Maximum likelihood estimators and Cramér-Rao bounds for the localization of an acoustical source with asynchronous arrays

Gilles Chardon

▶ **To cite this version:**

Gilles Chardon. Maximum likelihood estimators and Cramér-Rao bounds for the localization of an acoustical source with asynchronous arrays. Journal of Sound and Vibration, 2023, pp.117906. 10.1016/j.jsv.2023.117906 . hal-04149108

HAL Id: hal-04149108
https://hal.science/hal-04149108

Submitted on 3 Jul 2023

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.

Maximum likelihood estimators and Cramér-Rao bounds for the localization of an acoustical source with asynchronous arrays

Gilles Chardon

Université Paris-Saclay, CNRS, CentraleSupélec, Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France

Abstract

The performances of source localization using a microphone array can be improved by repeating the experiment with different array placements, assuming that the source position remains constant. In this article, two types of theoretical results on this setting are presented. The Maximum Likelihood Estimator (MLE) is derived, and Cramér-Rao bounds are computed, both for a strict model, where the power of the source is constant, and a relaxed model where the power of the source is allowed to change. Cramér-Rao bounds show that the performances (in terms of mean squared error of the estimation of the position and power of the source) of asynchronous arrays for the estimation of the position are, in some settings, significantly degraded compared to synchronous arrays. A particular example of such a setting is the case of two parallel arrays around the source. In contrast, the performances of power estimation are, in most cases, close to the performances of synchronous arrays. The obtained MLEs are compared to the state of the art in asynchronous array source localization using simulations, showing that the MLE for the strict model outperforms the state of the art. Experimental results illustrate the theoretical and numerical findings.

Keywords: source localization, maximum likelihood, beamforming, asynchronous arrays

1. Introduction

Beamforming is a popular method for source localization, owing to its simplicity of use [1, 2]. In the case of the localization of a unique Gaussian source and Gaussian noise, it has been shown that, with appropriate choices of steering vectors, beamforming is the Maximum Likelihood Estimator (MLE) for the parameters of the source (position and power) [3]. As a consequence, beamforming is an efficient method, in the sense that, when the duration of the measurement increases, estimations of the power and position are unbiased, and their variances reach the Cramér-Rao bounds (CRBs) [4].

Nevertheless, performances of beamforming are limited by the layout of the array. In particular, for planar arrays, estimation of the perpendicular distance from the array to the source is in general inaccurate [5]. Larger arrays yield more accurate results, but are costly and complex to implement.

As an alternative, asynchronous measurements can be used, where the same experiment is repeated, changing the position of the microphone array between each measurement. Under the assumption that the sources have not moved between the experiments, the asynchronous data can be combined to improve the estimation of the parameters of the source. However, in this case, the spatial covariance matrix of the measurements (SCM), or cross-spectral matrix, is not completely known, as coefficients related to the covariance between microphones from different arrays cannot be estimated.

Several methods have been proposed to process asynchronous acoustical data. A first methodology is to process the asynchronous data independently, e.g. by computing beamforming maps for each array, and then combine them in a global beamforming map [5–9]. Another approach is to first estimate a complete SCM, as if the data were collected synchronously, and compute a beamforming map from this completed SCM [10–14].

The goal of this paper is to provide theoretical results on source localization with asynchronous measurements, in the case of a unique source. This limited scope enables to obtain easily interpretable and implementable results. A similar approach was recently used to elucidate the question of steering vectors normalization in beamforming [3]. The MLE for the estimation of the parameters of the source (position and power) is derived, yielding an original source localization method for asynchronous arrays. CRBs, lower bounds on the variance of unbiased estimators, are computed. The CRBs can help to compare the performances of synchronous arrays and asynchronous arrays. The MLE and CRBs are also given for a relaxed model, where the power of the source is allowed to vary between measurements. MLE, in this case, is shown to be similar to combining asynchronous beamforming maps by arithmetic averaging. The proposed methods are compared to the state of the art for varied array configurations, number of snapshots, source positions and frequencies using simulations. The results are also illustrated using experimental meausurements.

The paper is structured as follows. Section 2 introduces the model and recalls the state of the art. The MLE for the strict and relaxed models is derived in section 3, and section 4 provides

Email address: gilles.chardon@centralesupelec.fr (Gilles Chardon)

the CRBs for the two models. Simulation results, with comparison to the state of the art, are given in section 5. Illustrative experimental results are given in section 6. Section 7 concludes the paper. The code necessary to reproduce the presented simulation results is available online [15].

1.1. Some notations

An estimator of a parameter θ is denoted as $\hat{\theta}$. The Mean Squared Error (MSE) for an estimator $\hat{p}$ of the power p is $\mathrm{E}\left(|p-\hat{p}|^2\right)$, where $\mathrm{E}(\cdot)$ is the mathematical expectation. The MSE for an estimator $\hat{\mathbf{x}}$ of the position $\mathbf{x}$ is $\mathrm{E}\left(\|\mathbf{x}-\hat{\mathbf{x}}\|^2\right)$. Hermitian conjugation is denoted by $\cdot^H$, the trace of a matrix by $\mathrm{tr}\,\cdot$. Superscripts $\cdot^r$, $\cdot^s$, $\cdot^a$, $\cdot^g$, $\cdot^c$ and $\cdot^m$ indicate a type of estimator or model. The probability density of the circular complex normal multivariate distribution $C\mathcal{N}(0,\Sigma)$ in dimension N is [16]

$$f(\mathbf{z}) = \frac{1}{\pi^N \det\Sigma} \exp\left(-\mathbf{z}^H\Sigma^{-1}\mathbf{z}\right). \tag{1}$$

2. Model and state of the art

Acoustical data are collected on J arrays $A_1,\ldots,A_j,\ldots,A_J$ at a given frequency f. For each array, S_j snapshots are measured. With a_{js} the amplitude of the source at the s-th snapshot for the j-th array, $\mathbf{g}_j(\mathbf{x})$ the vector modeling the propagation between the source point $\mathbf{x}$ and the microphones of the array A_j, the vector $\mathbf{m}_{js}$ of the measured complex amplitudes is

$$\mathbf{m}_{js} = a_{js}\mathbf{g}_j(\mathbf{x}) + \mathbf{n}_{js} \tag{2}$$

where $\mathbf{n}_{js}$ is a noise term, assumed to be temporally and spatially white. In the particular case of free-field propagation, the vector $\mathbf{g}_j(\mathbf{x})$ is given by its coefficients

$$g_{jm}(\mathbf{x}) = \frac{\exp\left(-\mathrm{i}\kappa\|\mathbf{x}-\mathbf{y}_{jm}\|_2\right)}{\|\mathbf{x}-\mathbf{y}_{jm}\|_2}, \tag{3}$$

where κ is the wavenumber, and $\mathbf{y}_{jm}$ is the position of the m-th microphone of the j-th array. This expression will be used for the numerical experiments, but the theoretical results hold for general propagation models.

