THE GEOMETRY OF SOUND-SOURCE LOCALIZATION USING NON-COPLANAR MICROPHONE ARRAYS

Xavier Alameda-Pineda[1,2], *Radu Horaud*[1] *and Bernard Mourrain*[3]

[1] INRIA Grenoble Rhône-Alpes, [2] Université de Grenoble, [3] INRIA Sophia Antipolis, France

ABSTRACT

This paper addresses the task of sound-source localization from time delay estimates using arbitrarily shaped non-coplanar microphone arrays. We fully exploit the direct path propagation model and our contribution is threefold: we provide a necessary and sufficient condition for a set of time delays to correspond to a sound source position, a proof of the uniqueness of this position, and a localization mapping to retrieve it. The time delay estimation task is casted into a non-linear multivariate optimization problem constrained by necessary and sufficient conditions on time delays. Two global optimization techniques to estimate time delays and localize the sound source are investigated. We report an extensive set of experiments and comparisons with state-of-the-art methods on simulated and real data in the presence of noise and reverberations.

Index Terms— Sound source localization, time delay estimate, constrained multivariate non-linear optimization.

1. INTRODUCTION

Source localization from time delay estimates (TDEs) has proven to be an extremely useful methodology with a variety of applications in such diverse fields as aeronautics, telecommunications and robotics. We focus on general-purpose TDE-based models for sound-source localization in indoor environments. This is extremely challenging because: (i) there may be several sound sources over time, (ii) regular rooms are echoic (leading to reverberations), and (iii) the microphones are often embedded (robot heads, smart phones, etc.), thus generating high noise.

The TDE problem has been very well investigated and a review can be found in [1]. The vast majority of existing approaches deal with one microphone pair, but it is not straightforward to extend most of these methods to more than two 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. [2] is illustrative of the first category where a method based on generalized eigenvalue decomposition is proposed. The second category is represented by [3] where a multichannel criterion based on cross-correlation is introduced to estimate time delays using a linear microphone array. A method using the geometric constraints from multiple microphone arrays is proposed in [4]. In [5], a multi-source counting and localizing algorithm is introduced using uniform circular arrays. In [6], the authors proposed a method for joint time delay estimation and source localization working on non-coplanar microphone arrays. Unfortunately, since it is based on local minimization, this method needs to be initialized on a huge grid to find the global minimum, thus increasing the computational complexity. In most of the cases, experiments are performed on speech data in a *simulated* indoor environment.

This paper has several contributions. We develop a geometric model for sound source localization from TDEs and non-coplanar microphones arrays (Section 3). This model enables the *characterization of the feasible time delays* (those corresponding to an actual source position), the *uniqueness of the source position* and the *localization mapping*, used to retrieve the source position in practice. Since the problem is cast into a non-linear constrained optimization problem, *two optimizers are proposed* (Section 4). The proposed approach is evaluated on simulated data (as is often the case in the multichannel TDE literature) as well as on *real data* (Section 5).

2. SIGNAL AND PROPAGATION MODELS

We consider a sound source placed at an unknown position $\boldsymbol{S} \in \mathbb{R}^3$. The emitted signal, $x(t)$, is received at M microphones, located at $\mathfrak{M} = \{\boldsymbol{M}_m\}_{m=1}^M \subset \mathbb{R}^3$. The signal received at the m^{th} microphone is $x_m(t) = x(t - t_m) + n_m(t)$, where n_m, the noise of the m^{th} microphone, is a zero-mean Gaussian random process, and t_m is the time-of-arrival. Assuming direct-path sound propagation at constant speed ν, we have $t_m = \|\boldsymbol{S} - \boldsymbol{M}_m\|/\nu$. The time delay between microphones m and n is:

$$t_{m,n}(\boldsymbol{S}) = t_n - t_m = \frac{\|\boldsymbol{S} - \boldsymbol{M}_n\| - \|\boldsymbol{S} - \boldsymbol{M}_m\|}{\nu}. \quad (1)$$

3. PROBLEM GEOMETRY

We recall that the task is to estimate the time delays to further localize the sound source. In this section, we provide the three main theoretical results: (i) the conditions under which a set of *time delays correspond to a sound source* (such sets will be called *feasible* sets) and (ii) *the uniqueness of the sound source position* for any feasible set and (iii) a *closed-formula for localization*, i.e., to retrieve the position of the sound source from a feasible set. Even if, in practice the problem is set in the ambient space, $\mathbb{R}^3$, the theory presented here is valid in $\mathbb{R}^N, N \geq 2$. In the following, Section 3.1 studies the case of two microphones and Section 3.2 exploits the geometry of the M microphone case.

