EXPERIMENTAL VALIDATION OF TOA-BASED METHODS FOR MICROPHONES ARRAY POSITIONS CALIBRATION

T.-K. Le and N. Ono

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda-ku
101-8430 Tokyo, Japan
kien@nii.ac.jp

T. Nowakowski, L. Daudet, and J. De Rosny

ESPCI and CNRS, Paris Diderot Univ.
Institut Langevin, 1 rue Jussieu
75005 Paris, France
thibault.nowakowski@espci.fr

ABSTRACT

This study is an experimental validation of a new closed-form method for automatic array position calibration, based on time of arrival (TOA) measurements between sources and sensors. An experiment with a large array composed of 121 microphones and a dozen of sources has been set up. We first show that, when considering the whole array, this calibration method gives results on par with a reference state-of-the-art acoustic method. We then show experimentally that the new method provides significantly better results when the number of sources and microphones decreases, confirming numerical simulations. We conclude the paper with a discussion on methodological issues for array position calibration.

***Index Terms*—** position calibration methods, TOA-based array calibration, acoustic arrays

1. INTRODUCTION AND PRIOR WORK

A large number of microphone array techniques, such as source localization, noise reduction, source separation or acoustic wavefield analysis and synthesis, assume that the position of each sensor (microphone) is perfectly known. More generally, such sensor network techniques are of growing importance, with applications in navigation systems, aerospace, geophysics, and seismology, to name only a few domains.

To increase the performances of such systems, a new tendency is to increase the number of sensors. Arrays can now contain hundreds or even thousands of elements. Another current trend is the use of irregular arrays, due to physical constraints (e.g. ad-hoc arrays), or also to new signal processing methods such as sparse methods. The use of such very large arrays raises a number of challenges, amongst which the knowledge of the array geometry. Indeed, many of the above-mentioned applications are very sensitive to the microphone positions, and therefore a very accurate positioning is required - a fraction of the smallest wavelength, in the audible range, typically less than 1 cm. In such cases, manual measurement of positions is difficult to perform, and self-calibration methods are required, using acoustic informations to estimate the array geometry. Most of these methods rely on the measurement of the times-of-arrival (TOA) set between controlled sources, with known or unknown positions, and each microphone. From these measurements, multidimensional scaling (MDS) -based methods provide the relative positions of each source and microphone.

Recently, some closed-form solutions have been developed, and provide an unprecedented ease of use. When the positions of microphones are known and when TOA measurements are available, several closed-form methods have been proposed for estimating source positions, such as spherical interpolation [1, 2], hyperbolic intersection [3, 4], or linear intersection [5]. These methods show that with known microphones positions, source localization is not too complex in the sense that the constraints used to estimate their positions are simple. Moreover, these methods present a symmetry in the roles of sources and microphones. Inverting the emitters and receivers allows localization of microphones positions from known sources positions and TOA measurements. More general TOA-based calibration methods aim at jointly localizing unknown sources and microphones positions on the basis of TOA measurements [6, 7, 8, 9]. The main idea of these methods is to solve the least-squares problem :

$$(\hat{\mathbf{m}}_i, \hat{\mathbf{s}}_j) = \arg\min_{\mathbf{m}_i, \mathbf{s}_j} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\|\mathbf{m}_i - \mathbf{s}_j\| - d_{i,j} \right)^2 \qquad (1)$$

where $\mathbf{m}$ and $\mathbf{s}$ are respectively the unknown positions of the M microphones and the N sources, and $d_{i,j}$ is the distance between microphone i and source j, obtained by multiplying the measured TOA with the sound velocity. Because this cost function possesses a lot of local minima, the optimization problem is difficult to solve with standard or iterative [6, 10] optimization techniques.

Hence, designing closed-form solutions to this problem is an active field of research. Amongst recent contributions we can cite Crocco's closed-form (C-CF) [6] and Le's closed-form solutions (L-CF) [11]. One difference between these two methods is that the (C-CF) approach requires that the position of one microphone is coincident with one of the sources, contrarily to the (L-CF), that replaces this assumption by the estimation of one supplementary parameter. However, besides numerical tests, no experimental validation had been performed yet for the (L-CF) and its comparison with other state-of-the-art methods like (C-CF) was still to achieve.

