- A linear correction for least-squares approach is presented in [22]. Just on simulated data.
- An outlier detection algorithm from TDOA noisi measurements by parity check is presented in [16].
- Numerical relaxation methods for the hard-least-squares problem are presented in [1] to provide an approximate solution. Just on simulations.
- A method is presented for non-line-of-sight TDOA localization in [8].
- A method is presented based on TOA (not TDOA) in [12].
- Two methods are presented in [13] to extend the adaptive eigenvalue decomposition (see [10] for a nice review on methods): stochastic gradient and data prewhitening. Simulated data.
- In [14] the problem of localization from TDOA with more than four microphones is casted into a first order Taylor series approximation, which geometrically corresponds to a hyperbolic asymptote localiation algorithm. Simulated data.
- A really nice survey on time difference estimation (TDE = TDOA) methods is presented in [10]. The comparisons are run extensively. Two microphone, N-microphone and adaptive methods are presented.
- A method for multi-microphone TDOA estimation is presented in [9]. Simulated data. Linear array.
- Here a multichannel cross-correlation coefficient os presented (applied to TDE with a linear array). Varechoich chamber.
- A method for ML estimation in the near-field and cross-bearing estimation from separated arrays is presented in [11]. Single/Multiple source. Real experiments in a semi-anechouc chamber. 2D localization.
- In [19, 18] a study in physics on how waves are refracted.
- Localization from ILD in [20].
- TDOA & FDOA based localization and velocity estimation of sound sources in [21].
- Humanoid binaural sound trakcing using kalman filtering and HRTF's (simulated and real data) in [23, 24].

- A very interesting paper on kalman filters for speaker tracking using real data in [25].
- There is a theoretical study of the close-dorm solution to a really limited (almost ridiculous) extent in [28].
- A method for sound source localization using ILD and ITD with less than four microphones is presented in [31].
- A nice survey of mathematical methods for indoor localization is presented in [32].
- ML framework for localization in [33]. Distributed sensors. TDOA errors are independent and Gaussian.
- CRLB optimal mapping from all TDOAs to N-1 TDOAs. TDOA errors are independent and Gaussian. Simulated data. ([35])
- [36] proposes a ML method to classify estimated TDOAs based on learning conditional pdfs. Tested with real data from an anechoic chamber.
- In [37] the authors propose a new ML estimator (I did not get what is new on it). The TDOA measurements are independent. Simulated data.
- Optimal microphone arrays design to minimize the CRB (interesting). [39].
- Theoretical study of 2D sensor arrays in [40].
- Relaxation methods for localization using TDOAs and simulated data in [41].
- A ML framework for localization using directional microphones is presented in [42, 43] and tested with real data.
- Modification of the PHAT algorithm to discard frequencies in which the signal is not present. Localization and tracking with real data in [38].

2.1 Closest work

There are some papers among the previous list which are more related to what I would like to do:

1. The method presented in [9] derives a multi-microphone TDOA estimator. That is, it estimates all the TDOA values at ones, instead of each TDOA separately. It has been just tested on simulated data!!!!
2. The method in [11] presents a method for sigle/multiple source TDOA estimator. This is interesting, because it is not just the single source case.
3. The method in [38] modifies the PHAT algorithm to discard frequencies in which the signal is not present.

3 Signal model

In this section, the signal acquisition model is described. The most important elements of this model are: the sound source position ($\boldsymbol{S}$), the number of microphones (M) and their positions ($\boldsymbol{M}_1, \ldots, \boldsymbol{M}_M$) and the emited signal ($x(t)$).

From this basic elements, we can write the expression for the received signal at the m-th microphone:

$$x_m(t) = x(t - t_m) + n_m(t), \tag{1}$$

where n_m is the microphone's noise and t_m is the time-of-arrival. The physical model for t_m assumes constant sound propagation speed, denoted by ν. Hence we write $t_m = \|\boldsymbol{S} - \boldsymbol{M}_m\|/\nu$. Using this model, the expression for the time dealy between the m-th and the n-th microphones arises naturally. Denoted by $t_{m,n}$, this TDE is generated by the following geometric model:

$$t_{m,n} = t_n - t_m = \frac{\|\boldsymbol{S} - \boldsymbol{M}_n\| - \|\boldsymbol{S} - \boldsymbol{M}_m\|}{\nu}. \tag{2}$$

Notice that, for a fixed value of $t_{m,n}$ and given the microphones' positions, the sound source generating $t_{m,n}$ lies in a two-sheet hyperboloid with focci $\boldsymbol{M}_n$ and $\boldsymbol{M}_m$. It is very important to notice also, that the set of all $M(M-1)/2$ time delay values corresponding to the M microphones are not independent, since $t_{m,n} = t_{m,k} + t_{k,n}$.

In the following section we develop a method for time delay estimation from the signal model described in (1).

4 Multichannel Time Delay Estimation

In this section we present the framework used to estimate the time delay between microphone signals. As in [9], the signal acquired by one of the microphones is set as the reference signal. The other signals are properly scaled and delayed in order to predict the reference signal. The best prediction, that is the best scaling and delaying values, correpond to the time delay estimates we are looking for. More precisely, given the signals acquired by the M microphones, $x_1(t), \ldots, x_M(t)$, we can set x_1 as the reference signal and write the following linear prediction error:

$$e_{\boldsymbol{c},\boldsymbol{t}}(t) = x_1(t) - \sum_{m=2}^{M} c_{1,m}\, x_m(t + t_{1,m}), \tag{3}$$

where $\boldsymbol{c} = (c_{1,2}, \ldots, c_{1,M})^{\mathsf{T}}$ are the prediction coefficients and $\boldsymbol{t} = (t_{1,2}, \ldots, t_{1,M})^{\mathsf{T}}$ are the prediction time delays. Notice that $t_{1,m} = t_m - t_1$ where t_m is the time arrival to the m-th microphone. Notice also that the signals $x_m(t + t_{1,m})$ and $x_n(t + t_{1,n})$ are on phase. The best prediction values correspond to those minimizing the expected energy of the prediction error, providing the following optimisation problem:

$$(\boldsymbol{c}^*, \boldsymbol{t}^*) = \arg\min_{\boldsymbol{c},\boldsymbol{t}} J(\boldsymbol{c}, \boldsymbol{t}) = \arg\min_{\boldsymbol{c},\boldsymbol{t}} \mathrm{E}\left\{ e^2_{\boldsymbol{c},\boldsymbol{t}}(t) \right\}. \tag{4}$$

If we develop the expression above and derive with respect to $\boldsymbol{c}$, we can show that $\boldsymbol{c}^*(\boldsymbol{t}) = \mathbf{R}^{-1}(\boldsymbol{t})\boldsymbol{r}(\boldsymbol{t})$, with:

$$\mathbf{R}(\boldsymbol{t}) = \begin{pmatrix} R_{2,2}(0) & R_{2,3}(t_{1,3}-t_{1,2}) & \cdots \\ R_{2,3}(t_{1,3}-t_{1,2}) & R_{3,3}(0) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \boldsymbol{r}(\boldsymbol{t}) = \begin{pmatrix} R_{1,2}(-t_{1,2}) \\ R_{1,3}(-t_{1,3}) \\ \vdots \end{pmatrix}.$$

In the expressions above $R_{i,j}(\tau) = \mathrm{E}\left\{x_i(t)x_j(t-\tau)\right\}$. Notice that $R_{i,j}(\tau)$ has its maxima at the true value of $t_{i,j}$. In some sense we are looking for the combination of $t_{1,k}$ that maximize the criteria, that turns out to maximize the cross-correlation functions. The optimization problem is:

$$\boldsymbol{t}^* = \arg\min_{\boldsymbol{t}} J(\boldsymbol{c}^*(\boldsymbol{t}), \boldsymbol{t}) = \arg\min_{\boldsymbol{t}} \left\{R_{1,1}(0) - \boldsymbol{r}^\mathsf{T}\mathbf{R}^{-1}\boldsymbol{r}\right\}, \tag{5}$$

where we suppressed the explicit dependency of $\boldsymbol{r}$ and $\mathbf{R}$ on $\boldsymbol{t}$. Also, it can be shown (see Appendix A) that this problem is equivalent to the following one:

$$\boldsymbol{t}^* = \arg\min_{\boldsymbol{t}} \tilde{J}(\boldsymbol{t}) = \arg\min_{\boldsymbol{t}} \det\left(\tilde{\mathbf{R}}\right), \tag{6}$$

where $\tilde{\mathbf{R}}$ is the matrix of normalized cross-correlation functions evaluated at $\boldsymbol{t}$:

$$\tilde{\mathbf{R}} = \begin{pmatrix} 1 & \rho_{1,2}(-t_{1,2}) & \rho_{1,3}(-t_{1,3}) & \cdots \\ \rho_{1,2}(-t_{1,2}) & 1 & \rho_{2,3}(t_{1,3}-t_{1,2}) & \cdots \\ \rho_{1,3}(-t_{1,3}) & \rho_{2,3}(t_{1,3}-t_{1,2}) & 1 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \tag{7}$$

As in [9] the time delay estimation problem is cast into a non-linear optimisation problem. However, since no assumption has been done about the geometry of the microphone array, the problem is multidimenional ($M-1$ variables) instead of unidimensional.

