

Received April 8, 2020, accepted April 24, 2020, date of publication April 30, 2020, date of current version May 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2991606

# A High-Resolution and Low-Frequency Acoustic Beamforming Based on Bayesian Inference and Non-Synchronous Measurements

NING CHU<sup>1</sup>, YUE NING<sup>1</sup>, LIANG YU<sup>2</sup>, QIAN HUANG<sup>1</sup>, AND DAZHUAN WU<sup>1,3</sup>

<sup>1</sup>Institute of Process Equipment, College of Energy Engineering, Zhejiang University, Hangzhou 310027, China

<sup>2</sup>State Key Laboratory of Mechanical System and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China

<sup>3</sup>State Key Laboratory of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou 310027, China

Corresponding authors: Ning Chu (chuning@zju.edu.cn) and Liang Yu (liang.yu@sjtu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61701440 and Grant 11704248, in part by the State Key Laboratory of Mechanical System and Vibration under Grant MSV202001, in part by the Science and Technology on Sonar Laboratory under Grant 6142109KF201901, in part by the Science and Technology on Underwater Test and Control Laboratory, State Key Laboratory of Compressor Technology under Grant SKL-YSL201812 and Grant SKL-YSJ201903, and in part by the National Program on Key Research Project of China under Grant 2016YFF0203300.


**ABSTRACT** Beamforming is a powerful technique to achieve acoustic imaging in far-field. However, its spatial resolution is strongly blurred by the point spread function (PSF) of phased microphone array. Due to the limitation of array aperture and microphone density, the PSF is far from Dirac delta function, so that it is difficult to obtain a high-resolution beamforming image at low-frequencies (e.g.500-1500Hz). This paper proposes a Bayesian inference method based on Non-synchronous Array Measurements (Bi-NAM) so as to refine the PSF and break through the beamforming limitation for low-frequency source imaging. Firstly, by sequentially moving prototype array at different positions, the non-synchronous measurements can get a sizeable synthetic aperture and high density of microphones. The synthetic cross-spectrum matrix (CSM) can significantly improve the beamforming performance. To confine the approximation error of synthetic CSM and the uncertainty of forward model, as well as the noise interference, a Bayesian inference based on joint maximum a posterior (JMAP) is proposed to solve an ill-posed inverse problem. A Student-t prior is employed to enforce the sparse property of acoustic strength distribution. The background noise can be adaptively modeled by the Student-t distribution, which is related to some of the typical symmetric distributions. Then the hyper-parameters in JMAP inference are efficiently estimated by the Bayesian hierarchical framework. Through experimental data, the proposed Bi-NAM approach is confirmed to achieve high-resolution acoustic imaging at 1000Hz and 800Hz, respectively, even under the Laplace noise interference.

**INDEX TERMS** Acoustic imaging, Bayesian inference, beamforming, cross-spectrum matrix, joint maximum a posterior, non-synchronous measurements, point spread function, student-t prior.

## I. INTRODUCTION

Acoustic source imaging has widespread applications such as city-noise mapping, industry noise monitoring and mechanical fault diagnosis, etc. The spatial position and strength distribution can be visualized by back-projecting the acoustic pressure measurements of phased microphone array. In the far-field, the Conventional Beamforming (CBF) is a fast and robust method to get an intensity map of acoustic distribu-

tion [1]–[4]. However, the spatial resolution of the CBF map is strongly blurred by the spatial impulse response or point spread function (PSF) of phased microphone array [5], [6]. Due to the limitation of array aperture and microphone density, the PSF is far from an ideal impulse response, which should be as sharp and abrupt as the Dirac delta function. In this sense, the CBF map can be regarded as a convolution result of the PSF (modeled by a convolution kernel) and a ground-truth image of source intensity. In fact, the PSF spectrum can be expressed as a spatial low-frequency filter. As for the small aperture-array and low-frequency sources,

The associate editor coordinating the review of this manuscript and approving it for publication was Stavros Ntalampiras .

the cut-off frequency of PSF spectrum becomes so narrow that it sacrifices both high and low frequency information of the acoustic source. Meanwhile, the spatial shape of PSF distribution becomes so fat and smooth that it not only makes the CBF map quite ambiguous due to the convolution effect, but also causes the PSF spatially-shifted, which cannot be easily modeled by a single convolution kernel for different source positions.

To improve the spatial resolution of the CBF at low frequencies, the fundamental approach is to clean up the PSF influence or make the PSF close to Dirac delta function. There are at least two ways to follow.