3.1. The Case of Two Microphones

We start by formally characterizing the set of possible sound-source locations in the case of two microphones located at $\boldsymbol{M}_m$ and $\boldsymbol{M}_n$.

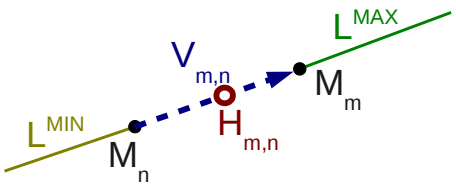

Figure 1: Geometry associated with the two microphone case, located at $\boldsymbol{M}_m$ and $\boldsymbol{M}_n$ (see Lemma 1). $H_{m,n}$ is the mid-point of the the microphones (in red) and $V_{m,n}$ the vector $\boldsymbol{M}_m - \boldsymbol{M}_n$ (in dashed-blue). $L^{\text{MAX}}_{m,n}$ and $L^{\text{MIN}}_{m,n}$ are the two half lines represented in green and yellow respectively.

For a given time delay $\hat{t}_{m,n}$, we characterize $\boldsymbol{S}$ satisfying $\hat{t}_{m,n} = t_{m,n}(\boldsymbol{S})$. Because (1) is a hyperboloid in $\mathbb{R}^N$, this equation embeds the *hyperbolic geometry* of the problem. For completeness, we state the following lemma (figure 1):

Lemma 1 *The set of sound-source locations $\boldsymbol{S} \in \mathbb{R}^N$ satisfying $t_{m,n}(\boldsymbol{S}) = \hat{t}_{m,n}$ is:*

(i). empty if $|\hat{t}_{m,n}| > t^*_{m,n}$, where $t^*_{m,n} = \|\boldsymbol{M}_m - \boldsymbol{M}_n\|/\nu$,

(ii). the half line $\boldsymbol{L}^{MAX}_{m,n}$ (or $\boldsymbol{L}^{MIN}_{m,n}$), if $\hat{t}_{m,n} = t^*_{m,n}$ (or if $\hat{t}_{m,n} = -t^*_{m,n}$), where $\boldsymbol{L}^{MAX}_{m,n} = \{\boldsymbol{H}_{m,n} + \mu \boldsymbol{V}_{m,n}\}$, $\boldsymbol{L}^{MIN}_{m,n} = \{\boldsymbol{H}_{m,n} - \mu \boldsymbol{V}_{m,n}\}$, $\mu \geq 1/2$, $\boldsymbol{H}_{m,n} = (\boldsymbol{M}_m + \boldsymbol{M}_n)/2$ and $\boldsymbol{V}_{m,n} = \boldsymbol{M}_m - \boldsymbol{M}_n$,

(iii). the hyperplane passing by $\boldsymbol{H}_{m,n}$ perpendicular to $\boldsymbol{V}_{m,n}$, if $\hat{t}_{m,n} = 0$ or

(iv). one sheet of a two-sheet hyperboloid with foci $\boldsymbol{M}_m$ and $\boldsymbol{M}_n$ for other values of $\hat{t}_{m,n}$.

Lemma 1[1] and characterizes the positions associated to one microphone pair and sets the basis for next section, where we analyze the geometry of the most general microphone setup.

3.2. The Case of M Microphones in General Position

In this section we characterize the set of possible sound-source locations in the case of M microphones. We first notice that if a set of time delays $\hat{\boldsymbol{t}} = \{\hat{t}_{m,n}\}_{m=1,n=1}^{m=M,n=M} \in \mathbb{R}^{M^2}$ satisfies (1) $\forall m, n$, then the time delays are coupled by $\hat{t}_{m,n} = -\hat{t}_{1,m} + \hat{t}_{1,n}$. Hence, we only need to consider the time delays $\boldsymbol{t} = (t_{1,2}, \ldots, t_{1,M})$ which lie in a $(M-1)$-dimensional vector subspace $\mathcal{W} \subset \mathbb{R}^{M^2}$.