This work proposes an experimental validation and comparison of both methods. Some experimental difficulties such as knowing the ground-truth to compare the different results are discussed in a first section, together with our evaluation framework. A large experiment with 121 microphones and 12 sources irregularly spaced is described in a second part and allows us to compare our method (L-CF) with the state-of-the-art method (C-CF), when varying the number of sources and microphones of the array.

2. EVALUATION FRAMEWORK

2.1. Principle

One experimental issue in position self-calibration methods for microphone arrays, is that they require a precise measurement of the

TOA set. The distance obtained from the experimental TOA measurement between microphone i and source j is always corrupted by some error η_{ij}, for example arising from uncertainties in the sound velocity of the medium, when the source is not sufficiently wideband (resulting in some uncertainties in the exact time of the first signal arrival), when microphones are close to obstacles like walls, or when ambient noise corrupts the measurements.

Another experimental difficulty is to obtain the ground-truth positions to compare the different methods, especially for randomly distributed arrays. One evaluation framework can be used, based on a few assumptions. Assuming that the measurements errors corrupting the TOA have bounded amplitude, when the number of sources and microphones gets very large, there is sufficient information to precisely determine the position of N sources and M microphones. In this case the number $N \times M$ of TOA measurements grows faster than the number of unknowns ($3(N + M)$), averaging out the measurement errors. Under this assumption, any suitable method for the TOA-based calibration problem, such as the (C-CF) and (L-CF), gives positions that asymptotically approach the ground-truth, when the number of sources and microphones tends to infinity, as will be discussed in the next section. This estimation, called reference positions, can be used to estimate and replace the role of the unknown ground-truth. Once this reference is estimated, the (C-CF) and (L-CF) are evaluated again by computing the differences between the reference positions and the new estimations obtained when using a subset of the TOA measurements (i.e. by decreasing the number of sources and microphones).

Two arguments based on previous work can be formulated to support this methodology. Firstly, numerical studies have shown that the (C-CF) gives good calibration results for small errors on the initial distance estimation between sources and receivers [6]. Further synthetic experiments showed that the estimated positions given by (C-CF) and (L-CF) are very similar for large number of sensors and sources [12]. Secondly, experimental validations based on the (C-CF) [9, 13] and its use to experimentally validate several acoustic problems such as narrowband source localization [14], have already validated this method experimentally. Thereby, estimating the positions with the two methods will validate the relevance of the (L-CF) approach, if its results are close to the ones obtained with the (C-CF) solution.

2.2. Comparison method

In this section, we define the metric used to compare the different methods and give some numerical results to check their validity with large arrays. Thereafter, the word "*estimator*" has the meaning of the word "*method*".

Definition 1. *An estimator f is called a **TOA-based calibration estimator** if for any*

(i) $\mathcal{S} \equiv \{\mathbf{s}_1, \dots, \mathbf{s}_N\} \subset \mathbb{R}^3$ a group of N unknown sources, $\mathcal{M} \equiv \{\mathbf{m}_1, \dots, \mathbf{m}_M\} \subset \mathbb{R}^3$ a group of M unknown microphones, and

(ii) $\mathbf{D} = (d_{ij})_{N\times M}$ an observed distance matrix of $\mathcal{S}$ and $\mathcal{M}$, where $d_{ij} = \|\mathbf{s}_i - \mathbf{m}_j\|_2 + \eta_{ij}$, the Euclidean distance between $\mathbf{s}_i$ and $\mathbf{m}_j$ adding a measurement error η_{ij},

then f gives estimations of positions $\mathbf{s}_i, \mathbf{m}_j$ on the basis of $\mathbf{D}$.

