Self-calibration of microphone arrays from measurement of Times Of Arrival of acoustic signals

A.Contini[†], A. Canclini[†], F.Antonacci[†], M.Compagnoni[§], A.Sarti[†], S.Tubaro[†] [*†]

Abstract

In this paper we present a methodology for the self-calibration of microphone arrays using Times Of Arrival of acoustic signals produced by acoustic sources. Geometric constraints are derived from the measurements of the Times Of Arrival. Microphone locations are then found by combining the geometric constraints. When multiple arrays are present in the acoustic scene, in many applications the knowledge of their mutual positions is required. If a priori information about the geometry of the array is given, the same framework can be easily adapted to infer the mutual positions of the arrays. Some experimental results and simulations prove the accuracy of the current work.

1. Introduction

The use of microphone arrays is widespread whenever we need to acquire the space-time structure of acoustic wavefronts: interesting applications of microphone arrays are localization; tracking; and separation of acoustic sources. Moreover, in the last years, applications that rely on the use of multiple arrays are appearing.

In some cases, e.g. in acoustic source localization and tracking, the location and pose of the arrays with respect to the environment and the position of the microphones within each array need to be available or they need to be somehow estimated. This problem is referred in the literature as self-calibration. Calibration is a fairly common operation in 3D vision, as it is a preliminary step to any application that uses multiple cameras to extract 3D geometric information from the imaged scene [1]. Only in recent times the multichannel audio processing community started investigating the problem of self-calibration of microphone arrays. We can categorize the existing algorithms of self-calibration in two classes: we refer to intra-calibration and inter-calibration techniques.

As far as intra-calibration is concerned, the goal is to accurately localize sensors within the same array and it is required when space-time processing algortihms cannot rely on approximate measurements of the locations of the sensors. In [2] the authors approach the problem of source localization with non-calibrated microphone arrays with a two-step iterative algorithm: at alternate steps source and microphone locations are localized using the knowledge acquired at the previous step. In [3] the authors tackle the intra-calibration problem using the Multi-Dimensional Scaling (MDS) [4] framework. More specifically, given the tape measures of the distances between all possibile pair of sensors, the locations of microphones are estimated with respect to a reference one. In [5] the authors use the same approach but the measurement of the distances between sensors is accomplished through the analysis of the complex coherence function between microphones when diffuse noise field is present in the environment. As a result, the calibration procedure turns out to be greatly simplified with respect to [3].

Inter-calibration is required when multiple arrays are present in the acoustic scene and the knowledge of their mutual positions is needed. Some solutions in this direction have recently appeared in the literature. In [6] the authors address the inter-calibration problem by jointly localizing with two microphone arrays an acoustic source. The mutual positions are found with tools borrowed by computer vision that jointly analyze the localization results for several locations of the acoustic source.

In [7] the authors combine intra and inter-calibrations using a hierarchical approach: first each array is "intra-calibrated" using diffuse-noise field as input signal, then arrays are "inter-calibrated" with an approach similar to [6]. We notice, however, that intra-

[*†] Dipartimento di Elettronica ed Informazione Politecnico di Milano § Dipartimento di Matematica F.Brioschi canclini@elet.polimi.it

[†]The work presented in this paper was partially realized under the funding of the SCENIC project. The SCENIC project acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number: 226007

calibration and inter-calibration need two different procedures, as different input signals are used in the two phases.

In this paper we develop a novel solution for the intra-calibration that additionally may be used to inter-calibrate the arrays. The input measurements are the Times Of Arrival (TOAs) of signals emitted by an acoustic source located at multiple positions. Let us consider, for a while, the presence of a single microphone in the environment. Given the source location and the measurement of the Distance of Flight (inferred from TOA), the microphone is bound to lie on a circle centered on the source and with radius equal to the Distance of Flight. This constraint takes a convenient and elegant form when we use homogeneous - scalable - coordinates to represent points. The location of the microphone is obtained as the “least squares intersection” of circles for several locations of the acoustic source. Positions of all the microphones in the array are obtained by replicating the above algorithm for all sensors. The only requirement in the measurement procedure is that mutual positions of sources are known. This can be easily obtained by bounding the source to lie on positions on a “calibration pattern”, as it is commonly done in computer vision. An example of calibration pattern is in Section 4.