Hence, there are $M-1$ equations of the form (1). Geometrically, this is equivalent to seek the intersection of $M-1$ hyperboloids in $\mathbb{R}^N$ (see figure 2). Algebraically, this is equivalent to solve a system on $M-1$ non-linear equations in N unknowns. In general, this leads to search for the roots of a high-degree polynomial. However, in our case the hyperboloids share one focus, namely $\boldsymbol{M}_1$. As it will be shown below, the problem in this case reduces to solving a second-degree polynomial plus a linear system of equations. The $M-1$ equations write:

$$\begin{cases} \nu\hat{t}_{1,2} &= \|\boldsymbol{S} - \boldsymbol{M}_2\| - \|\boldsymbol{S} - \boldsymbol{M}_1\| \\ &\vdots \\ \nu\hat{t}_{1,M} &= \|\boldsymbol{S} - \boldsymbol{M}_M\| - \|\boldsymbol{S} - \boldsymbol{M}_1\| \end{cases}. \quad (2)$$

[1]Proven here: `http://hal.inria.fr/docs/00/84/88/76/ANNEX/Alameda-WASPAA-2013-Annex_1_.pdf`

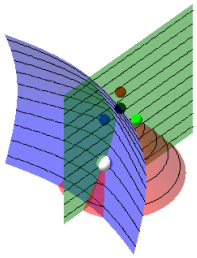

Figure 2: Localization of the source using four microphones. Their position is shown in black ($\boldsymbol{M}_1$), blue ($\boldsymbol{M}_2$), red ($\boldsymbol{M}_3$) and green ($\boldsymbol{M}_4$). The sound source is placed in the white marker. The blue hyperboloid corresponds to $\hat{t}_{1,2}$, the red to $\hat{t}_{1,3}$ and the green to $\hat{t}_{1,4}$. The intersection of the hyperboloids corresponds to the sound source position.

Because the M microphones are in general position (they do not lie in the same hyperplane), we have $M \geq N+1$, hence the number of equations is greater or equal than the number of unkowns. We now provide the conditions on $\hat{\boldsymbol{t}}$ under which (2) yields a real and unique solution for $\boldsymbol{S}$. More precisely, firstly we provide a necessary condition on $\hat{\boldsymbol{t}}$ for (2) to have real solutions, secondly we prove the uniqueness of the solution and build a mapping to recover the solution $\boldsymbol{S}$, and thirdly we provide a necessary and sufficient condition on $\hat{\boldsymbol{t}}$ for (2) to have a real and unique solution.

Notice that each equation in (2) is equivalent to $(\nu\hat{t}_{1,m} + \|\boldsymbol{S} - \boldsymbol{M}_1\|)^2 = \|\boldsymbol{S} - \boldsymbol{M}_m\|^2$, from which we obtain $-2(\boldsymbol{M}_1 - \boldsymbol{M}_m)^{\mathsf{T}}\boldsymbol{S} + p_{1,m}\|\boldsymbol{S} - \boldsymbol{M}_1\| + q_{1,m} = 0$, where $p_{1,m} = 2\nu\hat{t}_{1,m}$ and $q_{1,m} = \nu^2(\hat{t}_{1,m})^2 + \|\boldsymbol{M}_1\|^2 - \|\boldsymbol{M}_m\|^2$. Hence, (2) can now be written in matrix form:

$$\mathbf{M}\boldsymbol{S} + \boldsymbol{P}\|\boldsymbol{S} - \boldsymbol{M}_1\| + \boldsymbol{Q} = 0, \quad (3)$$

where $\mathbf{M} \in \mathbb{R}^{(M-1)\times N}$ is a matrix with its m^{th} row, $1 \leq m \leq M-1$, equal to $(\boldsymbol{M}_{m+1} - \boldsymbol{M}_1)^{\mathsf{T}}$, $\boldsymbol{P} = (p_{1,2}, \ldots, p_{1,M})^{\mathsf{T}}$ and $\boldsymbol{Q} = (q_{1,2}, \ldots, q_{1,M})^{\mathsf{T}}$. Notice that $\boldsymbol{P}$ and $\boldsymbol{Q}$ depend on $\hat{\boldsymbol{t}}$.