Under the assumption that the amplitudes a_{js} are Gaussian, independent and identically distributed (i.i.d), of zero mean and variance p (the power of the source), and the noises $\mathbf{n}_{js}$ are Gaussian i.i.d. with covariance matrix $\sigma^2\mathbf{I}$, the measurements $\mathbf{m}_{js}$ are independent and distributed according to

$$\mathbf{m}_{js} \sim C\mathcal{N}(0,\Sigma_{j,p,\mathbf{x}}) \tag{4}$$

with

$$\Sigma_{j,p,\mathbf{x}} = p\mathbf{g}_j(\mathbf{x})\mathbf{g}_j(\mathbf{x})^H + \sigma^2\mathbf{I}. \tag{5}$$

Beamforming maps are computed from the estimated covariance matrix of the measurements. Here, for each array A_j, the estimated covariance matrix $\hat{\Sigma}_j$ is obtained by

$$\hat{\Sigma}_j = \frac{1}{S_j}\sum_{s=1}^{S_j}\mathbf{m}_{js}\mathbf{m}_{js}^H \tag{6}$$

We note that we do not make any assumption of the number of snapshots S_j. In other words the duration of each asynchronous measurement do not need to be equal.

2.1. Beamforming fusion

In beamforming fusion methods [6, 9], beamforming maps $B_j(\mathbf{x})$ are first computed for each array. Then, the beamforming maps are combined according to the following possible rules :

- arithmetic mean [7]: $B^a(\mathbf{x}) = \frac{1}{J}\sum_{j=1}^{J} B_j(\mathbf{x})$
- geometric mean [8]: $B^g(\mathbf{x}) = \sqrt[J]{\prod_{j=1}^{J} B_j(\mathbf{x})}$
- minimum [5]: $B^m(\mathbf{x}) = \min_{j=1\ldots J} B_j(\mathbf{x})$

The beamforming criterion with formulation IV of steering vectors (following the nomenclature introduced in [17]) is used to estimate the position, i.e.

$$B_j(\mathbf{x}) = \frac{\mathbf{g}_j(\mathbf{x})^H\hat{\Sigma}_j\mathbf{g}_j(\mathbf{x})}{\|\mathbf{g}_j(\mathbf{x})\|_2^2}, \tag{7}$$

as its maximum is known to be an unbiased estimate of the position, with smaller MSE than formulation I [3]. Then formulation III is used to estimate the power of the source for each asynchronous array, with

$$\hat{p}_j = \frac{1}{\|\mathbf{g}_j(\hat{\mathbf{x}})\|_2^2}\left(\frac{\mathbf{g}_j(\hat{\mathbf{x}})^H\hat{\Sigma}_j\mathbf{g}_j(\hat{\mathbf{x}})}{\|\mathbf{g}_j(\hat{\mathbf{x}})\|_2^2} - \sigma^2\right). \tag{8}$$

The power is then estimated by averaging, or taking the minimum of, the estimates $\hat{p}_j$, according to the chosen rule.

2.2. Matrix completion

In matrix completion methods, a complete SCM is estimated. In [10], a Bayesian method based on the expansion of the distribution of the sources to be estimated is proposed. In [11, 12, 18], methods based on rank minimization and spatial continuity of the acoustical field are proposed. In [14], the continuity condition is dropped, and a simpler algorithm is introduced. After completion of the matrix, beamforming can be applied as if synchronous measurements have been obtained. Here, we will make no assumption on the relative positions of the arrays, and spatial continuity cannot be exploited. Only the latter method will be considered here. Moreover, methods based on spatial continuity of the measurements necessitates the setting of several parameters (dimension of a subspace, regularization parameters, etc.). Likewise, the method proposed in [10], based on the expansion of the distribution of sources in a basis, cannot be applied for the case of a unique source.

3. Maximum likelihood estimators for two asynchronous source models

In this section, we derive the MLE for source localization with asynchronous arrays. Two models will be considered: the *strict model*, where the power of the source is constant between the measurements, and the *relaxed model*, where the power is allow to vary. In both models, the position of the source is assumed to be constant.

3.1. Strict model

As the measurements $\mathbf{m}_{js}$ are independent, the joint probability density of the measurements $\mathbf{M} = ((\mathbf{m}_{js})_{s=1\dots S_j})_{j=1\dots J}$ is the product

$$f_{\mathbf{x},p}(\mathbf{M}) = \prod_{j=1}^{J}\prod_{s=1}^{S_j} \frac{1}{\pi^{N_j}\det(\Sigma_{j,p,\mathbf{x}})} \exp\left(-\mathbf{m}_{js}^H \Sigma_{j,p,\mathbf{x}}^{-1}\mathbf{m}_{js}\right) \quad (9)$$

Following [3] Appendix C.2, the log-likelihood $L(\mathbf{x}, p)$ is given by

$$L(\mathbf{x}, p) = \log f_{\mathbf{x},p}(\mathbf{M}) \quad (10)$$

$$= -\sum_{j=1}^{J} S_j \operatorname{tr}(\Sigma_{j,p,\mathbf{x}}^{-1}\hat{\Sigma}_j) - S_j \log\det(\Sigma_{j,p,\mathbf{x}}) - N_j S_j \log\pi \quad (11)$$

and the MLE $(\hat{\mathbf{x}}^s, \hat{p}^s)$ for the position and power is

$$(\hat{\mathbf{x}}^s, \hat{p}^s) = \underset{\mathbf{x}\in\Omega, p\in\mathbf{R}_+}{\operatorname{argmin}} \sum_{j=1}^{J} S_j \left(-\frac{p\mathbf{g}_j(\mathbf{x})^H\hat{\Sigma}_j\mathbf{g}_j(\mathbf{x})}{\sigma^2(\sigma^2+p\|\mathbf{g}_j(\mathbf{x})\|^2)} + \log(\sigma^2+p\|\mathbf{g}_j(\mathbf{x})\|^2)\right) \quad (12)$$

where Ω is the spatial domain where the source is assumed to be located.

In the standard case of a unique array ($J = 1$), the optimization problem can be solved for fixed position $\mathbf{x}$, allowing to estimate the position by maximizing the beamforming criterion

$$B(\mathbf{x}) = \frac{\mathbf{g}(\mathbf{x})^H\hat{\Sigma}\mathbf{g}(\mathbf{x})}{\|\mathbf{g}(\mathbf{x})\|^2\sigma^2} - \log\frac{\mathbf{g}(\mathbf{x})^H\hat{\Sigma}\mathbf{g}(\mathbf{x})}{\|\mathbf{g}(\mathbf{x})\|^2\sigma^2} \quad (13)$$

which depends on the position only. The log term is usually removed, which does not change the location of the maximum of $B(\mathbf{x})$ [3]. Such a simple form cannot be obtained for the asynchronous array case, and the problem (12) has to be solved for position and power jointly.