5 Geometric model

Depending on the geometry of the microphone array, some values of $\boldsymbol{t}$ do not correspond to a position in the sound source space. This was fully exploited by Chen in the particular case of linear arrays by constraining the signal model. Since this is not possible in general, we need a change of paradigm when dealing with arbitraty geometry arrays. Hence, we propose to modify the estimation procedure by adding a non-linear constraint. This constraint describes the geometric feasibility of the values in $\boldsymbol{t}$. That is why the proposed algorithm is called geometrically-constrained time delay estimation (gTDE).

In this section we present the geometric model for TDE, given the sound source's position. We will call that model the "Direct". After we derive a method to inverse this direct model, i.e., to obtain the sound source's position from the TDEs. We call that the "Inverse" model.

5.1 Direct model

Given the microphones' position, $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_M$, and the sound source's position, $\boldsymbol{S}$, the geometric generative model for the TDEs is:

$$t_{ij}(\boldsymbol{S}) = \frac{\|\boldsymbol{S} - \boldsymbol{M}_j\| - \|\boldsymbol{S} - \boldsymbol{M}_i\|}{\nu}, \tag{8}$$

where ν is the sound speed. We can also think about the following application:

$$\begin{array}{rccl} \boldsymbol{F}: & \mathbb{R}^3 & \rightarrow & \mathbb{R}^{M(M-1)/2} \\ & \boldsymbol{S} & \mapsto & \boldsymbol{F}(\boldsymbol{S}) := (t_{ij}(\boldsymbol{S}))_{i<j}\,. \end{array} \tag{9}$$

Notice that the $M(M-1)/2$ TDEs are not independent. Actually the following relation holds

$$t_{ij} = t_{kj} - t_{ki}. \tag{10}$$

From this relation we observe that the maximal subsets of independent TDEs are of size $M-1$. In other words, all subsets of size M have at least one TDE which depends on the other $M-1$. For example, we can think on the subset of all TDEs related to one microphone: $\{t_{k1}, \ldots, t_{kk-1}, t_{kk+1}, \ldots, t_{kM}\}$.

5.2 Inverse model

The model presented before is invertible under some hypothesis. This means that we can recover the 3D position of the sound source ($\boldsymbol{S}$) from a set of TDEs. In the next we will derive this inverse geometric model. As explained in the previous section we can use just $M-1$ TDEs since the other are dependent on them. We will use the TDEs related tot he first microphone: $\{t_{12}, \ldots, t_{1M}\}$. To be precise we will use a scaled version of the TDE for the geometric derivation: the DDE (Distance Difference of Arrival), noted by $\{d_{12}, \ldots, d_{1M}\}$, where $d_{ij} = \nu\, t_{ij}$.

Let us denote by $d_m(\boldsymbol{S})$ the distance from the sound source $\boldsymbol{S}$ to the m-th microphone, which can be written as:

$$d_m^2(\boldsymbol{S}) = \langle \boldsymbol{S} - \boldsymbol{M}_m, \boldsymbol{S} - \boldsymbol{M}_m \rangle = \|\boldsymbol{S}\|^2 - 2\langle \boldsymbol{S}, \boldsymbol{M}_m \rangle + \|\boldsymbol{M}_m\|^2\,. \tag{11}$$

Using the fact that $d_m(\boldsymbol{S}) = d_{1m}(\boldsymbol{S}) + d_1(\boldsymbol{S})$, we can write:

$$d_{1m}^2(\boldsymbol{S}) + 2d_{1m}(\boldsymbol{S})d_1(\boldsymbol{S}) + d_1^2(\boldsymbol{S}) = \|\boldsymbol{S}\|^2 - 2\langle \boldsymbol{S}, \boldsymbol{M}_m \rangle + \|\boldsymbol{M}_m\|^2\,, \tag{12}$$

and subtract the following equation:

$$d_1^2(\boldsymbol{S}) = \|\boldsymbol{S}\|^2 - 2\langle \boldsymbol{S}, \boldsymbol{M}_1 \rangle + \|\boldsymbol{M}_1\|^2\,. \tag{13}$$

This leads to:

$$d_{1m}^2 + 2d_{1m}d_1 = \|\boldsymbol{M}_m\|^2 - \|\boldsymbol{M}_1\|^2 - 2\langle \boldsymbol{S}, \boldsymbol{M}_m - \boldsymbol{M}_1 \rangle\,, \tag{14}$$

where we suppressed the explicit dependency of d_{m1} and d_1 on $\boldsymbol{S}$. Since this equation is non-trivial for $m \geq 1$, we can write the following system of equations:

$$\left\{\begin{array}{rcl} d_{12}^2 + 2d_{12}d_1 & = & \|\boldsymbol{M}_2\|^2 - \|\boldsymbol{M}_1\|^2 - 2\left\langle \boldsymbol{S}, \boldsymbol{M}_2 - \boldsymbol{M}_1 \right\rangle \\ & \vdots & \\ d_{1M}^2 + 2d_{1M}d_1 & = & \|\boldsymbol{M}_M\|^2 - \|\boldsymbol{M}_1\|^2 - 2\left\langle \boldsymbol{S}, \boldsymbol{M}_M - \boldsymbol{M}_1 \right\rangle \end{array}\right. \tag{15}$$

Notice that the system of equations is linear in $\boldsymbol{S}$. Indeed, we can rewrite it in the following form:

$$\mathbf{M} \cdot \boldsymbol{S} = \boldsymbol{K} - \boldsymbol{D}^2 - 2\boldsymbol{D}\, d_1 \tag{16}$$

where

$$\mathbf{M} = 2\begin{pmatrix} (\boldsymbol{M}_2 - \boldsymbol{M}_1)^\mathsf{T} \\ \vdots \\ (\boldsymbol{M}_M - \boldsymbol{M}_1)^\mathsf{T} \end{pmatrix}, \boldsymbol{K} = \begin{pmatrix} \|\boldsymbol{M}_2\|^2 - \|\boldsymbol{M}_1\|^2 \\ \vdots \\ \|\boldsymbol{M}_M\|^2 - \|\boldsymbol{M}_1\|^2 \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} d_{12} \\ \vdots \\ d_{1M} \end{pmatrix}, \tag{17}$$

and $\boldsymbol{D}^2$ is the component-wise square power of $\boldsymbol{D}$.

Notice that the matrix $\mathbf{M}$ is full rank if and only if the M microphones do not lie all in the same plane. This means that the problem has non-degenerated solution when the microphones are non co-planar. We will refer to that as the non-degeneracy condition of the problem. Assuming that the non-degeneracy condition is satisfied, we can solve the system of equations (linear in $\boldsymbol{S}$), by inversion:

$$\boldsymbol{S} = \mathbf{M}^\ddagger \left(\boldsymbol{K} - \boldsymbol{D}^2\right) - 2\mathbf{M}^\ddagger \boldsymbol{D}\, d_1 = \boldsymbol{A} d_1 + \boldsymbol{B}, \tag{18}$$

where $\boldsymbol{A} = -2\mathbf{M}^\ddagger \boldsymbol{D}$ and $\boldsymbol{B} = \mathbf{M}^\ddagger \left(\boldsymbol{K} - \boldsymbol{D}^2\right)$, and $\mathbf{M}^\ddagger$ is the pseudoinverse matrix of $\mathbf{M}$. We now substitute this solution inside the equation for d_1:

$$\begin{array}{rcl} d_1^2 & = & \left\langle \boldsymbol{S} - \boldsymbol{M}_1, \boldsymbol{S} - \boldsymbol{M}_1 \right\rangle \\ & = & \left\langle \boldsymbol{A} d_1 + \boldsymbol{B} - \boldsymbol{M}_1, \boldsymbol{A} d_1 + \boldsymbol{B} - \boldsymbol{M}_1 \right\rangle \\ & = & \|\boldsymbol{A}\|^2 d_1^2 + 2\left\langle \boldsymbol{A}, \boldsymbol{B} - \boldsymbol{M}_1 \right\rangle d_1 + \|\boldsymbol{B} - \boldsymbol{M}_1\|^2. \end{array} \tag{19}$$

This leads to a second order equation for d_1 which we can solve analytically and substitute in the equation for $\boldsymbol{S}$.