When a-priori information about the internal structure of the array is given or it is obtained in a previous step, this knowledge can be easily embedded into the geometric constraints. The unknown, in this case, turns out to be the rotation matrix and the translation vector that bring the array from the local reference frame to the world one. Using the same theoretical framework, therefore, we are able to shift the target of the self-calibration procedure from intra-array to the inter-array calibration. The rotation matrix and the translation vector, in turn, can be used to find the actual positions of the microphones.

The use of tools borrowed from computer vision for audio processing is not completely novel in the literature. As an example, in [8] the author reconstructs the acoustic scene (the position of sources and microphones) analyzing the problem using tools typically used in computer vision. In [9] the authors formulate the problem of inferring the geometry of the environment from acoustic measurements using the projective geometry in a similar way to what is done in this paper.

The rest of the paper is structured as follows: in Section 2 we derive the geometric constraints about the microphone locations from the knowledge of the source location and of the Time Of Arrival. In Section 3 we describe the self-calibration procedure of the arrays for both “intra” and “inter” cases. Section describes the hardware and software implementations and presents some experimental results with both real and simulated signals to confirm the validity of the approach. Finally, Section 5 draws some conclusions.

2. Geometric Constraints from Times Of Arrival

For notational simplicity, we will consider here two-dimensional geometries. However, the algorithm can be easily adapted to 3D as the mathematical tools we will use throughout the rest of the paper have a three-dimensional counterpart. Let us consider the presence of a loudspeaker placed at $\boldsymbol{x}_n = [x_n, y_n]^T$ and of a microphone (synchronized with the loudspeaker) at the unknown location $\boldsymbol{x}_m = [x_m, y_m]^T$. The impulse response from $\boldsymbol{x}_n$ to $\boldsymbol{x}_m$ is measured with a suitable technique (e.g. through Maximum Length Sequences). We assume that source and microphone are mutually visible, therefore the impulse response always contains the direct path among them. We denote the TOA of the direct path from $\boldsymbol{x}_n$ to $\boldsymbol{x}_m$ with the symbol τ_{nm}. With this information at hand, $\boldsymbol{x}_m$ is bound to lie on a circle centered in $\boldsymbol{x}_n$ and with radius $\tau_{nm}s$, where s is the speed of sound (343 m/s at standard conditions). The equation of a generic conic passing through (x_m, y_m) is

$$a x_m^2 + b x_m y_m + c y_m^2 + d x_m + e y_m + f = 0 \ . \qquad (1)$$

Given the measurement of the Distance Of Flight $\tau_{nm}s$ and the knowledge of $\boldsymbol{x}_n$, we can specialize the vector parameter $[a, b, c, d, e, f]^T$ to our problem by replacing the equation of a circle centered in $\boldsymbol{x}_n$ and with radius $\tau_{nm}s$ into (1). After some passages we obtain

$$\begin{cases} a = & 1 \\ b = & 0 \\ c = & 1 \\ d = & -2x_n \\ e = & -2y_n \\ f = & x_n^2 + y_n^2 - (\tau_{nm}s)^2 \ . \end{cases} \qquad (2)$$

Equation (1) is easily expressed using homogeneous - scalable - coordinates. In particular, the homogeneous representation of the point $\boldsymbol{x}_m = [x_m, y_m]^T$ is $\mathbf{x}_m = [kx_m, ky_m, k]^T$, where $k \in \mathbb{R}$ and $k \neq 0$. With this notation, the circle equation in (1) becomes

$$\mathbf{x}_m^T \mathbf{C}_{nm} \mathbf{x}_m = 0 \ , \qquad (3)$$

where

$$\mathbf{C}_{nm} = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix} \qquad (4)$$

is the conic matrix for the circle centered in $\boldsymbol{x}_n$ and with radius $\tau_{nm}s$. Eq. (3) describes the constraint on the microphone position in an elegant and convenient way. In the next Section we will exploit the constraint on the microphone position in eq. (3) for the problems of intra and inter-calibration.