3.2. Relaxed model

A simpler estimator is obtained by relaxing the model, allowing the powers to differ between the measurements. In that case, the MLE for the position $\mathbf{x}$ and the powers $\mathbf{p} = (p_1, \dots, p_J)$ is given by

$$(\hat{\mathbf{x}}^r, \hat{\mathbf{p}}^r) = \underset{\mathbf{x}\in\Omega, \mathbf{p}\in\mathbf{R}_+^J}{\operatorname{argmin}} \sum_{j=1}^{J} S_j \left(-\frac{p_j\mathbf{g}_j(\mathbf{x})^H\hat{\Sigma}_j\mathbf{g}_j(\mathbf{x})}{\sigma^2(\sigma^2+p_j\|\mathbf{g}_j(\mathbf{x})\|^2)} + \log(\sigma^2+p_j\|\mathbf{g}_j(\mathbf{x})\|^2)\right) \quad (14)$$

Here, the problem can be solved for the powers p_j at fixed position, as in the unique array case, and the beamforming criterion for the position is

$$B(\mathbf{x}) = \sum_{j=1}^{J} S_j \left(\frac{\mathbf{g}_j(\mathbf{x})^H\hat{\Sigma}_j\mathbf{g}_j(\mathbf{x})}{\|\mathbf{g}_j(\mathbf{x})\|^2\sigma^2} - \log\frac{\mathbf{g}_j(\mathbf{x})^H\hat{\Sigma}_j\mathbf{g}_j(\mathbf{x})}{\|\mathbf{g}_j(\mathbf{x})\|^2\sigma^2}\right) \quad (15)$$

We recognize the sum of the beamforming criteria for each array. Here, the log term cannot be removed without, slightly, changing the location of the maximum. Nevertheless, neglecting the log terms and with constant number of snapshots, the beamforming fusion by arithmetic averaging can be interpreted as an approximation of the MLE for the relaxed model. As such, it is expected to be not as accurate as the MLE for the strict model, which take advantage of the additional prior information that the power of the source is constant. We note that the beamforming criteria are weighted by the number of snapshots S_j for each array. The estimated powers $\hat{p}_j$ are

$$\hat{p}_j = \frac{1}{\|\mathbf{g}_j(\hat{\mathbf{x}}^r)\|_2^2}\left(\frac{\mathbf{g}_j(\hat{\mathbf{x}}^r)^H\hat{\Sigma}_j\mathbf{g}_j(\hat{\mathbf{x}}^r)}{\|\mathbf{g}_j(\hat{\mathbf{x}}^r)\|_2^2} - \sigma^2\right), \quad (16)$$

and a natural estimator for the power of the source is to average the estimated powers $\hat{p}_j$, using the same weights as in Eq. (15):

$$\hat{p}^r = \frac{\sum_{j=1}^{J} S_j\hat{p}_j}{\sum_{j=1}^{J} S_j}. \quad (17)$$

In addition to its simplicity compared to the strict model, the relaxed model can also take into account sources that have non-isotropic directivity patterns, assuming that the directivity diagram remains approximately constant when restricted to an array.

4. Cramér-Rao bounds

Cramér-Rao bounds are lower bounds to the variance of unbiased estimators. Moreover, the performances of the MLE reach the CRBs at increasing number of snapshots. CRBs are obtained by inverting the Fisher Information Matrix (FIM), a square matrix of dimension the number of parameters to be estimated [19]. In cases where the data are modeled as complex circular normal centered variables with covariance matrix Σ_θ depending on the vector θ of parameters to be estimated, the FIM is easily obtained by its coefficients [19, 20]

$$F_{uv} = S\operatorname{tr}\left(\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_u}\Sigma_\theta^{-1}\frac{\partial\Sigma_\theta}{\partial\theta_v}\right). \quad (18)$$

Here, for a synchronous array, the parameters to be estimated are the power p and the coordinates x, y, z, and the FIM is

$$\mathbf{F} = \left(\begin{array}{ccc|c} F_{xx} & F_{xy} & F_{xz} & F_{xp} \\ F_{yx} & F_{yy} & F_{yz} & F_{yp} \\ F_{zx} & F_{zy} & F_{zz} & F_{zp} \\ \hline F_{px} & F_{py} & F_{pz} & F_{pp} \end{array}\right) = \left(\begin{array}{cc} \mathbf{F}_{\mathbf{xx}} & \mathbf{f}_{\mathbf{x}p} \\ \mathbf{f}_{\mathbf{x}p}^H & F_{pp} \end{array}\right) \quad (19)$$

The inverse and partial derivatives of $\Sigma_{p,\mathbf{x}} = p\mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^H + \sigma^2\mathbf{I}$ are

$$\Sigma_{p,\mathbf{x}}^{-1} = \frac{1}{\sigma^2}\left(1 - \frac{p\mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^H}{\sigma^2+p\|\mathbf{g}(\mathbf{x})\|^2}\right) \quad (20)$$

$$\frac{\partial\Sigma_{p,\mathbf{x}}}{\partial p} = \mathbf{g}(\mathbf{x})\mathbf{g}(\mathbf{x})^H \quad (21)$$

$$\frac{\partial\Sigma_{p,\mathbf{x}}}{\partial x} = p\left(\frac{\partial\mathbf{g}(\mathbf{x})}{\partial x}\mathbf{g}(\mathbf{x})^H + \mathbf{g}(\mathbf{x})\frac{\partial\mathbf{g}(\mathbf{x})}{\partial x}^H\right) \quad (22)$$

and likewise for y and z. The CRB are then given by the inverse of the FIM:

$$\mathbf{C} = \mathbf{F}^{-1} \quad (23)$$

By decomposing $\mathbf{C}$ as

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{\mathbf{xx}} & \mathbf{c}_{\mathbf{x}p} \\ \mathbf{c}_{\mathbf{x}p}^H & C_{pp} \end{pmatrix} \tag{24}$$

a lower bound for the variance (and the MSE) of an unbiased estimator of the power of the source is $B_p^c = C_{pp}$, and a lower bound of the variance (and the MSE) of an unbiased estimator of the position is $B_{\mathbf{x}}^c = \mathrm{tr}\,\mathbf{C}_{\mathbf{xx}}$.