Note that, since the observed distance matrix is invariant under reflection, translation, and rotation, without loss of generality, we can assume $\mathbf{s}_1 \equiv (0, 0, 0)^T$, $\mathbf{s}_2 \equiv (0, 0, \alpha)^T$, and $\mathbf{s}_3 \equiv (0, \beta, \gamma)^T$,

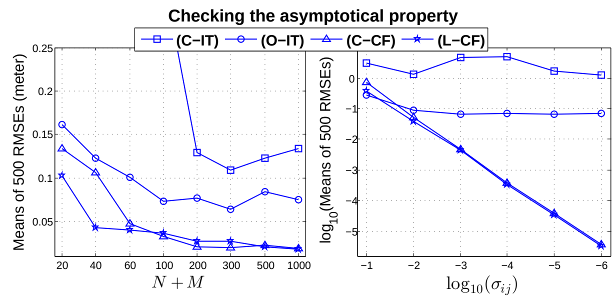

Fig. 1. Means of RMSEs for 500 independent experiments, for (*left*) fixed $\sigma_{ij} = 10^{-2}$, and (*right*) fixed $N = 10, M = 25$.

where $\alpha, \beta \geqslant 0$. The estimations by f are also based on this assumption. We denote $\mathbf{s}_i^f$ and $\mathbf{m}_j^f$ the estimations of $\mathbf{s}_i$ and $\mathbf{m}_j$ by f, respectively. The *root mean square error* (RMSE) of this estimation is defined by:

$$E_f(\mathbf{D}) = \left[\frac{1}{N+M}\left(\sum_{i=1}^{N}\|\mathbf{s}_i^f - \mathbf{s}_i\|_2^2 + \sum_{j=1}^{M}\|\mathbf{m}_j^f - \mathbf{m}_j\|_2^2\right)\right]^{\frac{1}{2}}. \tag{2}$$

Numerical experiments are first achieved with simulated data to check the asymptotical properties of the calibration methods.

Definition 2. *A TOA-based calibration estimator f is **asymptotical** if for all $\mathcal{S}, \mathcal{M}$ and $\mathbf{D}$ for which $\forall i, j$, η_{ij} is a Gaussian noise with zero mean and σ_{ij} standard deviation, and independent with other noises $\eta_{i'j'}$, then*

(i) given σ_{ij} for all i, j, $E_f(\mathbf{D})$ is decreasing when N and M are increasing, and

(ii) given N and M, $\lim\limits_{\forall i,j:\sigma_{ij}\to 0} E_f(\mathbf{D}) = 0$.

On figure 1, two parameters are studied to compare four TOA-based calibration methods: the two closed-forms (C-CF), and (L-CF), and two iterative methods (O-IT) [10], and (C-IT) [6]. In the left part of the figure, the standard deviation is fixed, $\sigma_{ij} = 10^{-2}$ for all m, n. For each value of $N + M$, 500 independent numerical experiments are executed, in which N is chosen randomly and uniformly from 7 to $(N + M) - 7$. The N sources and M microphones are simulated as uniformly distributed and independent points in a virtual box of size $1m \times 1m \times 1m$. The measurement errors η_{ij} are i.i.d. Gaussian noises with zero mean and σ_{ij} standard deviation. RMSEs are computed for each experiments, and their *mean* is represented. In the right part of figure 1, we fix the number of sources and microphones $N = 10$, $M = 25$, and consider a variable $\sigma_{ij} = 10^{-k}$, where $k = 1, 2, \dots, 6$.

The results in Fig. 1 infer that (O-IT) and (C-IT) are not asymptotical. One reason could come from the difficulty of initializing the iterative algorithms. Although these results are not sufficient to guarantee the conditions in Def. 2, they can confirm that (C-CF) and (L-CF) are asymptotical.

If the TOA-based calibration estimator f is asymptotical, when N, M are large and the measurement errors η_{ij} are small (possible in real applications), the estimations $\mathbf{s}_1^f, \dots, \mathbf{s}_N^f$ and $\mathbf{m}_1^f, \dots, \mathbf{m}_M^f$ are very close to the ground-truth $\mathbf{s}_1, \dots, \mathbf{s}_N$ and $\mathbf{m}_1, \dots, \mathbf{m}_M$. The estimated positions $\mathbf{s}_1^f, \dots, \mathbf{s}_N^f$ and $\mathbf{m}_1^f, \dots, \mathbf{m}_M^f$ are used as *reference positions*. We can then evaluate f based on the RMSEs between the reference positions and new estimations obtained for sub-matrix of $\mathbf{D}$, i.e. using information of subsets of sources and microphones.