3. Self-calibration from Geometric Constraints

3.1. Intra-calibration

Let us consider that Times Of Arrival $\tau_{nm}, n = 1\ldots,N$ are measured for multiple source locations $\boldsymbol{x}_n, n = 1,\ldots,N$. From a geometric point of view, find the microphone position $\boldsymbol{x}_m$ corresponds to find the intersection of the circles associated to the conic matrices $\mathbf{C}_{nm}, n = 1,\ldots,N$, which means finding the solution of the system of equations

$$\begin{cases} \mathbf{x}_m^T\mathbf{C}_{1m}\mathbf{x}_m &= 0 \\ \mathbf{x}_m^T\mathbf{C}_{2m}\mathbf{x}_m &= 0 \\ \quad\ldots & \\ \mathbf{x}_m^T\mathbf{C}_{nm}\mathbf{x}_m &= 0 \end{cases} \tag{5}$$

In real scenarios, however, time sampling and measurement errors tend to affect the solution of (5). We estimate $\mathbf{x}_m$, therehore, as the global minimum of a cost function obtained by combining the individual constraints in (5). We define the cost function

$$J(\mathbf{x}_m) = \sum_{n=1}^{N} (\mathbf{x}_m^T\mathbf{C}_n\mathbf{x}_m)^2 \ , \tag{6}$$

and we search for the point $\hat{\mathbf{x}}_m$ that minimizes $J(\mathbf{x}_m)$:

$$\hat{\mathbf{x}}_m = \arg\min_{\mathbf{x}_m} J(\mathbf{x}_m) \ . \tag{7}$$

We now concentrate on the solution the minimization problem. We first expand the expression of $J(\mathbf{x}_m)$ inserting into eq.(6) the coefficients of the matrices relative to the conic matrices $\mathbf{C}_n$. This way, we obtain a fourth order polynomial, whose coefficients are obtained from the elements of $\mathbf{C}_n$ and whose unknown variable is $\mathbf{x}_m$. We notice, however, that equations in (5) are homogeneous, therefore if $\hat{\mathbf{x}}$ is solution, then $k\hat{\mathbf{x}}, k \neq 0$ and $k \in \mathbb{R}$ are solutions as well. In order to find an unique solution, we search for the solution of (5) on a subspace of the projective space. In order to do so, we define a reduced cost function setting the third coordinate to $k = 1$. This operation is known in the literature as “de-homegenization” and in our case it gives

$$J_3(\mathbf{x}_m) = J(\mathbf{x}_m)|_{k_m=1} \ . \tag{8}$$

We now look for the stationary points of $J_3(\mathbf{x}_m)$ by imposing

$$\begin{cases} \frac{\partial J_3(\mathbf{x}_m)}{\partial x_m} &= 0 \\ & \\ \frac{\partial J_3(\mathbf{x}_m)}{\partial y_m} &= 0 \end{cases} \ . \tag{9}$$

We notice that the system in eq.(9) is composed of two 3-rd order equations whose unknowns are x_m and y_m. We solve the two equation through a symbolic solver. This yields to a set of 9 stationary points, which can be complex. The searched solution is therefore the global real minimum of the cost function $J_3(\mathbf{x}_m)$.

In order intra-calibrate the array, the minimization procedure in eq.(9) is replicated for all the microphones.

3.2. Global array calibration

When some information about the geometry of the array is given, we can easily embed this knowledge into the constraints. More specifically, we assume that the mutual positions of the microphones in the array are known. Let us denote with $\boldsymbol{x}'_m$ the nominal position of the m-th microphone (i.e. its position in the local reference system), while $\boldsymbol{x}_m$ is its position in the world coordinate system. We intend to estimate the rotation matrix and the translation vector that relate the world and the local coordinate systems. The homogeneous points $\mathbf{x}'_m$ and $\mathbf{x}_m$ corresponding to $\boldsymbol{x}'_m$ and $\boldsymbol{x}_m$, respectively, are related by an isometry $\mathbf{H}$ that contains the rotation matrix and the translation vector that bring $\mathbf{x}'_m$ to $\mathbf{x}_m$:

$$\mathbf{x}_m = \mathbf{H}\mathbf{x}'_m \ , \tag{10}$$

We recall that an isometry is given by

$$\mathbf{H} = \begin{bmatrix} \mathbf{R}(\theta) & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \ , \tag{11}$$

where $\mathbf{R}(\theta)$ is the matrix operating a rotation of the reference frame by an angle θ and $\mathbf{t}$ is the translation vector.