Without loss of generality and because the points $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_M$ do not lie in the same hyperplane, we assume that $\mathbf{M}$ can be written as a concatenation of an invertible matrix $\mathbf{M}_L \in \mathbb{R}^{N\times N}$ and a matrix $\mathbf{M}_E \in \mathbb{R}^{(M-N-1)\times N}$ such that $\mathbf{M} = \begin{pmatrix} \mathbf{M}_L \\ \mathbf{M}_E \end{pmatrix}$. Similarly $\boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_L \\ \boldsymbol{P}_E \end{pmatrix}$ and $\boldsymbol{Q} = \begin{pmatrix} \boldsymbol{Q}_L \\ \boldsymbol{Q}_E \end{pmatrix}$. Thus, (3) rewrites:

$$\mathbf{M}_L\boldsymbol{S} + \boldsymbol{P}_L\|\boldsymbol{S} - \boldsymbol{M}_1\| + \boldsymbol{Q}_L = 0, \quad (4)$$

$$\mathbf{M}_E\boldsymbol{S} + \boldsymbol{P}_E\|\boldsymbol{S} - \boldsymbol{M}_1\| + \boldsymbol{Q}_E = 0, \quad (5)$$

where $\boldsymbol{P}_L$, $\boldsymbol{Q}_L$ are vectors in $\mathbb{R}^N$ and $\boldsymbol{P}_E$, $\boldsymbol{Q}_E$ are vectors in $\mathbb{R}^{M-N-1}$. Notice that (2) is strictly equivalent to (4)-(5). In the following, (4) will be used for defining the necessary conditions on $\hat{\boldsymbol{t}}$ as well as localizing the sound source. The study of (5) is reported further on. By introducing a scalar variable w, (4) can be written as:

$$\mathbf{M}_L\boldsymbol{S} + w\boldsymbol{P}_L + \boldsymbol{Q}_L = 0, \quad (6)$$

$$\|\boldsymbol{S} - \boldsymbol{M}_1\|^2 - w^2 = 0. \quad (7)$$

We remark that the system (6)-(7) is defined in the $(\boldsymbol{S}, w)$ space. Notice that (6) represents a straight line and (7) represents quadric.

Hence the solution to (6)-(7) is the intersection of a straight line and a quadric. In such systems there are two possible configurations: (i) the quadric contains the straight line, and there are an infinite number of solutions, or (ii) the straight line crosses the quadric, and there are two (maybe complex) solutions. In fact, the first case, (i), does not occur. Notice that the quadric is a two-sheet hyperboloid. Because two-sheet hyperboloids are not ruled surfaces, (7) does not contain any straight line. Consequently the system has two (maybe complex) solutions.

In order to solve (6)-(7), we first rerwite (6) as

$$\boldsymbol{S} = \boldsymbol{A}w + \boldsymbol{B}, \tag{8}$$

where $\boldsymbol{A} = -\mathbf{M}_L^{-1}\boldsymbol{P}_L$ and $\boldsymbol{B} = -\mathbf{M}_L^{-1}\boldsymbol{Q}_L$, and then substitute $\boldsymbol{S}$ from (8) into (7) obtaining:

$$(\|\boldsymbol{A}\|^2 - 1)w^2 + 2\langle \boldsymbol{A}, \boldsymbol{B} - \boldsymbol{M}_1\rangle w + \|\boldsymbol{B} - \boldsymbol{M}_1\|^2 = 0. \tag{9}$$

We are interested in the real solutions, that is, $\boldsymbol{S} \in \mathbb{R}^N$. Because $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^N$, the solutions of (6)-(7) are real, if and only if, the solutions to (9) are real too. Equivalently, the discriminant of (9) has to be non-negative. Hence the solutions to (6)-(7) are real if and only if $\hat{\boldsymbol{t}}$ satisfies:

$$\Delta(\hat{\boldsymbol{t}}) = \langle \boldsymbol{A}, \boldsymbol{B} - \boldsymbol{M}_1\rangle^2 - \|\boldsymbol{B} - \boldsymbol{M}_1\|^2(\|\boldsymbol{A}\|^2 - 1) \geq 0. \tag{10}$$