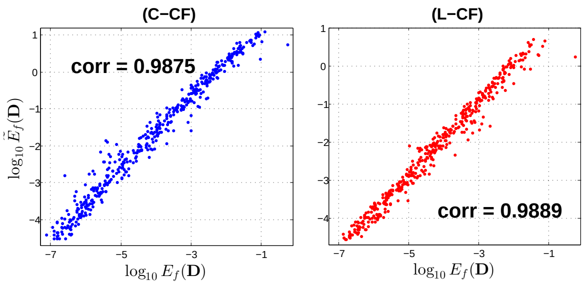

Fig. 2. Correlations between $E_f(\mathbf{D})$ and $\widetilde{E}_f(\mathbf{D})$, where f are (C-CF) and (L-CF).

Let $\mathcal{I} = \{i_1, \ldots, i_U\} \subset \{1, \ldots, N\}$ and $\mathcal{J} = \{j_1, \ldots, j_V\} \subset \{1, \ldots, M\}$ be the sub-indexes of the indexes of sources and microphones. Let $\mathbf{D}_{\mathcal{I},\mathcal{J}} = (d_{i_u j_v})_{U\times V}$ be a new observed distance matrix when keeping only the measured distances for sources and microphones in the new indexes $\mathcal{I}, \mathcal{J}$. And let $\mathbf{s}_{i_u}^{f,\mathcal{I},\mathcal{J}}$ and $\mathbf{m}_{j_v}^{f,\mathcal{I},\mathcal{J}}$ are estimations of $\mathbf{s}_{i_u}$ and $\mathbf{m}_{j_v}$ by f and $\mathbf{D}_{\mathcal{I},\mathcal{J}}$. Regarding the reference positions $\mathbf{s}_{i_u}^f$ and $\mathbf{s}_{j_v}^f$ we compare the *empirical local root mean square error* (EL-RMSE), defined by

$$\widetilde{E}_f(\mathbf{D}_{\mathcal{I},\mathcal{J}}) = \left[\frac{1}{U+V}\left(\sum_{u=1}^{U}\|\mathbf{s}_{i_u}^{f,\mathcal{I},\mathcal{J}} - \mathbf{s}_{i_u}^f\|_2^2 + \sum_{v=1}^{V}\|\mathbf{m}_{j_v}^{f,\mathcal{I},\mathcal{J}} - \mathbf{m}_{j_v}^f\|_2^2\right)\right]^{\frac{1}{2}}, \tag{3}$$

and the *empirical global root mean square error* (EG-RMSE), defined by

$$\widetilde{E}_f(\mathbf{D}|U,V) = \sum_{|\mathcal{I}|=U, |\mathcal{J}|=V} \widetilde{E}_f(\mathbf{D}_{\mathcal{I},\mathcal{J}}). \tag{4}$$

We propose to evaluate the estimator f as follows:

Definition 3. *Given two asymptotical TOA-based calibration estimators f and g. f is called* ***more stable than*** *g if*

(i) for all $\mathcal{S}$, $\mathcal{M}$, and their observed distance matrix $\mathbf{D}$ in which the noises η_{ij} have small amplitudes for all i, j,

(ii) for all $U < N$, $V < M$ such that U, V satisfy the conditions of f and g about the numbers of sources and microphones.

then $\widetilde{E}_f(\mathbf{D}|U,V) \leqslant \widetilde{E}_g(\mathbf{D}|U,V)$.

Def. 3 says that the asymptotical estimator f is stable if its estimations are good not only for large numbers but also for small numbers of sources and microphones. In a general way, for any $\mathcal{S}$, $\mathcal{M}$, and $\mathbf{D}$, the smaller RMSEs computed by the asymptotical estimator f are, the smaller EG-RMSEs computed by f are. This is confirmed by the following study on 500 independent and synthetic experiments. For each experiment, (i) N and M are chosen randomly from 30 to 80, (ii) N sources and M microphones are simulated as uniformly distributed and independent points in the virtual box with size $1 \times 1 \times 1$ meter, and (iii) η_{ij} are i.i.d. random variables with zero mean and σ standard deviation, i.e. Gaussian variables with zero mean and σ standard deviation or continuous uniform variables on the interval $[-\sqrt{3}\sigma, \sqrt{3}\sigma]$, where $\sigma = 10^{-k}$ and k is chosen uniformly on the interval $[1, 7]$. The RMSE $E_f(\mathbf{D})$ and the EG-RMSEs $\widetilde{E}_f(\mathbf{D}|U,V)$ for $U, V = 7, \ldots, 12$ are computed by (C-CF) and (L-CF). We are paying attention to the correlation between

$$E_f(\mathbf{D}) \quad \text{and} \quad \widetilde{E}_f(\mathbf{D}) := \sum_{U=7}^{12}\sum_{V=7}^{12} \widetilde{E}_f(\mathbf{D}|U,V)$$