The problem has now become the estimation of the isometry that brings the local coordinate system to the global one, i.e. find the rotation matrix and the translation vector that move $\mathbf{x}'_m$ to $\mathbf{x}_m$. Replacing eq.(10) into eq.(3) leads to a new form of the constraint:

$$\mathbf{x}'^T_m\mathbf{H}^T\mathbf{C}_{mn}\mathbf{H}\mathbf{x}'_m = 0 \ . \tag{12}$$

If the nominal positions of the microphones in the array have been measured, the only unknown in eq.(12) is the isometry matrix $\mathbf{H}$. As neither $\mathbf{C}_{mn}$ nor $\mathbf{x}'_m$ are

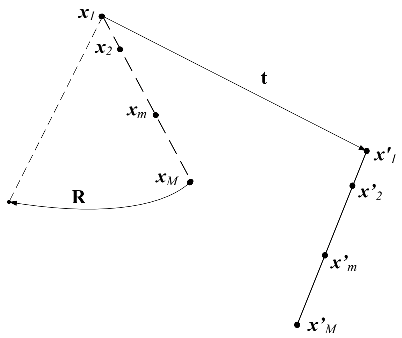

Figure 1. **Self-calibration problem when the geometry of the array is known in advance: our goal is to estimate the rotation matrix R and the translation vector t that bring the array from the nominal position to the actual one.**

unknowns in (12), we can put equations for all microphones in the same array in a single equation system:

$$\begin{cases} \mathbf{x}_1'^T\mathbf{H}^T\mathbf{C}_{11}\mathbf{H}\mathbf{x}_1' = 0 \\ \mathbf{x}_1'^T\mathbf{H}^T\mathbf{C}_{12}\mathbf{H}\mathbf{x}_1' = 0 \\ \cdots \\ \mathbf{x}_M'^T\mathbf{H}^T\mathbf{C}_{M_1N}\mathbf{H}\mathbf{x}_M' = 0 \end{cases}, \qquad (13)$$

The cost function for the global array calibration is obtained as the combination of all constraints

$$J_{gl}(\theta,\mathbf{t}) = \sum_{m=1}^{M_1}\sum_{n=1}^{N}(\mathbf{x}_m'^T\mathbf{H}^T\mathbf{C}_{mn}\mathbf{H}\mathbf{x}_m')^2 , \qquad (14)$$

and the solution is given by

$$(\hat{\theta},\hat{\mathbf{t}}) = \arg\min_{\theta,\mathbf{t}} J_{gl}(\theta,\mathbf{t}) . \qquad (15)$$

As far as the minimization of the cost function in (14) is concerned, we cannot use in this case a closed-form solution. At the same time, we notice that the cost function has a strongly nonlinear dependence on the rotation angle and the translation vector and therefore it presents multiple local minima. In order to guarantee that the iterative minimizer is not trapped in a local minimum, we need to initialize the minimum search algorithm. We do so by using the results from the intra-calibration. In particular, the rotation of the array with respect to the nominal position is estimated as the rotation of the line that interpolates the microphones within the same array, while the initial translation vector is estimated as the distance of the interpolating line from the origin. Provided a good initial point, many possible iterative mimimum search algorithms can be used: we use here nonlinear least-squares.

Finally, the position of the m-th microphone array is given by

$$\hat{\mathbf{x}}_m = \hat{\mathbf{H}}\mathbf{x}_m' ,$$

where $\hat{\mathbf{H}}$ is the homography estimated according to the minimization in eq.(15). If multiple arrays are present in the acoustic scene, the above equation is replicated for the two arrays, therefore we end up with the estimation of two rotation angles and two translation vectors. In this case, the two arrays can be inter-calibrated (according to the definition of inter-calibration given in [6]) finding the difference of the rotation angles and translation vectors.

3.3. Discussion

The equation systems in (13) and (5) are expected to present complementary results. The advantage of the cost function in (15) is that measurements related to all the microphones within the same array are collected in the same equation system to estimate the rotation matrix and the translation vector. As a consequence, an improved effectiveness is expected with respect to the cost function in (15) when a few measurements are available. On the other hand, when uncertainties about the geometry of the arrays (due, as an example, to an unperfect construction of it), the global calibration algorithm in (15) can suffer from a relevant error due to its nonlinearity. Experimental results confirm this complementarity.