$$d_1^+ = \frac{\left\langle \boldsymbol{A}, \boldsymbol{B} - \boldsymbol{M}_1 \right\rangle + \sqrt{\Delta}}{\|\boldsymbol{A}\|^2 - 1} \qquad d_1^- = \frac{\left\langle \boldsymbol{A}, \boldsymbol{B} - \boldsymbol{M}_1 \right\rangle - \sqrt{\Delta}}{\|\boldsymbol{A}\|^2 - 1}, \tag{20}$$

where

$$\Delta = \left\langle \boldsymbol{A}, \boldsymbol{B} - \boldsymbol{M}_1 \right\rangle^2 - \|\boldsymbol{B} - \boldsymbol{M}_1\|^2 (\|\boldsymbol{A}\|^2 - 1). \tag{21}$$

The two solutions are:

$$\boldsymbol{S}^+ = \boldsymbol{A} d_1^+ + \boldsymbol{B} \qquad \boldsymbol{S}^+ = \boldsymbol{A} d_1^- + \boldsymbol{B}. \tag{22}$$

Since $\boldsymbol{D} = \nu \boldsymbol{t}$, we can rewrite Δ depending on $\boldsymbol{t}$: $\Delta(\boldsymbol{t})$. Notice that the condition for the existence of a solution is $\Delta \geq 0$; in such case we will call the set of TDE consistent. Also if $\Delta > 0$ two solutions exist. However, just one of them is correct. Actually, we observed that the following statement holds:

$$\left\{ \begin{array}{l} \boldsymbol{F}(\boldsymbol{S}^+) = \boldsymbol{F}(\boldsymbol{S}) \\ \boldsymbol{F}(\boldsymbol{S}^-) = -\boldsymbol{F}(\boldsymbol{S}) \end{array} \right\} \quad \text{or} \quad \left\{ \begin{array}{l} \boldsymbol{F}(\boldsymbol{S}^+) = -\boldsymbol{F}(\boldsymbol{S}) \\ \boldsymbol{F}(\boldsymbol{S}^-) = \boldsymbol{F}(\boldsymbol{S}) \end{array} \right\}$$

Geometrically speaking, we are looking fot the intersection of $M-1$ sheets of $M-1$ two-sheet hyperboloids. These intersect just in one point. Notice, however, that we can change the sign of the TDEs and we will get the same solutions to the equation. This means that we are considering also the intersection of the other $M-1$ one-sheet hyperboloids (the other sheets of the two-sheet hyperboloids). This new intersection provides us the second and fake solution.

6 Geometrically constrained criterion

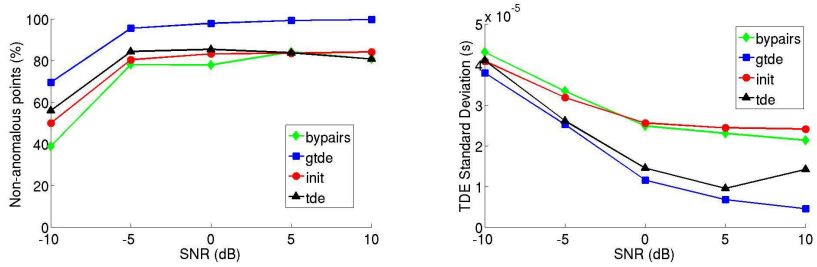
We would like, hence, to put together the signal prediction criterion subject to the geometric criterion. The optimization problem is set as:

$$\left\{ \begin{array}{l} (\boldsymbol{t}^*) = \arg\min\limits_{\boldsymbol{t}} J(\boldsymbol{t}) \\ \quad \text{s.t.} \quad \Delta(\boldsymbol{t}) \geq 0. \end{array} \right. \tag{23}$$

7 Implementation, Experiments & results

The minimization of (23) is carried out using a publicly available MATLAB implementation [6] of the *log-barrier interior point* method [2]. This method is designed for continuous convex optimization problems. On one hand, it is likely to fail in finding the global optimum of non-convex problems such as (23). To overcome this issue, the algorithm starts from several initial points, i.e., the set $\mathcal{S}^I = \{\boldsymbol{t}_i^I\}_{i=1}^P$. For each one of these initializations a local minimum is found, then the minimum over these local minima is selected. In our simulations, $P = 4096$ and $\mathcal{S}^I$ consists of points placed in a regular rectangular grid. Since the $t_{m,n}$'s have upper and lower bounds ($\|\boldsymbol{M}_m - \boldsymbol{M}_n\|$ and $-\|\boldsymbol{M}_m - \boldsymbol{M}_n\|$ respectively), the grid limits are defined by the geometry of the problem. On the other hand, the function to optimize must be continuous and the signals are discrete. We chose to compute the normalized cross-correlation function of the linear interpolation of the discrete signals.

In order to accurately evaluate and validate the proposed gTDE method, we developed a formal evaluation protocol using simulated data. A $3 \times 4 \times 2.5$ meter room, with uniform absorption coefficients, was simulated using the state-of-the-art Image-Source Model (ISM), [27] available at [26]. The main parameter of this model is T_{60}, which



(a) Good (non-anomalous) TDEs (b) TDE standard deviation

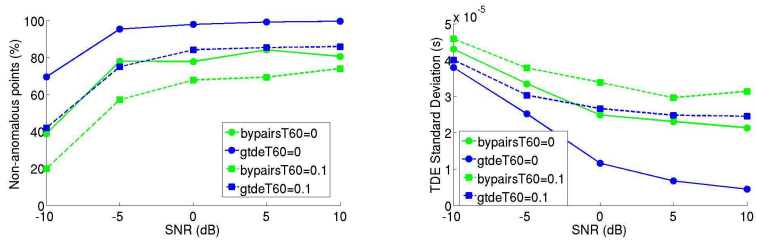
Figure 1: (Best seen in color) TDE method comparison based on random speech fragments. The plots show (a) the percentage of good TDEs and (c) the standard deviation as a function of SNR in an anechoic setup ($T_{60} = 0$).

corresponds to the time needed for an energy decay of 60 dB. We simulated four microphones placed at (in meters): $\boldsymbol{M}_1 = (2.35, 1.25, 1.179)^\mathsf{T}$, $\boldsymbol{M}_2 = (2.15, 1.25, 1.179)^\mathsf{T}$, $\boldsymbol{M}_3 = (2.25, 1.35, 1.32)^\mathsf{T}$ and $\boldsymbol{M}_4 = (2.25, 1.15, 1.32)^\mathsf{T}$, i.e., forming a regular tetrahedron. The sound source was placed at 27 different positions, namely all the possible 3-tuples $\boldsymbol{S} = (x, y, z)^\mathsf{T}$ with $x \in \{0.825, 1.5, 2.175\}$, $y \in \{1.1, 2, 2.9\}$ and $z \in \{0.6875, 1.25, 1.8125\}$ (in meters). The source emitted speech fragments randomly chosen from [17]. One hundred millisecond cuts of these sounds are the input of the evaluated method. We assume that only one source is emitting during these cuts. Finally the sensor noise, whose power depends on the chosen SNR, is added to these cuts.

Fig. 1 shows the results obtained with four different methods: pair-wise independent estimation of the time delays (*bypairs*), estimation using multichannel information *without* minimization (*init*), estimation based on unconstrained time delay minimization of (6) (*tde*), and the proposed geometrically-constrained minimization (*gtde*). Fig. 1(a) plots the percentage of good (non-anomalous) estimates, i.e., with an absolute error smaller than $100\,\mu$s and Fig. 1(b) plots the standard deviation of the good estimates as a function of the signal-to-noise ratio. These results correspond to an anechoic setup, $T_{60} = 0$. As expected, the proposed method significantly improves the percentage of good TDEs while it lowers down the standard deviation.

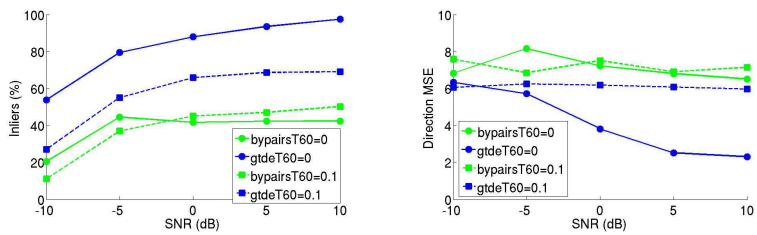
Additional simulations were carried out in order to precisely evaluate the performance of the gTDE method in the presence of noise and of reverberations. Figures 2 and 3 show the results on time delay estimation and sound source localization for different levels of noise and reverberations. In all the plots, the x-axis represents the SNR value (dB). The color corresponds to the method used: green for *bypairs* and blue for *gtde*; and the line style corresponds to the level of reverberation: solid-circle for $T_{60} = 0$ s and dashed-square for $T_{60} = 0.1$ s.

Regarding the TDE, in Fig. 2(a) the percentage of non-anomalous TDE is plot and in Fig. 2(b) the standard deviation of such time delays estimation is shown. Notice how the performance of both method improves with the SNR. Also, the proposed method



(a) Non-anomalous TDE (b) TDE Standard Deviation

Figure 2: Evaluation of the TDE performance of gTDE method. The x-axis corresponds to the SNR value (dB), the color to the methods (blue for *gtde*, green for *bypairs*), and the line style to the reverberation level (solid-circle for $T_{60} = 0$ and dashed-square for $T_{60} = 0.1\,\mathrm{s}$). (a) shows the percentage of non-anomalous TDE and (b) the standard deviation of this estimates.