Fig. 2 gives the plots of $\left(\log_{10} E_f(\mathbf{D}), \log_{10} \widetilde{E}_f(\mathbf{D})\right)$ and their correlations. These correlations infer that if the noises are i.i.d. Gaussian variables or i.i.d. continuous uniform variables, the EG-RMSEs strongly correlate with RMSEs for the asymptotical estimators, i.e. (C-CF) and (L-CF). Thus, in real experiments without a complete knowing of the ground-truth, we can use the EG-RMSEs to evaluate the estimations of sources and microphones positions by the asymptotical TOA-based estimators.

3. EXPERIMENTS

3.1. Experimental setup

To evaluate and compare the performances of the two asymptotical methods (C-CF) and (L-CF), an experiment is set up inside a room of dimensions 8.15 m $\times$ 3.9 m $\times$ 3.35 m, with a reverberation time $T_{60} \simeq 0.7$ s. An array composed of 120 microphones is placed in the room. The microphones are randomly distributed on the edges of a cubic structure of side 2 m (see left part of Fig. 3). A dozen of microphones is placed inside the cube. In addition, 12 sources (PC loudspeakers) are distributed around the array. To find the positions of the microphones and the sources using (C-CF), a 121^{st} microphone must be coincident with one of the loudspeakers. This constraint determines the origin of the array.

Each loudspeaker emits a chirp signal between 100 Hz and 15000 Hz recorded by the microphones at a sampling frequency $f_e = 32000$ Hz. By using the method of pulse compression [15], we are able to measure the impulse response (IR) between the sources and each microphone with a very good signal-to-noise ratio. Measuring the time of arrival of the first peak of each IR, we obtain a matrix of size (121×12) containing all the TOAs between the sources and the microphones. These TOAs are finally converted into distances by multiplying them with the measured speed of sound in the room $c_0 \approx 345$ m.s^{-1}. The resulting matrix is the observed distance matrix $\mathbf{D}$ described in Def. 1.

3.2. Experimental results and discussion

The microphones and sources positions estimated by (C-CF) and (L-CF) from the experimental distance matrix $\mathbf{D}$ are represented in the right part of Fig. 3. As expected, because the number of emitters and receivers is large, the estimations by (C-CF) and (L-CF) give equivalent results. The RMSE between the two methods is 4.4 cm. This small difference is partially due to some experimental errors on the TOA measurement that can be caused by uncertainties on the acoustic centers of microphones (for example the width of the electret microphone is about 1/2 " $\approx$ 13 mm) and loudspeakers (these centers even change with frequency), and to small uncertainties in the recorded signals due to the sampling frequency, finite bandwidth of the emitter, and remaining noise.

However, because M and N are high, it is possible to consider that the estimations of the positions by (C-CF) and (L-CF) reach their asymptotical properties (see Fig. 1). A third validation is done by using multi-view image-based 3D reconstruction. 14 pictures of the array are taken from different angles of view and the 2D microphones positions are manually located on each picture. From these 2D sets of positions, and knowing the camera parameters, it is possible to geometrically reconstruct the 3D coordinates and compare them with the acoustical methods. The RMSEs between the positions obtained from this optical method and (L-CF) or (C-CF) are respectively 4 cm and 5 cm. On the left part of Fig. 3, the obtained positions from the optical method and the (L-CF) are projected on