4. Implementation and evaluation

4.1. Acquisition device

An implementation of the above algotithms has been developed in Matlab. The block diagram of the self-calibration procedure is depicted in Figure 2. A

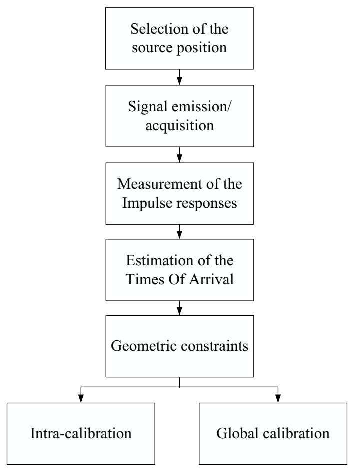

Figure 2. **Block diagram of the self-calibration system**

loudspeaker is positioned in any point on a 25×5 rectangular grid whose size is $0.9m \times 0.5m$. The origin of the grid is also the origin of the global reference frame, whose axes are parallel to those of the calibration pattern. A snapshot of the acquisition system (calibration pattern and microphone arrays) is shown in Figure 3. The loudspeaker produces a sequence of white noise

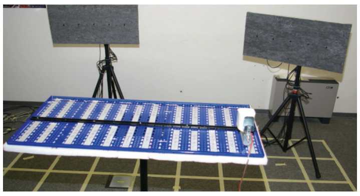

Figure 3. Snapshot of the acquisition system: the loudspeaker is placed on any position on the calibration pattern (in the foreground). The gray panels close to the walls, on the background, are the two arrays (each accommodating five sensors).

in the band $[0 - 22kHz]$ and it is synchronized with the microphones. The impulse response from the loudspeaker to each microphone in the arrays is measured through a cross-correlation. The Times Of Arrival τ_{mn} of the direct path are estimated by picking the first relevant maximum in the impulse responses. Then τ_{mn} is converted into the Distance Of Flight (DOF) D_{mn} with the knowledge (previously acquired) of the latency L of the soundcard (i.e. the component of the delay between sound emission and acquisition independent of the propagation) and of the soundspeed s as follows:

$$D_{mn} = \tau_{mn} s + L F_s \; , \qquad (16)$$

where F_s is the sampling frequency. The distance of flight is finally converted into the geometric constraint as described in equation (12).

4.2. Simulations

Before testing the system in a real scenario we tested the algorithm under simulative conditions. Microphones have been placed on two linear arrays, each accommodating five sensors distant $0.15m$. The positions of the arrays (in terms of rotation matrix and translation vector with respect to the calibration pattern) are given in Table 1. The acquisition of the signal has

Table 1. Displacement of the arrays with respect to the reference frame

Array #	Rotation angle	Translation vector
1	$-\pi/6$	$[-1.8, 0.5]^T$
2	$\pi/6$	$[1.8, 0.5]^T$

been simulated for up to $N = 19$ positions of the loudspeaker and with a sampling frequency $F_s = 44,1kHz$. In order to make the acqusitions as likely as possible, we have spatialized the acquired signals using the Fast Beam Tracing algorithm, described in [10]. For each position the Distances Of Flight have been corrupted with additive white noise whose standard deviation error is 0.01 m. For each position, 30 realizations have been simulated. The effectiveness of the intra and inter-calibration algorithms has been evaluated using the following metric:

$$E = \sqrt{\frac{1}{MK} \sum_{k=1}^{K} \sum_{m=1}^{M} (\hat{\mathbf{x}}_{m,k} - \mathbf{x}_m)^2} \; ,$$

where M is the number of microphones in both arrays, $K = 30$ is the number of repetitions of the experiment and $\hat{\mathbf{x}}_{mk}$ is the estimation of the position of the m-th microphone in the k-th simulation. In order to verify the increase of robustness of the algorithms to a variable number of constraints, the error has been computed for N ranging from 3 to 19. Table Table 2 confirms that

Table 2. Average root mean square error of localization of microphones for a variable number of TOA measurements

TOA meas.	Intra-cal. RMSE [m]	Global cal. RMSE [m]
3	0.0522	0.0331
4	0.0215	0.0187
5	0.0208	0.0194
8	0.0126	0.0086
12	0.0109	0.0077
16	0.0096	0.0063
19	0.0091	0.0056

when a few measurements are available (from three to five) the global calibration process is much more efficient than the intra-calibration. When more information is available, the global calibration presents only a little advantage over the intra-calibration. It is important to notice that in these simulations the Time Of Arrival of the direct path has been corrupted but the positions of the microphones in the two arrays have not been altered from their ideal position in the world coordinate system.