(a) Percentage of inliers (b) Direction MSE

Figure 3: Evaluation of the localization performance of gTDE method. The x-axis corresponds to the SNR value (dB), the color to the methods (blue for *gtde*, green for *bypairs*), and the line style to the reverberation level (solid-circle for $T_{60} = 0$ and dashed-square for $T_{60} = 0.1\,\mathrm{s}$). (a) shows the percentage of localization inliers and (b) the mean squared error of localization error.

clearly outperforms the baseline. This is not a surprise since it is using all the available information to consistently estimate the TDEs at a time. A remarkable fact is that the proposed method under reverberant conditions has similar performance than the baseline method in the anechoic case. Also, with higher percentage of non-anomalous estimates, the *gtde* method has lower error standard deviation.

Concerning the localization, Fig. 3(a) plots the percentage of localization inliers and Fig. 3(b) the standard deviation of the angular error. A sound source is considered to be an inlier if the angular absolute error is less than $30\,^{\circ}$. As in the case of the TDE, the methods' performance improve with the SNR. Notice also that the proposed method outperforms the baseline, even when the proposed method is under echoic conditions and the baseline is under anechoic conditions.

Generally speaking, the methods perform as expected. The higher the SNR value the better the methods estimate the time delays, the higher the percentage of inliers and the lower the localization error. We can also see a clear trend with respect to the reverberation level: the methods' performance decreases with T_{60}. However this two

parameters have different effects on the function to minimize. On one side, the sensor noise decorrelates the microphones' signals leading to much more (and randomly spread) local minima and increasing the value of the true minimum. If this effect is extreme, the hope for a good estimate decreases fast. On the other side, the reverberations produce a little amount of strong local minima. This perturbation is systematic given the source position in the room. Hence, there is hope to learn the effect of such reverberations in order to improve the quality of the estimates. These two types of perturbations of the function to minimize clearly have different effects on the results. Notice that the reverberation level has almost no effect on the quality of the estimates when the SNR is low. However, when this random effect disappears (i.e., higher values of SNR) a systematic and significative difference appears both in the time delay and in the localization estimates.

Finally, the authors would like to remark on the method's robustness. Notice that, for moderate levels of noise and reverberations ($T_{60} \leq 100\,\text{ms}$, SNR $\geq 0\,\text{dB}$) the method is able to localize the sound source with mean angular squared error of $6\,^{\circ}$ in more than $60\,\%$ of the cases.

8 Conclusions & future work

In this paper, a new method on time delay estimation for sound source localization working on non-coplanar microphone arrays is presented. The estimation is cast into a multivariate optimization problem. In addition, the geometric model for the time delays add a non-linear constraint. The optimal value of the problem is a consistent set of time delays, useful to localize the sound source. Experiments on simulated data show the quality of the method and validate the approach.

The are several ways to extend this work. As outlined before, it would be very useful to learn the effect the reverberations have on the objective function as in [30]. Also, it is worth to consider the multiple source case, following approaches like [11]. Besides that, a frequency decomposition stage may be useful to avoid the analysis in non-informative frequency bands ([38]). Also, experiments on more reverberant data, and on real data have to be done in order to explore the real extent of these initial results. Last but not least, it would be interesting to explore cases in which the microphones' positions have some error, and see how to adapt the method to estimate the time delays and correct the microphones' positions.

A Criteria equivalence

This appendix is devoted to prove the equivalence of both optimization problems: (5) and (6). We will start from the expression of the second criterion and get to the expres-

sion of the first (original) one. The objective function to minimize is:

$$\det\left(\tilde{\mathbf{R}}\right) = \det\left(\mathbf{R}\right) - \sum_{i=2}^{M}(-1)^{i+1}\rho_{1,i}\det\left(\mathbf{R}_i\right),$$

where

$$\mathbf{R}_i = \begin{pmatrix} \boldsymbol{r} & \boldsymbol{R}_1 & \cdots & \boldsymbol{R}_{i-1} & \boldsymbol{R}_{i+1} & \cdots & \boldsymbol{R}_M \end{pmatrix}.$$

begin the $\boldsymbol{R}_i$ the i-th column of the matrix $\mathbf{R}$. Notice now that:

$$\det\left(\mathbf{R}_i\right) = \sum_{j=2}^{M}(-1)^{j+1}\rho_{1,j}\det(\mathbf{R}_{i,j}),$$

where $\det(\mathbf{R}_{i,j})$ is the (i,j)-th minor of $\mathbf{R}$ (since $\mathbf{R}$ is symmetric). Finally,

$$\begin{aligned}
\det\left(\tilde{\mathbf{R}}\right) &= \det\left(\mathbf{R}\right) - \sum_{i,j=2}^{M}(-1)^{i+j}\rho_{1,i}\rho_{1,j}\det\left(\mathbf{R}_{i,j}\right) \\
&= \det\left(\mathbf{R}\right)\left(1 - \sum_{i,j=2}^{M}(-1)^{i+j}\rho_{1,i}\rho_{1,j}\frac{\det\left(\mathbf{R}_{i,j}\right)}{\det\left(\mathbf{R}\right)}\right) \\
&= \det\left(\mathbf{R}\right)\left(1 - \sum_{i,j=2}^{M}(-1)^{i+j}R_{1,i}R_{1,j}\frac{\det\left(\mathbf{R}_{i,j}\right)}{\det\left(\mathbf{R}\right)E_1\sqrt{E_i}\sqrt{E_j}}\right) \\
&= \frac{\det\left(\mathbf{R}\right)}{E_1}\left(E_1 - \sum_{i,j=2}^{M}(-1)^{i+j}R_{1,i}\left(\mathbf{R}^{-1}\right)_{i,j}R_{1,j}\right) \\
&= \frac{\det\left(\mathbf{R}\right)}{E_1}\left(E_1 - \boldsymbol{r}^{\mathsf{T}}\mathbf{R}^{-1}\boldsymbol{r}\right)
\end{aligned}$$

So both criteria are equivalent if $\det(\mathbf{R})$ and E_1 are well defined and different than zero. What does it mean? Explain that correctly.

B Deriving formulaes for $\nabla\tilde{J}$ and $\mathbf{H}\tilde{J}$

In this section we will derive the formulae for the gradient and the Hessian of the criterion in (6). To do that we will need some matrix calculus rules. Let $\mathbf{Y} : \mathbb{R} \Rightarrow \mathbb{R}^{M\times M}$, be a matrix function depending on y, the following formulas hold:

- $\dfrac{\partial\,\det\left(\mathbf{Y}(y)\right)}{\partial y} = \det(\mathbf{Y}(y))\text{trace}\left(\mathbf{Y}(y)^{-1}\dfrac{\partial\mathbf{Y}(y)}{\partial y}\right)$

- $\dfrac{\partial \,\text{trace}\,\left(\mathbf{Y}(y)\right)}{\partial y} = \text{trace}\left(\dfrac{\partial\,\mathbf{Y}(y)}{\partial y}\right)$
- $\dfrac{\partial\,\mathbf{Y}(y)^{-1}}{\partial y} = -\mathbf{Y}(y)^{-1}\dfrac{\partial \mathbf{Y}(y)}{\partial y}\mathbf{Y}(y)^{-1}$

Recall that the function we want to derivate is $\tilde{J} = \det\left(\tilde{\mathbf{R}}\right)$. We can then compute:

$$\frac{\partial\,\tilde{J}}{\partial t_{1,k}} = \frac{\partial\,\det\left(\tilde{\mathbf{R}}\right)}{\partial t_{1,k}} = \det\left(\tilde{\mathbf{R}}\right)\text{trace}\left(\tilde{\mathbf{R}}^{-1}\frac{\partial\,\tilde{\mathbf{R}}}{\partial t_{1,k}}\right). \tag{24}$$

We can also compute the second derivative of the criterion:

$$\begin{aligned}\frac{\partial^2\,\tilde{J}}{\partial t_{1,j}\partial t_{1,k}} &= \frac{\partial}{\partial t_{1,j}}\left(\det\left(\tilde{\mathbf{R}}\right)\text{trace}\left(\tilde{\mathbf{R}}^{-1}\frac{\partial\,\tilde{\mathbf{R}}}{\partial t_{1,k}}\right)\right)\\ &= \det\left(\tilde{\mathbf{R}}\right)\text{trace}\left(\tilde{\mathbf{R}}^{-1}\frac{\partial\,\tilde{\mathbf{R}}}{\partial t_{1,j}}\right)\text{trace}\left(\tilde{\mathbf{R}}^{-1}\frac{\partial\,\tilde{\mathbf{R}}}{\partial t_{1,k}}\right)+\\ &\quad \det\left(\tilde{\mathbf{R}}\right)\text{trace}\left(-\tilde{\mathbf{R}}^{-1}\frac{\partial\,\tilde{\mathbf{R}}}{\partial t_{1,j}}\tilde{\mathbf{R}}^{-1}\frac{\partial\,\tilde{\mathbf{R}}}{\partial t_{1,j}}+\tilde{\mathbf{R}}^{-1}\frac{\partial^2\,\tilde{\mathbf{R}}}{\partial t_{1,j}\partial t_{1,k}}\right).\end{aligned}$$