The previous equation is a *necessary condition* for (6)-(7) to have real solutions. Albeit, we are interested in the solutions of (4). Obviously, if $\boldsymbol{S}$ is a solution of (4), then $(\boldsymbol{S}, \|\boldsymbol{S} - \boldsymbol{M}_1\|)$ is a solution of (6)-(7). However, the reciprocal is not true; these two systems are not equivalent. Indeed, since $\Delta(\hat{\boldsymbol{t}}) = \Delta(-\hat{\boldsymbol{t}})$, one of the solutions of (6)-(7) is the solution of (4) and the other is the solution of (4) replacing $\hat{\boldsymbol{t}}$ by $-\hat{\boldsymbol{t}}$. In other words, the two solutions of (6)-(7), namely $(\boldsymbol{S}^+, w^+)$ and $(\boldsymbol{S}^-, w^-)$, satisfy either:

$$\begin{cases} \boldsymbol{t}(\boldsymbol{S}^+) = \hat{\boldsymbol{t}} \\ \boldsymbol{t}(\boldsymbol{S}^-) = -\hat{\boldsymbol{t}} \end{cases} \text{ or } \begin{cases} \boldsymbol{t}(\boldsymbol{S}^+) = -\hat{\boldsymbol{t}} \\ \boldsymbol{t}(\boldsymbol{S}^-) = \hat{\boldsymbol{t}} \end{cases}$$

Consequently, the solution to (4) is *unique*. Moreover, we can use (8) to define the following *localization mapping*, which retrieves the sound-source position from a feasible $\hat{\boldsymbol{t}}$:

$$L(\hat{\boldsymbol{t}}) = \begin{cases} \boldsymbol{S}^+ = \boldsymbol{A}w^+ + \boldsymbol{B} & \text{if } \boldsymbol{t}(\boldsymbol{S}^+) = \hat{\boldsymbol{t}} \\ \boldsymbol{S}^- = \boldsymbol{A}w^- + \boldsymbol{B} & \text{otherwise.} \end{cases} \tag{11}$$

Until now we provided the condition for equation (4) to have real solutions, the uniqueness of the solution and a localization mapping. However, the original system includes also equation (5). In fact, (5) adds $M-N-1$ constraints onto $\hat{\boldsymbol{t}}$. Indeed, if $(L(\hat{\boldsymbol{t}}), \|L(\hat{\boldsymbol{t}}) - \boldsymbol{M}_1\|)$ is the solutions to (4), then in order to be a solution of (4)-(5), it has to satisfy:

$$E(\hat{\boldsymbol{t}}) = \mathbf{M}_E L(\hat{\boldsymbol{t}}) + \boldsymbol{P}_E\|L(\hat{\boldsymbol{t}}) - \boldsymbol{M}_1\| + \boldsymbol{Q}_E = 0. \tag{12}$$

Moreover, the reciprocal is true. Summarizing, the system (4)-(5) has a unique solution $L(\hat{\boldsymbol{t}})$ if and only if $\Delta(\hat{\boldsymbol{t}}) \geq 0$ and $E(\hat{\boldsymbol{t}}) = 0$.

The mappings Δ, E and L are explicitly constructed solely from the microphone locations $\mathfrak{M}$. Hence, these mappings are not only an interesting mathematical finding in its own right, but also useful from a computational perspective. In addition, the mappings Δ and E can be understood from two points of view. Geometrically, they characterize the *time delays corresponding to a sound source*. Algebraically, Δ and E represent the *feasibility* constraint to the time delay estimation problem, i.e., the time delay estimate should satisfy the necessary and sufficient conditions for the existence of $\boldsymbol{S}$. L has to be understood as the *closed-form solution for localization*, allowing to recover $\boldsymbol{S}$ from any feasible $\hat{\boldsymbol{t}}$.

4. TIME DELAY ESTIMATION

In the previous section we characterized the *feasible* values of $\boldsymbol{t}$ (i.e., those corresponding to a sound source position). But, which is the best value for $\boldsymbol{t}$, among all the feasible ones? We need a criterion to choose the optimal value for $\boldsymbol{t}$ given the M received signals. This operation is called *time delay estimation* and the result is the *time delay estimator*. The criterion we have chosen, denoted by J, was presented in [3] in the framework of linear microphone arrays and extended in [6] to the non-coplanar case. J is the determinant of the matrix of normalized cross-correlation functions. That is, $J(\boldsymbol{t}) := \det\left([\rho_{i,j}(\boldsymbol{t})]_{i,j}\right)$, where $\rho_{i,j}(\boldsymbol{t}) = \mathrm{E}\left\{x_i(t+t_{1,i})x_j(t+t_{1,j})\right\} / \sqrt{\mathrm{E}\left\{x_i^2(t)\right\}\mathrm{E}\left\{x_j^2(t)\right\}}$, E being the expectation. Notice that J increases with the dissimilarity of the signals $\{x_1(t), x_2(t+t_{1,2}), \ldots, x_M(t+t_{1,M})\}$.