4.3. Experiments

We tested the self-calibration algorithms in real scenarios in order to verify the robustness of the algorithm to deviations from ideality in the construction of the microphone array. We tested the algorithm in two rooms: the first one is a room whose walls have been paneled with acoustic absorbing material. As a result, the reverberation time is $0.05s$. The second room is a typical office-room, whose reverberation time $0.7s$. The arrays in these experiments are composed of five microphones disposed on a line. The distance between adjacent sensors is variable between $0.155m$ and $0.148m$ but the nominal positions $\mathbf{x}'_m$ in equation (10) envision an inter-distance of $0.15m$. The mutual positions of the arrays correspond to those in the simulative scenario (Table 1). Figure 4 shows the Root Mean Square Errors of intra and inter-calibration processes as a function of the number of observations, ranging from 3 to 19, in the first room. We notice that the intra-calibration over-

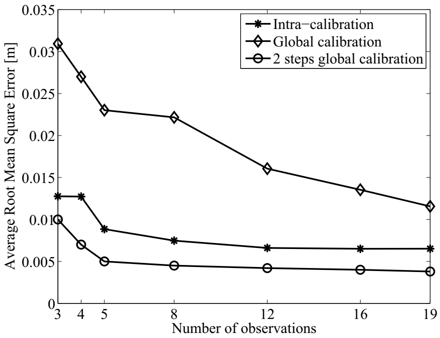

Figure 4. **Error of intra and inter-calibration as a function of the number of measurement of the Times Of Arrival. Experiments have been conducted in a semi-anechoic room.**

comes the inter-calibration algorithm. We interpret the different behavior with respect to the simulations as the consequence of the uncertainties of the exact positioning of the microphones within the array. In fact, the cost function in (14) make a strong hypothesis on the linear configuration of the arrays, which is not verified in the real case. Figure 5 shows the calibration results for the reverberant room. We notice that no main degrada-

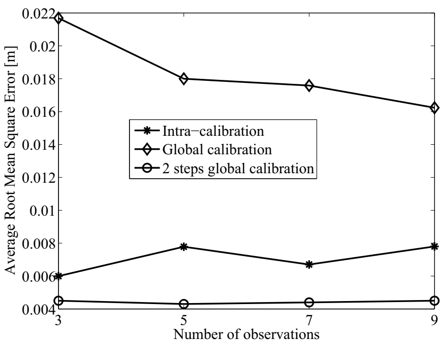

Figure 5. **Error of intra and inter-calibration as a function of the number of measurement of the Times Of Arrival. Experiments have been conducted in a reverberant room.**

tion of the performances is present when the algorithm is tested in a reverberant room. This is due to the fact that reverberations do not affect the measurement of the Time Of Arrival of the direct path.

5. Conclusions

In this paper we have shown a novel methodology for the self-calibration of microphone arrays that relies on the measurement from multiple positions of the Time Of Arrival of a synchronized source signal. The algorithm can easily embed a-priori information about the internal configuration of the array. Experimental results confirm the validity of the approach and highlight that the global self-calibration (which embeds the geometry of the array) is much sensitive to the exactness of the a-priori information.

References

[1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.

[2] A. Weiss and B. Friedlander, “Array shape calibration using sources in unknown locations - a maximum likelihood approach,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 1958–1966, Dec. 1989.

[3] S. T. Birchfield and A. Subramanya, “Microphone array position calibration by basis-point classical multidimensional scaling,” *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1025–1034, Sep. 2005.

[4] T. Cox and M. Cox, *Multidimensional Scaling*. Chapman and Hall, 2001.

[5] I. McCowan, M. Lincoln, and I. Himawan, “Microphone array shape calibration in diffuse noise fields,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 3, pp. 666–670, Mar. 2008.

[6] A. Redondi, M. Tagliasacchi, F. Antonacci, and A. Sarti, “Geometric calibration of distributed microphone arrays,” in *proc. of IEEE International Workshop on Multimedia Signal Processing*, Oct. 2009, pp. 1–5.

[7] M. Hennecke, T. Plotz, G. Fink, J. Schmalenstroer, and R. Hab-Umbach, “A hierarchical approach to unsupervised shape calibration of microphone array networks,” in *IEEE/SP 15th Workshop on Statistical Signal Processing, SSP ’09.*, Sept. 2009, pp. 257–260.

[8] S.Thrun, “Affine structure from sound,” in *Proc. of Conference on Neural Information Processing Systems (NIPS)*, 2005.

[9] F. Antonacci, A. Sarti, and S. Tubaro, “Geometric reconstruction of the environment from its response to multiple acoustic emissions,” in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010.

[10] F. Antonacci, M. Foco, A. Sarti, and S. Tubaro, “Fast tracing of acoustic beams and paths through visibility lookup,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 4, pp. 812–824, May 2008.