Hence, to be able to evaluate the gradient and the Hessian of $\tilde{J}$ we need to compute the first and second derivatives of the matrix $\tilde{\mathbf{R}}$. We will first compute these derivatives in function of the derivatives of $\mathbf{R}$ and $\boldsymbol{r}$ to finally compute the derivatives of $\mathbf{R}$ and $\boldsymbol{r}$. We can rewrite $\tilde{\mathbf{R}}$ as:

$$\tilde{\mathbf{R}} = \mathbf{D}\left(\begin{array}{c|c} E_1 & \boldsymbol{r}^{\mathsf{T}} \\ \hline \boldsymbol{r} & \mathbf{R}\end{array}\right)\mathbf{D}, \tag{25}$$

where

$$\mathbf{D} = \begin{pmatrix} E_1^{-1/2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & E_M^{-1/2}\end{pmatrix}. \tag{26}$$

Notice that the matrix $\mathbf{D}$ does not depend on $\boldsymbol{t}$. Hence the expressions for the derivatives of $\tilde{\mathbf{R}}$ are:

$$\frac{\partial\tilde{\mathbf{R}}}{\partial t_{1,k}} = \mathbf{D}\left(\begin{array}{c|c} 0 & \dfrac{\partial \boldsymbol{r}}{\partial t_{1,k}}^{\mathsf{T}} \\ \hline \dfrac{\partial \boldsymbol{r}}{\partial t_{1,k}} & \dfrac{\partial \mathbf{R}}{\partial t_{1,k}}\end{array}\right)\mathbf{D}, \tag{27}$$

and

$$\frac{\partial^2 \tilde{\mathbf{R}}}{\partial t_{1,j} \partial t_{1,k}} = \mathbf{D} \left(\begin{array}{c|c} 0 & \frac{\partial^2 \boldsymbol{r}}{\partial t_{1,j} \partial t_{1,k}}^{\mathsf{T}} \\ \hline \frac{\partial^2 \boldsymbol{r}}{\partial t_{1,j} \partial t_{1,k}} & \frac{\partial^2 \mathbf{R}}{\partial t_{1,j} \partial t_{1,k}} \end{array} \right) \mathbf{D}. \tag{28}$$

So we just need the first and second derivatives of the vector $\boldsymbol{r}$ and the matrix $\mathbf{R}$. Recall their expressions:

$$\mathbf{R} = \begin{pmatrix} R_{2,2}(0) & R_{2,3}(t_{1,3} - t_{1,2}) & \cdots & R_{2,M}(t_{1,M} - t_{1,2}) \\ R_{2,3}(t_{1,3} - t_{1,2}) & R_{3,3}(0) & \cdots & R_{3,M}(t_{1,M} - t_{1,3}) \\ \vdots & \vdots & \ddots & \vdots \\ R_{2,M}(t_{1,M} - t_{1,2}) & R_{3,M}(t_{1,M} - t_{1,3}) & \cdots & R_{M,M}(0) \end{pmatrix},$$

and

$$\boldsymbol{r} = \begin{pmatrix} R_{1,2}(-t_{1,2}) \\ R_{1,3}(-t_{1,3}) \\ \vdots \\ R_{1,M}(-t_{1,M}) \end{pmatrix}.$$

The partial derivative of $\mathbf{R}$ is matrix filled with zeros expect for its k-th row and its k-th column that are equal to the following vector:

$$\left(\frac{\partial \mathbf{R}}{\partial t_{1k}} \right)_k = \begin{pmatrix} R'_{2,k}(t_{1,k} - t_{1,2}) \\ \vdots \\ R'_{k-1,k}(t_{1,k} - t_{1,k-1}) \\ 0 \\ -R'_{k+1,k}(t_{1,k+1} - t_{1,k}) \\ \vdots \\ -R'_{M,k}(t_{1,M} - t_{1,k}) \end{pmatrix}. \tag{29}$$

The partial derivative of $\boldsymbol{r}$ is:

$$\frac{\partial \boldsymbol{r}}{\partial t_{1,k}} = \begin{pmatrix} 0 \\ \vdots \\ -R'_{1,k}(-t_{1,k}) \\ \vdots \\ 0 \end{pmatrix} \tag{30}$$

We will differentiate two cases when computing the second derivative:

$j = k$ This will fill the diagonal of the hessian matrix. Notice that:

$$\frac{\partial^2 \boldsymbol{r}}{\partial t_{1,k}^2} = \begin{pmatrix} 0 \\ \vdots \\ R''_{1,k}(-t_{1,k}) \\ \vdots \\ 0 \end{pmatrix} \tag{31}$$

and that the partial second derivative of $\mathbf{R}$ is matrix filled with zeros expect for its k-th row and its k-th column that are equal to the following vector:

$$\left(\frac{\partial^2 \mathbf{R}}{\partial t_{1k}^2}\right)_k = \begin{pmatrix} R''_{2,k}(t_{1,k} - t_{1,2}) \\ \vdots \\ R''_{k-1,k}(t_{1,k} - t_{1,k-1}) \\ 0 \\ R''_{k+1,k}(t_{1,k+1} - t_{1,k}) \\ \vdots \\ R''_{M,k}(t_{1,M} - t_{1,k}) \end{pmatrix}. \tag{32}$$

$j > k$ This fills the lower triangular matrix of the hessian (and the upper triangular since we assume that the hessian is symmetric, i.e., $\tilde{J}$ is twice continuously differentiable). The second derivative of $\boldsymbol{r}$ is null in this case, however the second derivative of $\mathbf{R}$ is not. Actually just two positions in the second derivative are not necessarily null: the jk-th and the kj-th being:

$$\left(\frac{\partial^2 \mathbf{R}}{\partial t_{1,j} \partial t_{1,k}}\right)_{kj} = -R''_{k,j}(t_{1,k} - t_{1,j}) \tag{33}$$

C Estimating of the cross-correlation function

The previous derivation needs the cross-correlation function to be twice continuously differentiable. However, we just have the discrete signals. A natural question arises: how can we estimate the cross-correlation function from two discrete signals? We have chosen to compute the cross-correlation function of the interpolated signals. The following prsents how to estimate the cross-correlation function from some polynomial interpolation of two discrete signals.

Let x and y be two continuous signals. We'll denote their discretizations by $x[n]$ and $y[n]$. We assume regular sampling at a rate F (and denote the sampling period $T = 1/F$). The n-th sampling time will be denoted by $t_n = nT$. We want to estimate the cross-correlation function of the signals x and y from the interpolation of $x[n]$ and $y[n]$.

Let us assume a D-degree polynomial interpolation for each time interval:

$$\hat{x}(t) = x_n(t) = \sum_{d=0}^{D} a_{n,d}\,(t-t_n)^d \qquad \hat{y}(t) = y_n(t) = \sum_{d=0}^{D} b_{n,d}\,(t-t_n)^d. \tag{34}$$

The task is to compute $R_{\hat{x}\hat{y}}(\tau)$. Notice that:

$$\begin{aligned} R_{\hat{x}\hat{y}}(\tau) &= \int \hat{x}(t)\hat{y}(t-\tau)\mathrm{d}t && (35)\\ &= \sum_n \int_{t_n}^{t_{n+1}} \hat{x}(t)\hat{y}(t-\tau)\mathrm{d}t && (36)\\ &= \sum_n \int_{t_n}^{t_{n+1}} x_n(t)\hat{y}(t-\tau)\mathrm{d}t. && (37) \end{aligned}$$

Let us now decompose the delay in $\tau = mT + \hat{\tau}$, with $0 \leq \hat{\tau} < T$. We have the following relations:

$$\begin{array}{lll} t \in (t_n, t_n+\hat{\tau}) & \Rightarrow t-\tau \in (t_{n-m}-\hat{\tau}, t_{n-m}) & \subset (t_{n-m-1}, t_{n-m}) \\ t \in (t_n+\hat{\tau}, t_{n+1}) & \Rightarrow t-\tau \in (t_{n-m}, t_{n+1}-t_m-\hat{\tau}) & \subset (t_{n-m}, t_{n+1-m}) \end{array} \tag{38}$$

That allows us to split the integral in two:

$$R_{\hat{x}\hat{y}}(t) = \sum_n \underbrace{\int_{t_n}^{t_n+\hat{\tau}} x_n(t)y_{n-m-1}(t-\tau)\mathrm{d}t}_{I_1} + \underbrace{\int_{t_n+\hat{\tau}}^{t_{n+1}} x_n(t)y_{n-m}(t-\tau)\mathrm{d}t}_{I_2} \tag{39}$$