One way is the soft-calibration-based method, which can attenuate the blurred effect of the PSF directly from the CBF map without changing the topology of the microphone array. In this way, soft-calibration methods can be generally classified into three categories: (1) Spatial filtering methods [2]–[7], such as adaptive beamforming based on the Minimum Variance Distortionless Response (MVDR), and the eigenvalue decomposition of Cross-Spectral Matrix (CSM) based on the Multiple Signal Classification (MUSIC). (2) Deconvolution-based methods [8]–[10] such as CLEAN-SC based on the sparsity constraint, the Deconvolution Approach for the Mapping of Acoustic Sources (DAMAS) and its extensions. (3) Regularization-based methods [11]–[23] such as Tikhonov regularization for denoising, Total Variation (TV) regularization for artifact flaw removal, compressive sensing (CS) beamforming for high-resolution, and Bayesian inference approaches for super-resolution. Though the filtering methods can efficiently improve the CBF resolution, the selections of filter parameters need the knowledge of the acoustic source and noise interference. Deconvolution methods can get a high spatial resolution without knowing the source priors. Still, the deconvolution is prone to cause artifact flaws if the PSF is not known, and sometimes it is sensitive to strong background noise and not fast to converge. By adding the sparsity constraints, the sparse regularization methods are capable of obtaining much higher resolution and fewer artifacts than the deconvolution in the case of strong noise interference. But regularization parameters should be tuned carefully. Sometimes they are not adaptive to scenario changes or non-Gaussian background noise.

The other way is the non-synchronous measurement beamforming method, which can get a large aperture of the synthetic array and high density of microphones, and make the PSF more ideal than that of the prototype array. Non-synchronous measurements can be implemented in two steps. Firstly, a series of multi-angle observations can be obtained by sequentially moving the prototype array at the different positions and scanning the same source plane. But due to the non-synchronous measurements, the phase relationships between consecutive snapshots from different array positions are missing, and it results in the missing items at the synthetic cross-spectral matrix (CSM), as Yu *et al.* [24] pointed out. In other words, in this large dimension of synthetic CSM, its diagonal blocks consist of small-size CSMs of

prototype array measurements at different array positions, but other non-diagonal blocks are the missing items. Therefore, in the second step, the synthetic CSM can be completed to improve the beamforming performance. This synthetic CSM completion problem can be solved by using the property of the low-rank matrix and the continuity of the acoustic fields. However, even if the acoustic sources are ideal and stationary monopoles, it is inevitably introducing interpolation errors during the filling completion of synthetic CSM.

Most of these drawbacks can be overcome by Bayesian inference methods [15], [20], [25]. It can adaptively estimate both unknown random variables and unknown model parameters by applying the Bayes' rule in updating the probability law, in which, a posterior probability can be obtained from the likelihood and prior models. And the likelihood can be derived from the forward model using measured data. The prior models can be assigned according to prior information on the unknowns. The priors serve to promote useful regularizations on ill-posed inverse problems caused by non-synchronous measurements. For example, the priors on acoustic sources can introduce valuable information such as sparsity property of monopole source distribution, so that the optimal estimation can be derived from many possible solutions due to ill-posed inversion. Moreover, to describe the forward model uncertainty, both the model approximation error and array noise interference can be modeled by a proper probability density function, so that the likelihood probability function can be derived to build up a mathematical bridge from priors to posteriors. Finally, hyper-parameters (including source strength, prior model parameters, likelihood model parameters, and related hidden variables) can be estimated alternatively under the Bayesian optimization criteria such as Maximum A Posterior (MAP), minimum Kullback-Leibler (KL) divergence, etc. Though Bayesian inference methods can breakthrough the CBF limitation, it still requires tremendous calculation for optimization and convergence.

The contributions in this paper are (1) to propose Bayesian inference method based on Non-synchronous microphone Array Measurements (Bi-NAM) in order to refine the PSF and achieve high-resolution imaging at low-frequencies; (2) to develop Student-t distribution as sparsity-enforcing prior so as to promote the spatial resolution. Meanwhile, the background noise is also modeled by Student-t distribution which can be adaptively fit for Gaussian and non-Gaussian noise interference such as Cauchy or Laplace noise. (3) to implement the Joint Maximum A Posterior (JMAP) based on alternate iteration optimization in the Bayesian hierarchical framework so as to efficiently solve an ill-posed inverse problem.