4.1. Application to asynchronous arrays

For the strict model of asynchronous arrays, the measurements are assumed to be independent, and the FIM of the complete set of measurements is obtained by summing the FIMs $\mathbf{F}_j$ of the arrays [21]:

$$\mathbf{F}^s = \sum_{j=1}^{J} \mathbf{F}_j \tag{25}$$

The FIMs $\mathbf{F}_j$ are obtained by replacing $\mathbf{g}(\mathbf{x})$ by $\mathbf{g}_j(\mathbf{x})$ in Eqs. (20), (21) and (22). The lower bounds on the variance in the strict model are $B_p^s = C_{pp}^s$ and $B_{\mathbf{x}}^s = \mathrm{tr}\,\mathbf{C}_{\mathbf{xx}}^s$, with C_{pp}^s and $\mathbf{C}_{\mathbf{xx}}^s$ defined as above, for the FIM $\mathbf{F}^s$.

In the relaxed model, a power for each array is to be estimated. The position is common, and the FIM coefficients related to the position are summed. The FIM for the parameters $(x, y, z, p_1, \ldots, p_j, \ldots p_J)$ is in this case

$$\mathbf{F}^r = \left(\begin{array}{c|ccc} \sum_{j=1}^{J} \mathbf{F}_{j,\mathbf{xx}} & \mathbf{f}_{1,p\mathbf{x}} & \cdots & \mathbf{f}_{J,p\mathbf{x}} \\ \hline \mathbf{f}_{1,p\mathbf{x}}^H & F_{1,pp} & & \\ \vdots & & \ddots & \\ \mathbf{f}_{J,p\mathbf{x}}^H & & & F_{J,pp} \end{array}\right) \tag{26}$$

The CRB matrix is here structured as

$$\mathbf{C}^r = \begin{pmatrix} \mathbf{C}_{\mathbf{xx}}^r & \mathbf{C}_{\mathbf{px}}^r \\ \mathbf{C}_{\mathbf{px}}^{rH} & \mathbf{C}_{\mathbf{pp}}^r \end{pmatrix} \tag{27}$$

As above, a lower bound of the variance of the estimator of the position is $B_{\mathbf{x}}^r = \mathrm{tr}\,\mathbf{C}_{\mathbf{xx}}^r$. A lower bound B_p^r of the variance of the estimator $\hat{p}^r$ of the power in Eq. (17) is

$$\mathrm{Var}(\hat{p}^r) \geq \mathbf{w}^H \mathbf{C}_{\mathbf{pp}}^r \mathbf{w} = B_p^r \tag{28}$$

where $w_j = \frac{S_j}{\sum S_j}$.

4.2. Examples

CRBs can be used to compare the expected performances of source localization, given the array shape, frequency, position of the source, number of snapshots, and the model used. CRBs are here given for three particular examples, comparing a synchronous array, and asynchronous arrays with the strict and relaxed models. The number of snapshots for the synchronous array is the sum of the number of snapshots for the asynchronous arrays, keeping the total duration of the measurement constant. The three tested configurations are shown on Fig. 1.

The influence of the existence of reference microphones, common to the arrays, is also investigated.

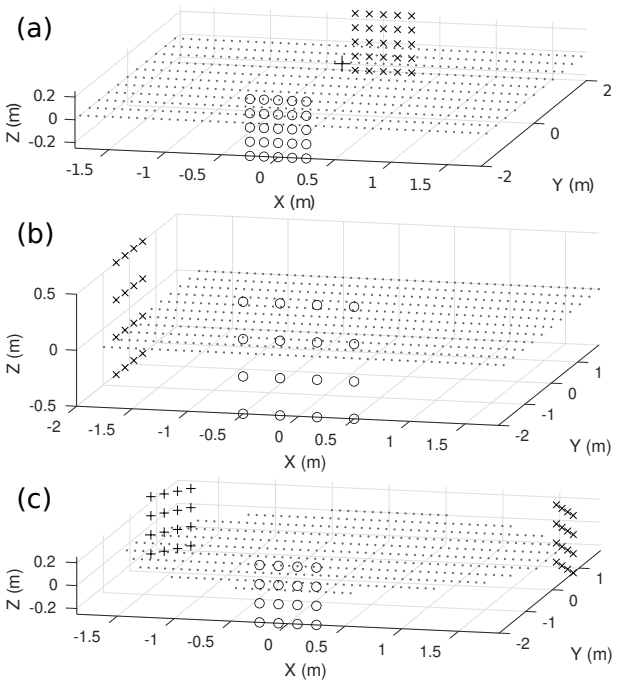

Figure 1: Array configurations used in the simulations: (a) opposite arrays, (b) L-configuration, (c) three arrays. In (a), the cross indicates the position of the source in Fig. 2

Opposite arrays. The CRBs for two arrays placed at opposite position with respect to the domain of interest are given on figures 2 and 3. Two regular square arrays of 25 microphones are used, with aperture 0.5m, placed at $y = 2$m and $y = -2$m. The power of the acoustical field generated by the source at a distance of one meter is $p = 1$ Pa^2, and the noise level $\sigma^2 = 0.5$ Pa^2. 100 snapshots are used on each array.

On figure 2 the CRBs for the position (a), and power (b), are given in function of the frequency for a source at $(0.1, 0.2, 0.0)$. The figure also shows the ratios between the CRBs for synchronous arrays and asynchronous arrays with the strict model, as well as the ratio between the bounds for the strict and relaxed asynchronous models, for position (c) and power (d). CRBs are

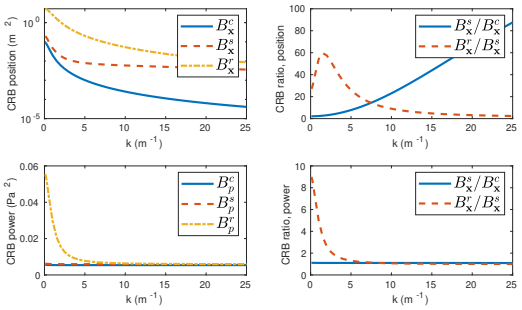

Figure 2: CRBs for the case of opposite arrays, in function of the frequency, for a source at $(0.1, 0.2, 0.0)$. Top row: position. Bottom row: power. Left column: CRBs of synchronous arrays, asynchronous arrays with strict and relaxed models. Right column: ratio between the CRBs of the strict model of asynchronous arrays, and CRBs of synchronous arrays, and ratio between the CRBs of the relaxed and strict models.