We will know develop the two terms of the sum. The first term:

$$\begin{aligned} I_1 &= \int_{t_n}^{t_n+\hat{\tau}} x_n(t)y_{n-m-1}(t-\tau)\mathrm{d}t && (40)\\ &= \sum_{p,q=0}^{D} a_{n,p}b_{n-m-1,q} \int_{t_n}^{t_n+\hat{\tau}} (t-t_n)^p(t-\tau-t_{n-m-1})^q\mathrm{d}t && (41)\\ &= \sum_{p,q=0}^{D} a_{n,p}b_{n-m-1,q} \int_{t_n}^{t_n+\hat{\tau}} (t-t_n)^p(t-t_m-\hat{\tau}-t_n+_m+T)^q\mathrm{d}t && (42)\\ &= \sum_{p,q=0}^{D} a_{n,p}b_{n-m-1,q} \int_{0}^{\hat{\tau}} t^p(t+T-\hat{\tau})^q\mathrm{d}t && (43)\\ &= \sum_{p,q=0}^{D} a_{n,p}b_{n-m-1,q} \sum_{k=0}^{q}\binom{q}{k} \int_{0}^{\hat{\tau}} t^p t^k(T-\hat{\tau})^{q-k}\mathrm{d}t && (44)\\ &= \sum_{p,q=0}^{D} a_{n,p}b_{n-m-1,q} \sum_{k=0}^{q}\binom{q}{k}(T-\hat{\tau})^{q-k}\frac{\hat{\tau}^{k+p+1}}{k+p+1} && (45)\\ &= \sum_{p,q=0}^{D} a_{n,p}b_{n-m-1,q}\, K_1(q,p,\hat{\tau},T) && (46) \end{aligned}$$

And the second term:

$$
\begin{aligned}
I_2 &= \int_{t_n+\hat{\tau}}^{t_{n+1}} x_n(t) y_{n-m}(t-\tau)\mathrm{d}t && (47)\\
&= \sum_{p,q=0}^{D} a_{n,p} b_{n-m,q} \int_{t_n+\hat{\tau}}^{t_{n+1}} (t-t_n)^p (t-\tau-t_{n-m})^q \mathrm{d}t && (48)\\
&= \sum_{p,q=0}^{D} a_{n,p} b_{n-m,q} \int_{t_n+\hat{\tau}}^{t_{n+1}} (t-t_n)^p (t-\hat{\tau}-t_n)^q \mathrm{d}t && (49)\\
&= \sum_{p,q=0}^{D} a_{n,p} b_{n-m,q} \int_{\hat{\tau}}^{T} t^p (t-\hat{\tau})^q \mathrm{d}t && (50)\\
&= \sum_{p,q=0}^{D} a_{n,p} b_{n-m,q} \sum_{k=0}^{q} \binom{q}{k} \int_{\hat{\tau}}^{T} t^p t^k (-\hat{\tau})^{q-k} \mathrm{d}t && (51)\\
&= \sum_{p,q=0}^{D} a_{n,p} b_{n-m,q} \sum_{k=0}^{q} \binom{q}{k} (-\hat{\tau})^{q-k} \frac{T^{k+p+1}-\hat{\tau}^{k+p+1}}{k+p+1} && (52)\\
&= \sum_{p,q=0}^{D} a_{n,p} b_{n-m,q}\, K_2(q,p,\hat{\tau},T) && (53)
\end{aligned}
$$

We can plug these two equations into the original cross-correlation equation and we get:

$$
\begin{aligned}
R_{\hat{x}\hat{y}}(\tau) &= \sum_n \left(\sum_{p,q=0}^{D} a_{n,p} b_{n-m-1,q}\, K_1(q,p,\hat{\tau},T) + \right.\\
&+ \left. \sum_{p,q=0}^{D} a_{n,p} b_{n-m,q}\, K_2(q,p,\hat{\tau},T) \right)\\
&= \sum_{p,q=0}^{D} R_{a_p,b_q}[m+1]\, K_1(q,p,\hat{\tau},T) + R_{a_p,b_q}[m]\, K_2(q,p,\hat{\tau},T)
\end{aligned}
$$

From this equation we can also compute the derivative of the cross-correlation:

$$
R'_{\hat{x}\hat{y}}(\tau) = \sum_{p,q=0}^{D} R_{a_p,b_q}[m+1]\, \frac{\partial K_1(q,p,\hat{\tau},T)}{\partial \hat{\tau}} + R_{a_p,b_q}[m]\, \frac{\partial K_2(q,p,\hat{\tau},T)}{\partial \hat{\tau}} \quad (54)
$$

as well as the second derivative:

$$
R''_{\hat{x}\hat{y}}(\tau) = \sum_{p,q=0}^{D} R_{a_p,b_q}[m+1]\, \frac{\partial^2 K_1(q,p,\hat{\tau},T)}{\partial \hat{\tau}^2} + R_{a_p,b_q}[m]\, \frac{\partial^2 K_2(q,p,\hat{\tau},T)}{\partial \hat{\tau}^2}. \quad (55)
$$

The formulas for the partial derivatives are:

$$
\frac{\partial K_1}{\partial \hat{\tau}} = \sum_{k=0}^{q} \binom{q}{k} (T-\hat{\tau})^{q-k-1} \hat{\tau}^{k+p} \left(T - \hat{\tau}\frac{q+p+1}{k+p+1} \right) \quad (56)
$$

$$\frac{\partial^2 K_1}{\partial \hat{\tau}^2} = \sum_{k=0}^{q} \binom{q}{k} (T-\hat{\tau})^{q-k-2} \hat{\tau}^{k+p-1} \left(\hat{\tau}^2 \frac{(q+p+1)(q+p)}{k+p+1} - 2\hat{\tau}T(q+p) + T^2(k+p) \right) \tag{57}$$

$$\frac{\partial K_2}{\partial \hat{\tau}} = \sum_{k=0}^{q} \binom{q}{k} (-\hat{\tau})^{q-k-1} \left((k-q) \frac{T^{k+p+1} - \hat{\tau}^{k+p+1}}{k+p+1} + \hat{\tau}^{k+p+1} \right). \tag{58}$$

$$\frac{\partial^2 K_2}{\partial \hat{\tau}^2} = \sum_{k=0}^{q} \binom{q}{k} (-\hat{\tau})^{q-k-2} \left(\frac{(1+k-q)(k-q)}{k+p+1} (T^{k+p+1} - \hat{\tau}^{k+p+1}) + (k-p-2q)\hat{\tau}^{k+p+1} \right) \tag{59}$$

D Deriving the $\nabla\Delta$ and $\mathbf{H}\Delta$

$$\begin{aligned} \nabla\Delta &= 2\Big(\langle \boldsymbol{A}, \boldsymbol{B} - \boldsymbol{M}_1 \rangle \left(\mathbf{J}_{\boldsymbol{A}}^{\mathsf{T}} (\boldsymbol{B} - \boldsymbol{M}_1) + \mathbf{J}_{\boldsymbol{B}}^{\mathsf{T}} \boldsymbol{A} \right) - \\ &- (\|\boldsymbol{A}\|^2 - 1) \mathbf{J}_{\boldsymbol{B}}^{\mathsf{T}} (\boldsymbol{B} - \boldsymbol{M}_1) - \|\boldsymbol{B} - \boldsymbol{M}_1\|^2 \mathbf{J}_{\boldsymbol{A}}^{\mathsf{T}} \boldsymbol{A} \Big) \end{aligned}$$

where $\mathbf{J}_{\boldsymbol{A}} = -2\nu \mathbf{M}^{\ddagger}$ and $\mathbf{J}_{\boldsymbol{B}} = -2\nu^2 \mathbf{M}^{\ddagger} \operatorname{diag}(\boldsymbol{t})$, being ν the sound speed. We can also compute the Hessian of Δ:

$$\begin{aligned} \mathbf{H}\Delta &= 2\Bigg(\left(\mathbf{J}_{\boldsymbol{A}}^{\mathsf{T}} (\boldsymbol{B} - \boldsymbol{M}_1) + \mathbf{J}_{\boldsymbol{B}}^{\mathsf{T}} \boldsymbol{A} \right) \left(\mathbf{J}_{\boldsymbol{A}}^{\mathsf{T}} (\boldsymbol{B} - \boldsymbol{M}_1) + \mathbf{J}_{\boldsymbol{B}}^{\mathsf{T}} \boldsymbol{A} \right)^{\mathsf{T}} + \\ &+ \langle \boldsymbol{A}, \boldsymbol{B} - \boldsymbol{M}_1 \rangle \left(\mathbf{J}_{\boldsymbol{A}}^{\mathsf{T}} \mathbf{J}_{\boldsymbol{B}} + \mathbf{D} + \mathbf{J}_{\boldsymbol{B}}^{\mathsf{T}} \mathbf{J}_{\boldsymbol{A}} \right) - \\ &- \Big[2(\mathbf{J}_{\boldsymbol{B}}^{\mathsf{T}} (\boldsymbol{B} - \boldsymbol{M}_1))(\mathbf{J}_{\boldsymbol{A}}^{\mathsf{T}} \boldsymbol{A})^{\mathsf{T}} + (\|\boldsymbol{A}\|^2 - 1)(\mathbf{E} + \mathbf{J}_{\boldsymbol{B}}^{\mathsf{T}} \mathbf{J}_{\boldsymbol{B}}) + \\ &+ 2(\mathbf{J}_{\boldsymbol{A}}^{\mathsf{T}} \boldsymbol{A})(\mathbf{J}_{\boldsymbol{B}}^{\mathsf{T}} (\boldsymbol{B} - \boldsymbol{M}_1))^{\mathsf{T}} + \|\boldsymbol{B} - \boldsymbol{M}_1\|^2 \mathbf{J}_{\boldsymbol{A}}^{\mathsf{T}} \mathbf{J}_{\boldsymbol{A}} \Big] \Bigg) \end{aligned}$$

where

$$\mathbf{D} = -2\nu^2 \operatorname{Diag}(\mathbf{M}^{\dagger} \boldsymbol{A}) \qquad \mathbf{E} = -2\nu^2 \operatorname{Diag}(\mathbf{M}^{\dagger} (\boldsymbol{B} - \boldsymbol{M}_1)). \tag{60}$$

E Analytic cross-correlation function

In this section we compute the closed-form cross-correlation function for two different signals:

The exponentially decaied sinusoidal:

$$x(t) = \exp^{-t/\lambda} \sin(Ft). \tag{61}$$

The full spectrum sinusoidal:

$$y(t) = \sum_{n=1}^{N} a_i \sin(F_i t + \phi_i). \tag{62}$$

We want to compute $R_{z_1,z_2}(\tau)$, where z is either x or y and z_i means that the signal is delaied by t_i. To be more precise, given the original signal z, we write the two following "acquisitions" of the signal:

$$z_i(t) = \begin{cases} x(t-t_i) & t_i \leq t \leq L + t_i \\ 0 & \text{otherwise} \end{cases} \tag{63}$$

and the goal is to derive a closed-form solution for $R_{z_1,z_2}(\tau)$.

E.1 Exponentially decaied sinusoidal

We write, then:

$$\begin{aligned} R_{x_1,x_2}(\tau) &= \int_{\mathbb{R}} x_1(t)x_2(t-\tau)\mathrm{d}t \\ &= \int_m^M x_1(t)x_2(t-\tau)\mathrm{d}t \\ &= \int_m^M \exp^{-(2t-t_1-t_2-\tau)/\lambda} \sin(F(t-t_1))\sin(F(t-t_2-\tau))\mathrm{d}t \\ &= \frac{1}{2}\int_m^M \exp^{-(2t-t_1-t_2-\tau)/\lambda} \cos(F(t_2+\tau-t_1))\mathrm{d}t - \\ &- \frac{1}{2}\int_m^M \exp^{-(2t-t_1-t_2-\tau)/\lambda} \cos(F(2t-t_1-t_2-\tau))\mathrm{d}t \end{aligned}$$

where

$$m = \max\{t_1, t_2+\tau\}$$
$$M = \min\{T+t_1, T+t_2+\tau\}$$

We just need to compute the following general integral:

$$\begin{aligned} I &= \int_m^M \cos(F(t-t_c)) \exp^{-(t-t_c)/\lambda} \mathrm{d}t \\ &= \int_{m-t_c}^{M-t_c} \cos(F\,t') \exp^{-t'/\lambda} \mathrm{d}t. \end{aligned}$$

The integral above is computed by applyint integration by parts twice:

$$\int \cos(F\,t) \exp^{-t/\lambda} \mathrm{d}t = \frac{\lambda}{1+(\lambda F)^2} \exp^{-t/\lambda}(\cos(F\,t) - \lambda F \sin(F\,t)). \tag{64}$$

We finally have all the ingredients to compute the cross-correlation function:

$$\begin{aligned} R_{x_1,x_2}(\tau) &= \frac{\lambda}{4}\left.\frac{\exp^{-t/\lambda}(\cos(Ft)-F\lambda\sin(Ft))}{1+(\lambda F)^2}\right|_{2m-t_1-t_2-\tau}^{2M-t_1-t_2-\tau} - \\ &\quad -\frac{\lambda}{4}\exp^{(t_1+t_2+\tau)/\lambda}\cos(F(t_2+\tau-t_1))\left.\exp^{-t/\lambda}\right|_{2m}^{2M} \end{aligned}$$

E.2 The full spectrum sinusoidal

The full spectrum signal we consider has the following expression:

$$y(t) = \sum_{n=1}^{N} a_i \sin(F_i t + \phi_i) \tag{65}$$

where the a_i are drawn from a standard uniform distribution, the F_i correspond to divisors of the sampling frequency $F_i = iF/L$, where F is the sampling frequency and L is the acquisition length in samples (notice that in this case $N < L/2$ since higher frequencies will not be present due to the Nyquist theorem).

Notice that, we do not want to compute the expected cross-correlation of this signal, but the correlation of the acquired signals as if they were deterministic. This computation aims to have the ground truth to test the criterion we proposed, it is different from deriving general properties of this signal.

As in the previous case we can write:

$$\begin{aligned} R_{y_1,y_2}(\tau) &= \int_m^M y_1(t)y_2(t-\tau)\mathrm{d}t \\ &= \sum_{i,j=1}^{N} a_i a_j \int_m^M \cos(F_i(t-t_1)+\phi_i)\cos(F_j(t-t_2-\tau)+\phi_j)\mathrm{d}t \end{aligned}$$

So we need to solve the following general integral:

$$\begin{aligned} I &= \int \cos(\alpha t+\beta)\cos(\gamma(t+\delta)+\epsilon)\mathrm{d}t \\ &= \frac{1}{2}\underbrace{\int \cos((\alpha-\gamma)\,t+\beta-\gamma\delta-\epsilon)\mathrm{d}t}_{I_A} + \\ &\quad \frac{1}{2}\underbrace{\int \cos((\alpha+\gamma)\,t+\beta+\gamma\delta+\epsilon)\mathrm{d}t}_{I_B} \end{aligned}$$

Those two integrals can be expressed as:

$$I_A = \begin{cases} \dfrac{\sin((\alpha-\gamma)\,t+\beta-\gamma\delta-\epsilon)}{\alpha-\gamma} & \alpha \neq \gamma \\ t\,\cos(\beta-\gamma\delta-\epsilon) & \alpha = \gamma \end{cases} \tag{66}$$

$$I_B = \begin{cases} \dfrac{\sin((\alpha+\gamma)\,t+\beta+\gamma\delta+\epsilon)}{\alpha+\gamma} & \alpha+\gamma \neq 0 \\ t\,\cos(\beta+\gamma\delta+\epsilon) & \alpha+\gamma = 0 \end{cases} \tag{67}$$

We can now compute the cross-correlation function:

$$\begin{aligned}
R_{y_1,y_2}(\tau) &= \sum_{i,j=1}^{N} a_i a_j \int_m^M \cos(F_i(t-t_1)+\phi_i)\cos(F_j(t-t_2-\tau)+\phi_j)\mathrm{d}t \\
&= \sum_{i,j=1\ j\neq i}^{N} a_i a_j \int_m^M \cos(F_i(t-t_1)+\phi_i)\cos(F_j(t-t_2-\tau)+\phi_j)\mathrm{d}t + \\
&+ \sum_{i=1} a_i^2 \int_m^M \cos(F_i(t-t_1)+\phi_i)\cos(F_i(t-t_2-\tau)+\phi_i)\mathrm{d}t \\
&= \sum_{i,j=1\ j\neq i}^{N} a_i a_j \int_{m-t_1}^{M-t_1} \cos(F_i t+\phi_i)\cos(F_j(t+t_1-t_2-\tau)+\phi_j)\mathrm{d}t + \\
&+ \sum_{i=1} a_i^2 \int_{m-t_1}^{M-t_1} \cos(F_i t+\phi_i)\cos(F_i(t+t_1-t_2-\tau)+\phi_i)\mathrm{d}t \\
&= \sum_{i,j=1\ j\neq i}^{N} a_i a_j \left(\frac{\sin((F_i-F_j)\,t+\phi_i-F_m(t_1-t_2-\tau)-\phi_j)}{F_i-F_j}\Bigg|_{t=m-t_1}^{t=M-t_1} + \right. \\
&+ \left. \frac{\sin((F_i+F_j)\,t+\phi_i+F_j(t_1-t_2-\tau)+\phi_j)}{F_i+F_j}\Bigg|_{t=m-t_1}^{t=M-t_1} \right) + \\
&+ \sum_{i=1} a_i^2 \left(t\,\cos(F_i(t_1-t_2-\tau))\Big|_{t=m-t_1}^{t=M-t_1} + \right. \\
&+ \left. \frac{\sin(2(F_i\,t+\phi_i)+F_i(t_1-t_2-\tau))}{2F_i}\Bigg|_{t=m-t_1}^{t=M-t_1} \right)
\end{aligned}$$

References

[1] A. Beck, P. Stoica, and J. Li. Exact and approximate solutions of source localization problems. *Signal Processing, IEEE Transactions on*, 56(5):1770–1778, 2008.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] M. Brandstein, J. Adcock, and H. Silverman. A closed-form location estimator for use with room environment microphone arrays. *Speech and Audio Processing, IEEE Transactions on*, 5(1):45–50, 1997.