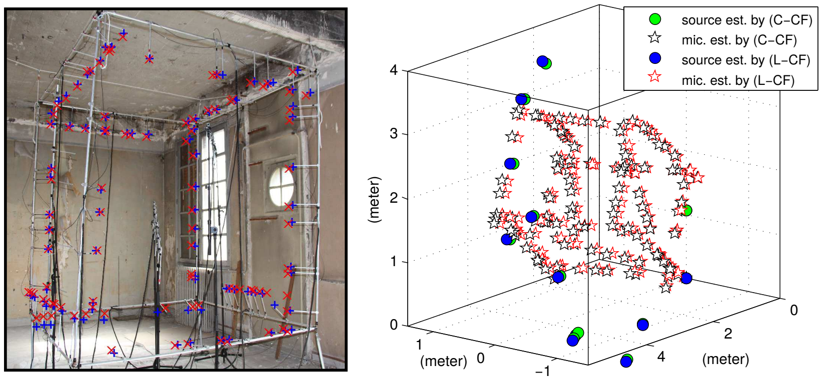

Fig. 3. (Left) Blue crosses: estimated positions from the multi-view optical reconstruction method. Red crosses: estimated positions from (L-CF). (Right) Estimations of sources and microphones positions by (C-CF) and (L-CF).

the array picture. For the sake of visibility, not all positions are represented. As seen, the positions obtained from the image reconstruction method are not perfect, as some uncertainties on the camera parameters and 2D localization on pictures remain. However this comparison shows that (C-CF) and (L-CF) estimate the positions with a good, and comparable, accuracy.

From these results, it is possible to experimentally compare the performance of the two methods when varying the number of sources and loudspeaker. Indeed, in a lot of acoustic or signal processing applications, the number of microphones is much lower. The behavior of the calibration methods when array are composed of only a few sensors is then important, especially as in some acoustic signal processing applications, errors on the microphones positions can have important repercussions. This is for example the case when performing source localization in reverberant or diffusive environments. Generally, the tolerated errors on the positions are linked with the smallest half-wavelength of the studied acoustic phenomenon. In practice, this often means that the errors on the microphone positions must be of the order of a centimeter. Based on the previous numerical study, we select successively sub-matrices of $\mathbf{D}$ by varying the numbers V and U of microphones and sources. Comparing the new experimental estimated positions with the ones obtained when considering the $121 + 12$ elements (eq. (3)), we are able to determine which method is more stable.

Figure 4 represents values of the EG-RMSEs, $\widetilde{E}_f(\mathbf{D}|U,V)$, computed by (C-CF) and (L-CF) for $U = 4,\ldots,12$ and $V = 9,\ldots,40$. For each value U and V, 300 independent realizations for the EL-RMSEs are chosen randomly, and their means are used to approximate the EG-RMSEs. The results $\widetilde{E}_f(\mathbf{D}|U,V)$ for the (L-CF) are smaller than the one's of (C-CF) for all couples $U,\ V$. This experimental validation shows that the (L-CF) is more stable than (C-CF), because it gives better estimated positions when decreasing the number of sources and microphones.

4. CONCLUSION

Two closed-form self-calibration methods for microphones localization have been experimentally compared. We have shown that their asymptotical properties guarantee a good estimation of positions for small errors on TOA measurements and large numbers of elements. This numerical study confirmed previous work and experimental validations based on the (C-CF) method. Designing an experiment with a large array composed of 121 microphones and 12 sources, we have validated that the new (L-CF) method also provides accurate localization, with experimental results similar to the (C-CF) solution. When the number of sources and microphones are decreased, the (L-CF) method appears more accurate than (C-CF), with larger difference as the number of sources and microphones get smaller.

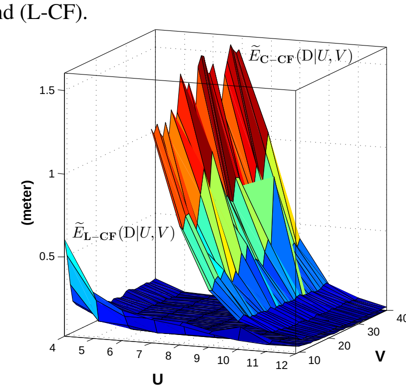

Fig. 4. The empirical global RMSEs of (L-CF) and (C-CF) for the real experiment. Since the values of $\widetilde{E}_{\text{C-CF}}(\mathbf{D}|U,V)$ are very large when $U = 4, 5$, and 6, we ignore these values in the figure. The minimum value of $\widetilde{E}_{\text{C-CF}}(\mathbf{D}|U,V) - \widetilde{E}_{\text{L-CF}}(\mathbf{D}|U,V)$ is 0.0184 (> 0) at $U = 12, V = 36$.