This paper is organized as follows. The forward model based on non-synchronous measurements is presented in section II. The proposed Bayesian inference is described in section III. Section IV provides experimental results at low-frequency localization to validate the proposed method. Finally, section V summarizes this paper and gives further prospects.

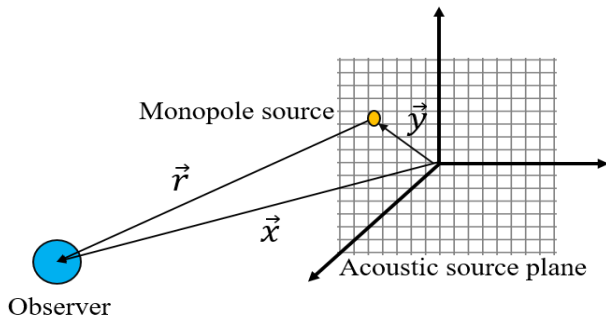


FIGURE 1. Propagation model of stationary monopole source in far-field.

II. FORWARD MODEL OF NON-SYNCHRONOUS MEASUREMENTS

A. NON-SYNCHRONOUS MEASUREMENTS

This paper mainly considers the 2D imaging of acoustic sources. For a static and stationary source in the far-field in Fig. 1, the acoustic pressure distribution at the observed position vector  $\vec{x}$  and time  $t$  can be given as:

$$\left(\frac{1}{c^2} \frac{\partial^2}{\partial t^2} - \nabla^2\right) p(\vec{x}, t) = s(\vec{x}, t) \tag{1}$$

where  $p(\vec{x}, t)$  denotes distribution function of acoustic pressure in the far-field and  $s(\vec{x}, t)$  denotes source distribution function in the planar plane.  $c$  is acoustic propagation speed in the uniform media, normally  $c = 340m/s$  in air.  $\frac{\partial^2(\cdot)}{\partial(\cdot)^2}$  denotes the second partial derivative. According to the superposition principle of the acoustic field, the measured acoustic pressure at an arbitrary position can be expressed as follows:

$$p(\vec{x}, t) = \int_{-\infty}^{+\infty} \int_{V_y} s(\vec{y}, t') G(\vec{x}, t|\vec{y}, t') dV_y dt' \tag{2}$$

where  $dV_y$  denotes the volume infinitesimal at the source position  $\vec{y}$ . And  $G(\vec{x}, t|\vec{y}, t')$  is the Green function defined in (3), which reflects acoustic field response from the source position  $\vec{y}$  and time  $t'$  to observed position  $\vec{x}$  and time  $t$ .

$$g(\vec{x}|\vec{y}) = \frac{1}{4\pi |\vec{x} - \vec{y}|} \exp^{-j\frac{2\pi f}{c} |\vec{x} - \vec{y}|} = \frac{1}{4\pi |\vec{r}|} \exp^{-jk|\vec{r}|} \tag{3}$$

where  $f$  denotes acoustic frequency and  $k = \frac{2\pi f}{c}$  denotes the number of waves.  $\exp(\cdot)$  denotes exponential function based on Euler's number.  $|\cdot|$  denotes the module of a vector.  $j$  is the imaginary unit.  $|\vec{r}| = |\vec{x} - \vec{y}|$  denotes the distance between source and observer, as shown in Fig. 1. And  $|t' - t| = \frac{|\vec{x} - \vec{y}|}{c}$  is also considered by  $G(\vec{x}|\vec{y})$  for a static and stationary source. So that the propagation model can be expressed in the frequency domain as:

$$p(\vec{x}, f) = g(\vec{x}|\vec{y}) s(\vec{y}, f) \tag{4}$$

A continuous propagation model is expressed in (4). When the microphone array is used to observe the acoustic signals, a discrete propagation model is derived from the continuous model. Assume that the source plane is divided into  $N$  grids,