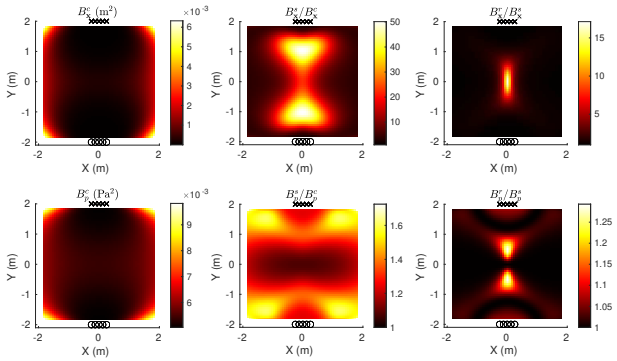

Figure 3: CRBs for the case of opposite arrays, $\kappa = 10\mathrm{m}^{-1}$. Top row: position. Bottom row: power. Left column: CRBs of synchronous arrays. Middle: ratio between the CRBs of the strict model of asynchronous arrays, and CRBs of synchronous arrays. Right: ratio between the CRBs of the relaxed and strict models. Positions of the microphones are indicated by circles and crosses.

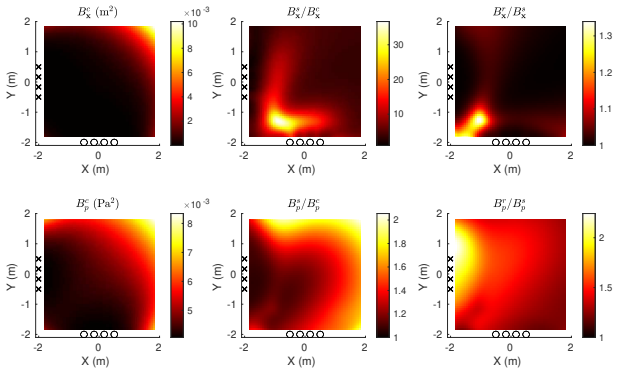

Figure 4: CRBs for the L configuration, $\kappa = 10\mathrm{m}^{-1}$. Top row: position. Bottom row: power. Left column: CRBs of synchronous arrays. Middle: ratio between the CRBs of the strict model of asynchronous arrays, and CRBs of synchronous arrays. Right: ratio between the CRBs of the relaxed and strict models. Positions of the microphones are indicated by circles and crosses.

decreasing with increasing frequency. The CRBs in position for the strict model are close to the CRBs of synchronous arrays at low frequencies. Indeed, at low frequencies, most of the information is given by the amplitude of the measurements, and not the phase. As the relaxed model cannot take into account the relative amplitudes of the measurements, its CRB is higher. At high frequencies, where most of the information is carried by the phase, the amplitude become less important, and the gap of performance between the relaxed and the strict models decreases, while the gap between synchronous and asynchronous arrays widens.

While the ratio between the CRBs in position between synchronous arrays and asynchronous arrays is large, the performance in power estimation are similar, except for the relaxed model at lower frequencies.

The CRBs are given on figure 3 for a fixed wave number $\kappa = 10\mathrm{m}^{-1}$ and varying positions, in position (a) and power (b). The ratio between the CRBs for the strict model and synchronous arrays are given in (c) and (d). It is shown that depending on the position, the loss in performance can range from negligible to substantial, in particular for position estimation. Using the relaxed model instead of the strict model decreases the performances in position, in particular near the center of the domain as shown on panel (e) showing the ratio between the CRBs for the strict and relaxed models. This is explained by the fact that the ratio between the powers received by each array is indicative of the position of the source between the arrays. This ratio cannot be exploited by the asynchronous model, where the power is allowed to change between measurements. The degradation of performances for the estimation of power is moderate, as seen on panel (f).

L configuration. CRBs in function of the position is shown on figure 4 for two regular square arrays looking at the domain of interest from orthogonal directions, in an L shape, with 16 microphones each, and an aperture of 1m, identical frequency, source power and noise level. To explore their effect on the performances, the numbers of snapshots are here 50 and 200 for the arrays along the X axis, and the Y axis respectively. Although the array configuration is symmetric along the $X = Y$ line, the CRBs are not. This is caused by the different number of snapshots used for the two arrays. Performances are here slightly better in front of the Y array than the X array.

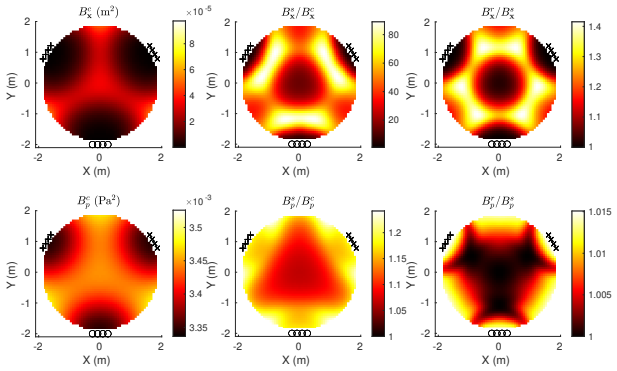

Figure 5: CRBs for the case of three arrays, $\kappa = 10\mathrm{m}^{-1}$. Top row: position. Bottom row: power. Left column: CRBs of synchronous arrays. Middle: ratio between the CRBs of the strict model of asynchronous arrays, and CRBs of synchronous arrays. Right: ratio between the CRBs of the relaxed and strict models. Positions of the microphones are indicated by circles and crosses.

The loss in performance of position estimation is large in regions between the arrays (around $(-1, -1, 0)$), that is, as in the previous case, where the directions of arrival are not sufficient for an accurate estimation of the position. Here also, the degradation in performances is amplified by using the relaxed model.

Performances in power estimation remain close to asynchronous arrays in both models.

Three arrays. Finally, the case of three arrays is considered. The array configuration and other parameters are identical to the previous case, with constant number of snapshots 100, and the arrays are regularly placed around the domain of interest. Bounds are given in figure 5. Similar conclusions as in the previous cases are reached, with a large degradation of the performances in position, but similar performances in power estimation.

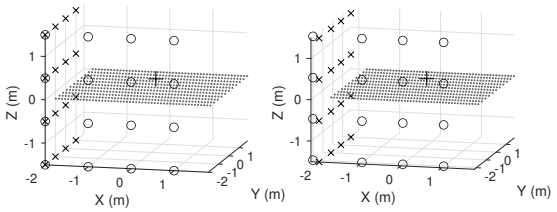

Figure 6: Arrays configurations with (left) and without (right) reference microphones.

Reference microphones. The use of reference microphones is now investigated, using perpendicular arrays shown on Fig. 6. In the first configuration, four microphone positions are common to both array, while in the second configuration, the reference microphones are slightly shifted so that no position is shared.

In the formulation of the MLE criterion (9), as well as in the computation of the CRBs, the existence of common microphones has no particular impact. It is therefore expected that having reference microphones do not have a particular influence on the attainable performance of source localization. This is confirmed by the CRBs for the two configurations shown on Fig. 7, with the CRBs for the configuration with reference microphones shown on the left panels, and the the CRBs with no reference microphone on the middle panels. The right panels show the ratio between the CRBs for the configuration without reference microphones, over the CRB for the configuration with reference microphones. CRBs are similar, and even better in position for the configuration without reference. This shows that having reference microphones has no impact on the performances, the difference in CRBs being caused by the slightly different placement of the microphones.