Many open questions still remain. First, this study has to be extended to more estimation methods, and more array geometries. Then, it would be important to test this method on large-scale arrays for which we can obtain ground-truth reference positions for the microphones, for instance by optical means (3-D reconstruction from multiple poses) or remote position sensors. Finally, all the methods discussed here rely on TOA estimations from experimentally recorded signals - a pre-processing step that might not always provide optimal results (at the sample or sub-sample accuracy) for narrowband sources, or in noisy / reverberant environments. Future work could focus on calibration methods that include explicit uncertainties on the TOA, or working directly on cross-correlation signals.

5. ACKNOWLEDGEMENTS

This work is partially supported by DGA and LABEX WIFI under references ANR-10-LABX-24 and ANR-10-IDEX-0001-02-PSL*, and also a Grant-in-Aid for Scientific Research (B) (Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 25280069). The authors would like to thank the the acoustic team MPIA from Institut d'Alembert (UPMC, Paris), for granting us access to their experimental equipment.

6. REFERENCES

[1] H. Schau and A. Robinson, "Passive source localization employing intersecting spherical surfaces from time-of-arrival differences," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-35, no. 8, pp. 1223–1225, Aug. 1987.

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

[3] Y. Chan and K. Ho, "A simple and efficient estimator for hyperbolic location," *IEEE Trans. Signal Processing*, vol. 42, no. 8, pp. 1905–1915, Aug. 1994.

[4] L. Yang and K. Ho, "An approximately efficient tdoa localization algorithm in closed-form for locating multiple disjoint sources with erroneous sensor positions," *IEEE Trans. Signal Processing*, vol. 57, no. 12, pp. 4598–4615, Dec. 2009.

[5] M. Brandstein, J. Adcock, and H. Silverman, "A closed-form location estimator for use with room environment microphone arrays," *IEEE Trans. Speech Audio Processing*, vol. 5, no. 1, pp. 45–50, Jan. 1997.

[6] M. Crocco, A. Bue, and V. Murino, "A bilinear approach to the position self-calibration of multiple sensors," *IEEE Trans. Signal Processing*, vol. 60, no. 2, pp. 660–673, Feb. 2012.

[7] V. Raykar, I. Kozintsev, and R. Lienhart, "Position calibration of microphones and loudspeakers distributed computing platform," *IEEE Trans. Speech Audio Processing*, vol. 13, no. 1, pp. 70–83, Jan. 2005.

[8] M. Crocco, A. Bue, M. Bustreo, and V. Murino, "A closed form solution to the microphone position self-calibration problem," in *IEEE ICASSP*, 2012, pp. 2597–2600.

[9] Y. Kuang, S. Burgess, A. Torstensson, and K. Åström, "A complete characterization and solution to the microphone position self-calibration problem," in *IEEE ICASSP*, 2013, pp. 3875–3879.

[10] N. Ono, H. Kohno, N. Ito, and S. Sagayama, "Blind alignment of asynchronously recorded signals for distributed microphone array," in *IEEE WASPAA*, 2009, pp. 161–164.

[11] T.-K. Le and N. Ono, "Numerical formulae for TOA-based microphone and source localization," in *IEEE IWAENC*, 2014, pp. 179–183.

[12] ——, "Reference-distance estimation approach for TDOA-based source and sensor localization," in *IEEE ICASSP*, 2015, pp. 2549–2553.

[13] N. Gaubitch, W. Kleijn, and R. Heusdens, "Auto-localization in ad-hoc microphone arrays," in *IEEE ICASSP*, 2013, pp. 106–110.

[14] G. Chardon, T. Nowakowski, J. D. Rosny, and L. Daudet, "A blind dereverberation method for narrowband source localization," *IEEE J. Sel. Topics Signal Process.*, vol. 9, no. 5, pp. 815–824, Aug. 2015.

[15] A. Hein, *Processing of SAR Data: Fundamentals, Signal Processing, Interferometry*. Springer, 2004.