Thus, the time delay estimation is casted into the following *non-linear constrained optimization* problem:

$$\begin{cases} \min\limits_{\boldsymbol{t}} J(\boldsymbol{t}), \\ \text{s.t.} \quad \boldsymbol{t} \in \mathcal{W}, \quad -\boldsymbol{t}^* \leq \boldsymbol{t} \leq \boldsymbol{t}^*, \\ \qquad \Delta(\boldsymbol{t}) \geq 0, \quad E(\boldsymbol{t}) = 0, \end{cases} \tag{13}$$

where $\mathcal{W}$, $\boldsymbol{t}^*$, Δ and E were defined in the previous section.

In order to solve this optimization problem, we investigate two distinct methods. First, if the functions $\rho_{i,j}$ are continuously differentiable, the cost function J is Lipschitz continuous in the compact set $-\boldsymbol{t}^* \leq \boldsymbol{t} \leq \boldsymbol{t}^*$, and hence a branch and bound (B&B) global optimization algorithm is appropriate. Its output is a list of points (ranked by the cost), from which we select the best among those satisfying the constraints. Second, we conjecture that the global minimum of J corresponds to local maxima of the functions $\rho_{1,m}$. Thus, for each microphone pair $(1, m)$, we extract K local maxima of $\rho_{1,m}$. We then construct a grid with all possible combinations of these values, ending up with K^{M-1} points. This point grid (which is sparser than the one used in [6]) is then used to initialize a log-barrier interior point method.

5. EXPERIMENTAL RESULTS

In order to accurately validate the proposed geometric model and the two optimization algorithms, we developed a formal evaluation protocol using simulated and real data. The setup is the same in both cases: a $4 \times 4 \times 4$ meter room with an array of four microphones at (in meters) $\boldsymbol{M}_1 = (2.0, 2.1, 1.83)^\mathsf{T}$, $\boldsymbol{M}_2 = (1.8, 2.1, 1.83)^\mathsf{T}$, $\boldsymbol{M}_3 = (1.9, 2.2, 1.97)^\mathsf{T}$ and $\boldsymbol{M}_4 = (1.9, 2.0, 1.97)^\mathsf{T}$ and the sound source at 189 different positions on a 1.7 m radius sphere around the microphones. The source emitted speech fragments randomly chosen from [7]. One hundred millisecond cuts of these sounds are the input of the evaluated methods. In the simulated case, we control two parameters. First, the SNR, regulating the amount of noise added to the received signals, and taking the following values (in dB): -10, -5 and 0. Secondly the T_{60}, used in the Image-Source Model [8] to control the amount of reverberations, taking the following values (in s): 0 (none), 0.2 (moderate), and 0.6 (severe). In the real case, we used the acquisition protocol defined in [9], replacing the dummy head by the tetrahedron microphone array. Several algorithms are compared: **D** is the method proposed in [6] (optimization solved by a log-barrier interior-point method), **I** solves independently for each microphone pair, **B** corresponds to

Table 1: Results obtained with simulated data. The first column corresponds to the values of SNR [dB]. The second column corresponds to the values of T_{60} [s]. The four last columns correspond to each of the methods. For each combination SNR-T_{60}-method there are three values: the proportion of inliers (angular error < 30 degrees), the inlier angular error mean and standard deviation.