[4] M. Brandstein and H. Silverman. A practical methodology for speech source localization with microphone arrays. *Computer Speech & Language*, 11(2):91–126, 1997.

[5] A. Brutti, M. Omologo, P. Svaizer, and F. Bruno. Comparison between different sound source localization techniques. In *Hands-Free Speech Communication and Microphone Arrays*, pages 69–72, 2008.

[6] P. Carbonetto. MATLAB primal-dual interior-point solver for convex programs with constraints, 2008. http://www.cs.ubc.ca/ pcarbo/convexprog.html.

[7] Y. Chan and K. Ho. A simple and efficient estimator for hyperbolic location. *Signal Processing, IEEE Transactions on*, 42(8):1905–1915, 1994.

[8] Y. Chan, W. Tsui, H. So, and P. Ching. Time-of-arrival based localization under NLOS conditions. *Vehicular Technology, IEEE Transactions on*, 55(1):17–24, 2006.

[9] J. Chen, J. Benesty, and Y. Huang. Robust time delay estimation exploiting redundancy among multiple microphones. *Speech and Audio Processing, IEEE Transactions on*, 11(6):549–557, 2003.

[10] J. Chen, J. Benesty, and Y. A. Huang. Time Delay Estimation in Room Acoustic Environments: An Overview. *EURASIP Journal on Advances in Signal Processing*, 2006(i):1–20, 2006.

[11] J. Chen, K. Yao, and R. Hudson. Acoustic source localization and beamforming: theory and practice. *EURASIP Journal on Applied Signal Processing*, pages 359–370, 2003.

[12] K. Cheung, H. So, W. Ma, and Y. Chan. Least squares algorithms for time-of-arrival-based mobile location. *Signal Processing, IEEE Transactions on*, 52(4):1121–1130, 2004.

[13] S. Doclo and M. Moonen. Robust Adaptive Time Delay Estimation for Speaker Localization in Noisy and Reverberant Acoustic Environments. *EURASIP Journal on Advances in Signal Processing*, 2003(11):1110–1124, 2003.

[14] S. Drake and K. Dogançay. Geolocation by time difference of arrival using hyperbolic asymptotes. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 2, pages ii–361. IEEE, 2004.

[15] B. Friedlander. A passive localization algorithm and its accuracy analysis. *Oceanic Engineering, IEEE Journal of*, 12(1):234–245, 1987.

[16] G. Galati, M. Gasbarra, P. Magaro, P. Marco, L. Mene, and M. Pici. New Approaches to Multilateration processing: analysis and field evaluation. In *2006 European Radar Conference*, volume 9, pages 116–119. Ieee, Sept. 2006.

[17] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue. Timit acoustic-phonetic continuous speech corpus, 1993. Linguistic Data Consortium, Philadelphia.

[18] A. A. Handzel. High Acuity Sound-Source Localization by means of a Triangular Spherical Array. In *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, volume 4, pages 1057–1060. Ieee, 2005.

[19] A. A. Handzel and P. Krishnaprasad. Biomimetic sound-source localization. *IEEE Sensors Journal*, 2(6):607–616, Dec. 2002.

[20] P. Hinich, S. Hinich, and S. Milutinovich. Localization of Sound Source in 3D Space. In *Signal Processing and Information Technology, 2008. ISSPIT 2008. IEEE International Symposium on*, pages 219–222, 2008.

[21] K. Ho, X. Lu, and L. Kovavisaruch. Source localization using TDOA and FDOA measurements in the presence of receiver location errors: Analysis and solution. *Signal Processing, IEEE Transactions on*, 55(2):684–696, 2007.

[22] Y. Huang, J. Benesty, G. Elko, and R. Mersereati. Real-time passive source localization: A practical linear-correction least-squares approach. *Speech and Audio Processing, IEEE Transactions on*, 9(8):943–956, 2001.

[23] F. Keyrouz and K. Diepold. An Enhanced Binaural 3D Sound Localization Algorithm. *2006 IEEE International Symposium on Signal Processing and Information Technology*, pages 662–665, Aug. 2006.

[24] F. Keyrouz, K. Diepold, and S. Keyrouz. Humanoid Binaural Sound Tracking Using Kalman Filtering and HRTFs. *Robot Motion and Control 2007*, pages 329–340, 2007.

[25] U. Klee, T. Gehrig, and J. McDonough. Kalman Filters for Time Delay of Arrival-Based Source Localization. *EURASIP Journal on Advances in Signal Processing*, 2006:1–16, 2006.

[26] E. A. Lehmann. Matlab code for image-source model in room acoustics. http://www.eric-lehmann.com/ism_code.html.

[27] E. A. Lehmann and A. M. Johansson. Prediction of energy decay in room impulse responses simulated with an image-source model. *The Journal of the Acoustical Society of*, 124(1):269–277, 2008.

[28] G. Mellen, M. Pachter, J. Raquet, and Others. Closed-form solution for determining emitter location using time difference of arrival measurements. *Aerospace and Electronic Systems, IEEE Transactions on*, 39(3):1056–1058, 2003.

[29] P. Pertilä. Acoustic Source Localization in a Room Environment and at Moderate Distances. *Tampereen teknillinen yliopisto. Julkaisu-Tampere University of Technology. Publication; 794*, 2009.

[30] F. Ribeiro, C. Zhang, D. Florêncio, and D. Ba. Using reverberation to improve range and elevation discrimination for small array sound source localization. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(7):1781–1792, 2010.

[31] S. Sam Ge, A. Poh Loh, and F. Guan. Robust Sound Localization using Small Number of Microphones. *International Journal of Information Acquisition*, 2(1):1–22, 2005.

[32] F. Seco, A. Jiménez, C. Prieto, J. Roa, and K. Koutsou. A survey of mathematical methods for indoor localization. In *Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on*, number x, pages 9–14. IEEE, 2009.

[33] X. Sheng and Y.-h. Hu. Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks. *IEEE Transactions on Signal Processing*, 53(1):44–53, Jan. 2005.

[34] J. Smith and J. Abel. Closed-form least-squares source location estimation from range-difference measurements. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 35(12):1661–1669, 1987.

[35] H. So, Y. Chan, and F. Chan. Closed-form formulae for time-difference-of-arrival estimation. *Signal Processing, IEEE Transactions on*, 56(6):2614–2620, 2008.

[36] N. Strobel and R. Rabenstein. Classification of time delay estimates for robust speaker localization. In *Acoustics, Speech, and Signal Processing, 1999. ICASSP'99. Proceedings., 1999 IEEE International Conference on*, volume 6, pages 3081–3084. IEEE, 1999.

[37] A. Urruela and J. Riba. Novel closed-form ML position estimator for hyperbolic location. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 2, pages ii–149. IEEE, 2004.

[38] J. Valin and F. Michaud. Robust 3D localization and tracking of sound sources using beamforming and particle filtering. *Acoustics, Speech and Signal*, 2(1), 2006.

[39] B. Yang and J. Scheuing. Cramer-Rao bound and optimum sensor array for source localization from time differences of arrival. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 4, pages iv–961. IEEE, 2005.

[40] B. Yang and J. Scheuing. A theoretical analysis of 2D sensor arrays for TDOA based localization. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 4, pages IV–IV. IEEE, 2006.

[41] K. Yang, G. Wang, and Z. Luo. Efficient convex relaxation methods for robust target localization by a sensor network using time differences of arrivals. *Signal Processing, IEEE Transactions on*, 57(7):2775–2784, 2009.

[42] C. Zhang, D. Florêncio, D. E. Ba, and Z. Zhang. Maximum Likelihood Sound Source Localization and Beamforming for Directional Microphone Arrays in Distributed Meetings. *IEEE Transactions on Multimedia*, 10(3):538–548, Apr. 2008.

[43] C. Zhang, Z. Zhang, and D. Florencio. Maximum Likelihood Sound Source Localization for Multiple Directional Microphones. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07*, number 6, pages 125–128. Ieee, 2007.

Centre de recherche INRIA Grenoble – Rhône-Alpes
655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Paris – Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399