5. Simulations

Simulations are used to compare the performances (MSE in position and power) of the proposed estimators compared to the state of the art. Results for the opposite arrays, L configuration, and three arrays are given, for two source position each, in function of the frequency, on figures 8-10 respectively.

The following estimators are tested:

- MLE with a synchronous array, i.e. beamforming (MLE sync)
- MLE with the strict model (MLEs)
- MLE with the relaxed model (MLEr)
- arithmetic averaging (arithm.)

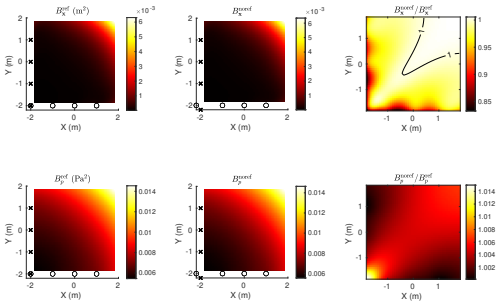

Figure 7: CRBs for arrays with and without reference, $\kappa = 10\text{m}^{-1}$. Top row: position. Bottom row: power. Left column: CRBs of arrays with five reference microphones. Middle: CRBs for arrays without reference microphones. Right: ratio between the CRBs of the configuration without reference over with references. Positions of the microphones are indicated by circles and crosses.

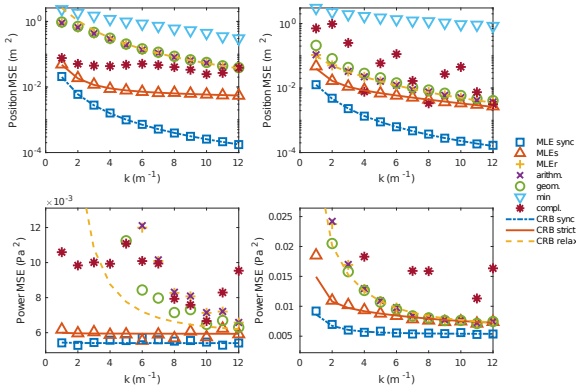

Figure 8: Mean squared errors for the position (top) and power (bottom) in function of the frequency, for opposite arrays. Left: source at $(0.1, 0, 0.2)$. Right: source at $(-1, 0, -1)$.

- geometric averaging (geom.)
- minimum (min)
- matrix completion (compl.)

In general, the simulations confirm the theoretical findings. MLE for synchronous arrays, the strict and the relaxed asynchronous models, follows the corresponding CRBs. Cases where the MLE MSE does not follows the CRBs are either due to box constraints on the position, or the fact that the plotted MSEs are an estimation of the theoretical MSE, obtained by Monte-Carlo simulations with 2000 samples.

The relaxed MLE outperforms the average mean. This improvement is negligible in cases where the number of snapshots are equal, where it is explained by the log term in Eq. (15). The improvement is more easily seen in the L configuration case, where the number of snapshots are different for the two arrays. Indeed, weighting the beamforming criterion by the number of snapshots, as in (15), gives more importance to the measurements obtained with a larger number of snapshots, which are expected to be more accurate. Similarly, geometric averaging is not capable of weighting the beamforming criteria.

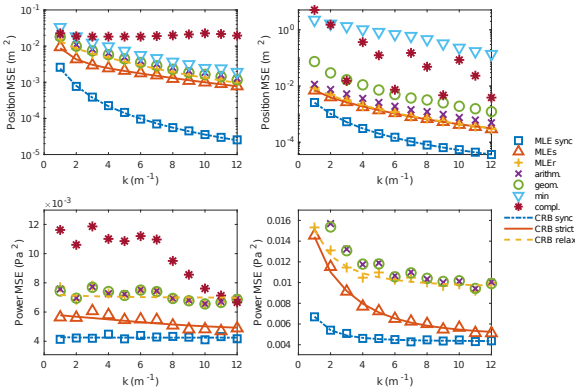

Figure 9: Mean squared errors for the position (top) and power (bottom) in function of the frequency, for the L configuration. Left: source at $(-1.35, 0.0, -0.2)$. Right: source at $(-0.43, 0.0, 0.14)$.

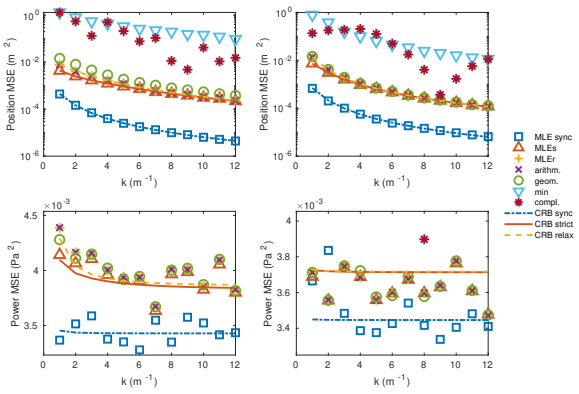

Figure 10: Mean squared errors for the position (top) and power (bottom) in function of the frequency, for three arrays. Left: source at $(0.1, 0, 0.2)$. Right: source at $(-1, 0, -1)$.

The gap in performances between the strict and relaxed models range from substantial (opposite arrays at low frequency for the position, L configuration for the power) to negligible (three arrays configuration).

Performances of minimum and matrix completion are mediocre, with markers above the top boundary of the plots for some of the configurations. For the minimum rule, this is explained by the fact that, in general, the beamforming criteria B_j do not have the same order of magnitude. As a consequence, only the smallest B_j is taken into account, which implies that only one array is used to estimate the position of the source. Numerical tests showed that using formulation III, which is an estimation of the power of the source, and should have similar values at the source location for the sub-arrays, does not improve the results. Indeed, estimation of the position of the source with formulation III is biased [3]. For matrix complexion, the algorithm in [14] is unable to recover the appropriate phase relationships between the arrays [11].

6. Experimental validation

In this section, experimental results are given to illustrate the results of the theoretical and numerical analysis of the problem. The setting of the experiment is as follows: a source is localized in a plane using two linear arrays of 8 microphones, with a step of 10cm. For each array configuration, two source positions were tested. The two configurations used are the localization of a source between two parallel arrays, and localization of a source between two perpendicular arrays, which were identified as cases where the gap in performance between the strict and relaxed MLE was the largest. The microphones are MEMS microphones (INVENSENSE - INMP441), and the source is a baffled Visaton-BF32 loudspeaker. Unwanted reflections in the room were treated by absorbing material (polyurethane foam) behind the arrays and on the floor. The sampling frequency is 20kHz, the length of each measurement is 2s, divided in 80 segments of 500 samples. The spatial covariance matrix is estimated after a Fourier analysis of the segments. Asynchronous measurements are simulated by processing the two halves of the signals separately.