SNR	T_{60}	**B**	**S**	**I**	**D**
0	0.0	82.1% 9.59 3.66	46.9% 11.63 5.54	53.7% 11.31 5.55	75.3% 10.54 4.57
	0.2	73.8% 12.65 6.14	40.9% 14.79 6.98	44.3% 14.60 6.92	67.5% 13.54 6.51
	0.6	35.7% 16.10 7.30	22.2% 17.67 7.13	23.6% 17.67 7.60	31.2% 16.58 7.24
-5	0.0	84.1% 10.46 4.64	39.3% 13.41 6.41	41.4% 13.24 6.11	80.4% 11.74 5.52
	0.2	68.6% 13.91 6.75	34.4% 16.60 7.21	32.7% 16.50 7.09	61.9% 14.74 6.91
	0.6	29.8% 16.97 7.35	18.7% 18.19 7.28	16.9% 18.08 7.43	28.0% 17.11 7.38
-10	0.0	77.5% 13.45 6.56	31.0% 17.13 7.36	29.6% 17.04 7.19	66.6% 14.69 6.90
	0.2	44.5% 16.53 7.36	22.1% 18.46 7.00	20.8% 18.92 7.49	38.3% 16.82 7.24
	0.6	19.0% 18.63 7.26	13.2% 19.65 7.26	12.5% 19.25 7.22	15.8% 18.61 7.20

the B&B method, and **S** corresponds to the log-barrier interior point methods initialized with the grid proposed in Section 4. All these algorithms provide a time delay estimate, $\hat{\boldsymbol{t}}$, used to retrieve the sound-source position using the localization mapping (11).

Table 1 shows the localization results obtained with simulated data. Each row consists on three subrows: the percentage of localization inliers (angular error less than 30^0), the angular error mean of inliers, and the standard deviation (in degrees). We first observe that all methods behave as expected with increasing levels of noise and reverberations. Secondly, we notice that methods **B** and **D** perform much better than **S** and **I**. Also, we remark that the global optimization procedure proposed in this paper (**B**) performs systematically better than the state-of-the-art method reported in [6] (**D**).

Table 2 presents the results on the real data. First of all, we observe that methods **D** and **B** outperform **S** and **I** as in the simulated case. Secondly, we remark that, contrary to the simulated data, **D** outperforms **B**. Third, we notice that the results on real data roughly correspond to the simulated case with $T_{60} = 0.6$ s and $SNR = -5$ dB, which is a very challenging scenario. A general remark is that, in all cases the performance notably decreases with reverberations, which is expected, since the signal model used does not explicitly handle the reverberations.

Table 2: Results obtained with real data. The rows have the same meaning as in Table 1.

B	**S**	**I**	**D**
21.98% 18.15 7.17	12.77% 19.01 6.83	13.14% 18.79 7.06	27.64% 16.16 7.58

6. CONCLUSIONS AND FUTURE WORK

In this paper we derived a geometric model for arbitrary shaped non-coplanar microphone arrays, providing a characterization of the feasible time delays and a localization mapping to recover the sound source position. The task is casted into a non-linear optimization problem constrained by the geometric model. Two algorithms are proposed to find the optimal solution and localize the sound source. Extensive experiments on both simulated and real data allow us to conclude that the the proposed model in conjunction with the B&B algorithm outperforms the state of the art, thus validating the geometric model as well as the optimization procedure.

We will extend this work in several directions. Firstly, learning the effect of the reverberations on the objective function. Secondly, by evaluating the model in the framework of dynamic sound sources. Thirdly, adapting the methodology into a calibration task, where the position of the sound source may be known, but not the microphones' position. Finally, performing experiments using a large number of microphones and evaluating the influence of their positions.

7. REFERENCES

[1] J. Chen, J. Benesty, and Y. A. Huang, "Time Delay Estimation in Room Acoustic Environments," *EURASIP*, 2006.

[2] S. Doclo and M. Moonen, "Robust Adaptive Time Delay Estimation for Speaker Localization in Noisy and Reverberant Acoustic Environments," *EURASIP*, 2003.

[3] J. Chen, J. Benesty, and Y. Huang, "Robust time delay estimation exploiting redundancy among multiple microphones," *IEEE Tran. on SAP*, 2003.

[4] A. Canclini, E. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Tran. on ASLP*, 2013.

[5] D. Pavlidi, A. Griffin, M. Puigt, and A. Mouchtaris, "Real-time multiple sound source localization and counting using a circular microphone array," *IEEE Tran. on ASLP*, 2013.

[6] X. Alameda-Pineda and R. P. Horaud, "Geometrically-constrained robust time delay estimation using non-coplanar microphone arrays," in *Proceedings of EUSIPCO*, 2012.

[7] 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, LDC.

[8] E. A. Lehmann and A. M. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *JASA*, 2008.

[9] A. Deleforge and R. P. Horaud, "The cocktail party robot: Sound source separation and localisation with an active binaural head," in *IEEE/ACM HRI*, 2012.