The configuration of the arrays and the sources are shown on Fig. 11. The positions of the source obtained by synchronous beamforming, relaxed and strict asynchronous beamforming, for 100 trials, are shown on Fig. 12 for the four cases, at $f = 1000$ Hz for the opposite arrays and $f = 800$ Hz for the L configuration.

The estimation errors averaged over 100 realizations and 50 frequencies ranging from 400 Hz to 2360 Hz are plotted on Figure 13. We note that these errors are not only caused by measurement noise, but also by modeling errors (calibration errors, microphone positioning, non-anechoic room, etc.). Such modeling errors are not taken into account by the analysis proposed here.

In particular, in the case of source 1 of the opposite configuration, strict MLE with asynchronous arrays has better performances than synchronous arrays. This surprising result can be

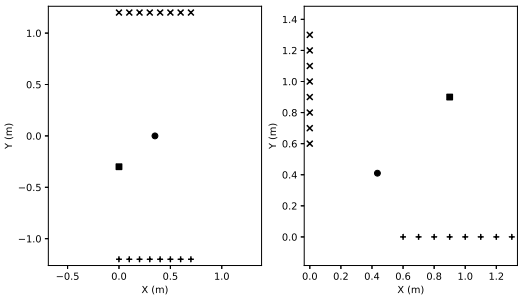

Figure 11: Experimental setups. Left: opposite arrays. Right: L configuration. Source 1 is indicated by a circle, source 2 by a square.

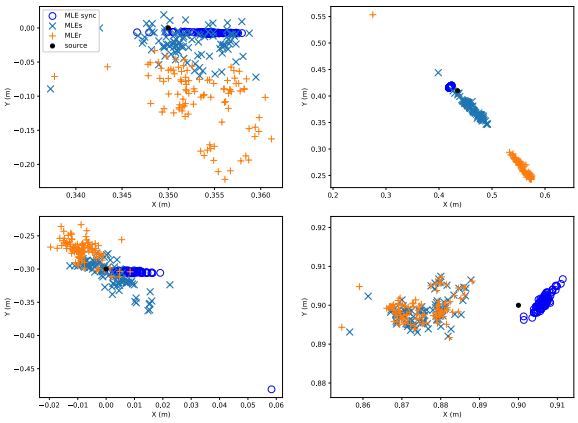

Figure 12: Scatter plots of estimated source positions. Left: opposite arrays. Right: L configuration. Top: source 1. Bottom: source 2. Note the non square axes for the opposite arrays.

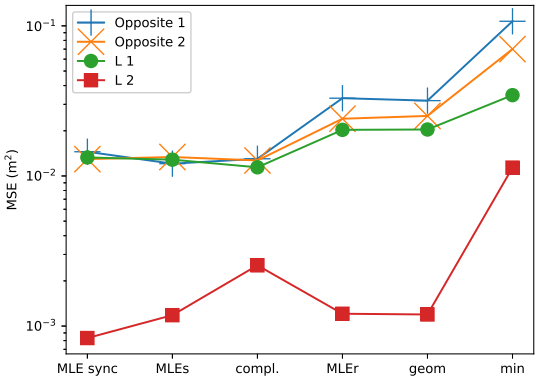

Figure 13: MSEs averaged over 100 trials and 50 frequencies.

explained by inspecting the behavior of the beamforming criterion in the asynchronous case. Fig. 14 shows the beamforming criterion, divided by its maximal value, at 880 Hz and 1200 Hz, for $X = 0.35$m and varying Y. The actual source is at $Y = 0$m. In both cases, the beamforming criterion has a local maximum close to the actual position of the source. However, because of noise, modeling errors, etc., a higher lobe is visible at $Y = 0.2$m at 880Hz, implying a large estimation error for the position of the source. Cramér-Rao bounds, which are sensitive to the local behavior of the model around the actual position, cannot explain this phenomenon.

Hermitian completion has similar performances as synchronous beamforming and the strict model in three cases, but has a larger error for the L configuration with source 2, where it is unable to recover the actual phase relationship between the measurements. As expected from the CRBs, the relaxed MLE has significantly larger errors than the strict MLE for the configurations where the source is between the arrays. Geometric averaging has similar performances as relaxed MLE, and fusion by miniming has the worst performances in all cases.

In conclusion, the best performances for asynchronous arrays are obtained by the MLE with the strict model.

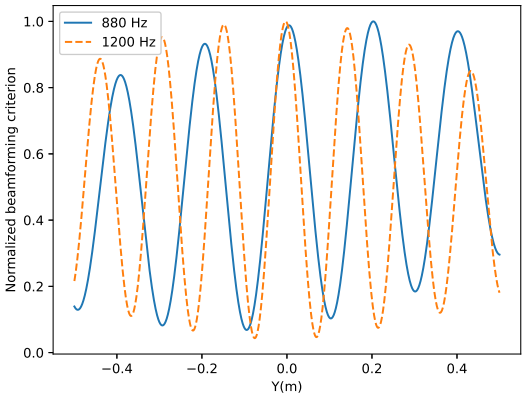

Figure 14: Bemaforming criterion for opposite arrays with source 1, in function of the Y coordinate, at 800 Hz and 1200 Hz.

7. Conclusion

Two models for source localization with asynchronous arrays were introduced, one assuming that the power of the source remains constant between the experiments, the other leaving the power free to change between experiments.

Two types of theoretical results were derived for these models. The CRBs for the estimation of the position and the power of a source were computed for several particular cases of array configurations, for synchronous arrays and the asynchronous arrays, with the two models. MLE estimators were also computed. Maximizing the MLE criterion for the strict model necessitates to maximize the criterion jointly with respect to the position and the power. In contrast, the MLE for the relaxed model can be obtained in two steps. The position is first estimated by maximizing a spatial criterion, and the power is then estimated.

The theoretical results and simulations obtained here have several practical implications. Analyzing the CRBs, it was found that the performances in position estimation were degraded by the loss of synchronicity between the arrays, and that using the relaxed model degrades the performance further, in particular at low frequencies. However, for power estimation, performances were not found to degrade significantly, except for the relaxed model at low frequencies, or varying numbers of snapshots. The MLE for the strict model have been found to yield the best performances among the tested methods. The simpler relaxed model, less accurate in general, reaches the performances of the strict model in some cases. Computing the CRBs can help identify such cases.

Simulations showed that other methods, such as geometric averaging, minimum, and matrix completion, are less accurate. Geometric averaging, as well as unweighted arithmetic averaging, cannot take into account the relative accuracy of the asynchronous beamforming maps. Fusion by minimum was found to have mediocre performances, as in general, its maximum only depends on one of the beamforming maps. The matrix completion method used here, which is not based on the continuity of the measured sound field as the arrays are spatially separated, cannot recover the phase relationship between the arrays, limiting its performances.

The conclusions differ from previous benchmarking between fusion methods [6, 9], where the fusion by minimum was found to be the most appropriate. This is explained by the different metrics used in these benchmarks, which mostly considered sidelobe levels and beamwidth, instead of the accuracy of the estimation of the parameters of the source. Geometric averaging was found to be the most accurate of the methods tested in [22]. However, geometric averaging was here shown to be less accurate in cases where the number of snapshots differs between the measurements, and less accurate that MLE with the strict model in all cases.

Experimental results confirm the superiority of the strict model compared to the relaxed model, as its performances in source localization are consistently better than that of the relaxed model. Beamforming using asynchronous arrays was found to be sensible to the experimental conditions, as high sidelobes can be source of large estimation errors, which cannot be explained by the analysis used here.

References

[1] B. Van Veen, K. Buckley, Beamforming: a versatile approach to spatial filtering, IEEE ASSP Magazine 5 (2) (1988) 4–24 doi:10.1109/53.665.

[2] R. Merino-Martínez, P. Sijtsma, M. Snellen, T. Ahlefeldt, J. Antoni, C. J. Bahr, D. Blacodon, D. Ernst, A. Finez, S. Funke, T. F. Geyer, S. Haxter, G. Herold, X. Huang, W. M. Humphreys, Q. Leclère, A. Malgoezar, U. Michel, T. Padois, A. Pereira, C. Picard, E. Sarradj, H. Siller, D. G. Simons, C. Spehr, A review of acoustic imaging methods using phased microphone arrays: Part of the "Aircraft Noise Generation and Assessment" Special Issue, CEAS Aeronautical Journal 10 (1) (2019) 197–230. doi:10.1007/s13272-019-00383-4.

[3] G. Chardon, Theoretical analysis of beamforming steering vector formulations for acoustic source localization, J. Sound Vib. 517 (2022) 116544. doi:10.1016/j.jsv.2021.116544.

[4] E. Lehmann, G. Casella, Theory of Point Estimation, 2nd Edition, Springer, 1998, chapter 6.

[5] D. Evans, M. Hartmann, J. Delfs, Beamforming for point force surface sources in numerical data, J. Sound Vib. 458 (2019) 303–319. doi:10.1016/j.jsv.2019.05.030.

[6] L. Kleine-Wächter, J. Ocker, D. Döbler, C. Puhle, G. Herold, Investigations on beamforming in the wind tunnel using multiple microphone array measurements, in: BEBEC 2018, Berlin, Germany, 2018.

[7] P. Castellini, A. Sassaroli, Acoustic source localization in a reverberant environment by average beamforming, Mech. Syst. Signal Process. 24 (3) (2010) 796–808. doi:10.1016/j.ymssp.2009.10.021.

[8] R. Porteous, Z. Prime, C. J. Doolan, D. J. Moreau, V. Valeau, Three-dimensional beamforming of dipolar aeroacoustic sources, J. Sound Vib. 355 (2015) 117–134. doi:10.1016/j.jsv.2015.06.030.

[9] L. T. Lima Pereira, R. Merino-Martínez, D. Ragni, D. Gómez-Ariza, M. Snellen, Combining asynchronous microphone array measurements for enhanced acoustic imaging and volumetric source mapping, Appl. Acoust. 182 (2021) 108247. doi:10.1016/j.apacoust.2021.108247.

[10] J. Antoni, Y. Liang, Q. Leclère, Reconstruction of sound quadratic properties from non-synchronous measurements with insufficient or without references: Proof of concept, J. Sound Vib. 349 (2015) 123–149. doi:10.1016/j.jsv.2015.03.008.

[11] L. Yu, J. Antoni, Q. Leclere, Spectral matrix completion by Cyclic Projection and application to sound source reconstruction from non-synchronous measurements, J. Sound Vib. 372 (2016) 31–49. doi:10.1016/j.jsv.2016.02.031.

[12] L. Yu, J. Antoni, H. Wu, Q. Leclere, W. Jiang, Fast iteration algorithms for implementing the acoustic beamforming of non-synchronous measurements, Mech. Syst. Signal Process. 134 (2019) 106309. doi:10.1016/j.ymssp.2019.106309.

[13] N. Chu, Y. Ning, L. Yu, Q. Huang, D. Wu, A High-Resolution and Low-Frequency Acoustic Beamforming Based on Bayesian Inference and Non-Synchronous Measurements, IEEE Access 8 (2020) 82500–82513, doi:10.1109/ACCESS.2020.2991606.

[14] F. Ning, J. Song, J. Hu, J. Wei, Sound source localization of non-synchronous measurements beamforming with block Hermitian matrix completion, Mech. Syst. Signal Process. 147 (2021) 107118. doi:10.1016/j.ymssp.2020.107118.

[15] G. Chardon, gilleschardon/asynchronous, doi:10.5281/zenodo.8041808 https://zenodo.org/record/8041808 last accessed 15/06/2023.

[16] B. Picinbono, Second-order complex random vectors and normal distributions, IEEE Trans. on Signal Process. 44 (10) (1996) 2637–2640, conference Name: IEEE Transactions on Signal Processing. doi:10.1109/78.539051.

[17] E. Sarradj, Three-Dimensional Acoustic Source Mapping with Different Beamforming Steering Vector Formulations, Adv. Acoust. Vib. 2012 (2012) e292695. doi:10.1155/2012/292695.

[18] F. Ning, J. Hu, H. Hou, K. Yao, J. Wei, B. Li, Sound source localization of non-synchronous measurements beamforming based on the truncated nuclear norm regularization, Appl. Acoust. 191 (2022) 108688. doi:10.1016/j.apacoust.2022.108688.

[19] H. L. V. Trees, Optimum Array Processing, John Wiley and Sons, Ltd, 2002, Ch. 8, pp. 917–1138. doi:https://doi.org/10.1002/0471221104.ch8.

[20] W. J. Bangs, Array processing with generalized beamformers, Ph.D. thesis, Yale University, New Haven, Connecticut (1971).
[21] W. Suleiman, P. Parvazi, M. Pesavento, A. M. Zoubir, Non-Coherent Direction-of-Arrival Estimation Using Partly Calibrated Arrays, IEEE Trans. on Signal Process. 66 (21) (2018) 5776–5788. doi:10.1109/TSP.2018.2867997.
[22] R. Merino-Martínez, B. von den Hoff, D. Morata, M. Snellen, Three-dimensional acoustic imaging using asynchronous microphone array measurements, in: BEBEC 2022, Berlin, Germany, 2022.