There are  $K$  incoherent monopole sources in the source plane, satisfying  $K \ll N$ . An array composed of  $M$  ( $K < M \ll N$ ) microphones is moved onto  $Z$  positions on the array plane, which is parallel to the source plane. The time-domain sampling is transformed to the frequency-domain analysis, and snapshots are used to make wide-band signal separated into many narrow-bands under the assumption of Gaussian wide-sense stationary process, then measured pressures  $\mathbf{P}^{[z]} = (p_1^{[z]}, \dots, p_m^{[z]}, \dots, p_M^{[z]})^T \in \mathcal{C}^{M \times 1}$  of the  $z$ -th array position ( $z = 1, \dots, Z$ ) at a frequency  $f$  corresponding to the  $l$ -th ( $l = 1, \dots, L$ ) snapshot can be expressed as:

$$\mathbf{P}_l^{[z]}(f) = \mathbf{G}^{*[z]} \mathbf{S}_l^*(f) + \mathbf{E}_l^{[z]}(f) \tag{5}$$

where  $\mathbf{S}_l^* = (s_1^*, \dots, s_k^*, \dots, s_K^*)^T \in \mathcal{C}^{K \times 1}$  denotes acoustic pressures of  $K$  uncorrelated monopole sources in the frequency domain. The symbol  $(\cdot)^T$  denotes matrix transpose.  $\mathbf{E}_l^{[z]} = (e_1^{[z]}, \dots, e_m^{[z]}, \dots, e_M^{[z]})^T \in \mathcal{C}^{M \times 1}$  denotes model uncertainty, including noise interference and model approximation.  $\mathbf{G}^* \{z\} \in \mathcal{C}^{M \times K}$  is the Green function matrix, whose element  $g(\vec{x}|\vec{y})$  is expressed in (3), reflecting the propagation response from sources to microphones. According to (3), the propagation vector  $g_{m,k}^{[z]}$  from the  $k$ -th source to the  $m$ -th microphone is derived as:

$$g_{m,k}^{*[z]} = \frac{\exp^{-j\frac{2\pi f}{c} r_{m,k}}}{4\pi r_{m,k}} \tag{6}$$

where  $r_{m,k}$  is the distance between  $k$ -th source and the  $m$ -th microphone.

However, it is a non-linear system of equations in (5), since source signals and source positions are both unknown. To transform (5) into a linear system, the classical approach is to discretize the acoustic source plane into a large number of identical grids, and each grid is regarded as a possible position for the acoustic source, as shown in Fig. 1.

To obtain the CBF result at the  $z$ -th position, a discrete forward model can be derived from (5) as:

$$\mathbf{P}_l^{[z]}(f) = \mathbf{G}^{[z]} \mathbf{S}_l(f) + \mathbf{E}_l^{[z]}(f) \tag{7}$$

where  $\mathbf{S}_l = (s_1, \dots, s_n, \dots, s_N)^T \in \mathcal{C}^{N \times 1}$  denotes the acoustic pressures of each grid. Here discrete source  $\mathbf{S}_l$  is supposed to contain physical source  $\mathbf{S}_l^*$ . Correspondingly,  $\mathbf{G}^{[z]} = (g_1^{[z]}, \dots, g_n^{[z]}, \dots, g_N^{[z]}) \in \mathcal{C}^{M \times N}$  is discrete Green function matrix, whose element  $g_{m,n}^{[z]}$  is discrete propagation vector, derived from (6) as:

$$g_{m,n}^{[z]} = \frac{\exp^{-j\frac{2\pi f}{c} r_{m,n}}}{4\pi r_{m,n}} \tag{8}$$

where  $r_{m,n}$  denotes the distance between the  $n$ -th grid and the  $m$ -th microphone. Then beamforming output at the  $n$ -th grid of the  $z$ -th array position is given as:

$$y_n(f) = \frac{(\mathbf{g}_n^{[z]})^H \widehat{\mathbf{R}}_{CSM}^{[z]}(f) \mathbf{g}_n^{[z]}}{\|\mathbf{g}_n^{[z]}\|_2^4}, \quad n = 1, \dots, N \tag{9}$$

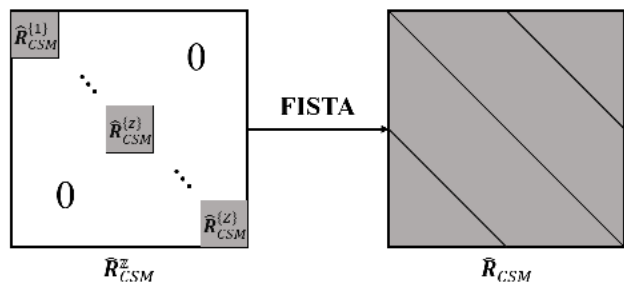


FIGURE 2. Item missing CSM  $\hat{\mathbf{R}}_{CSM}^z$  (left) and full CSM  $\hat{\mathbf{R}}_{CSM}$  (right) completed by FISTA.

where  $\hat{\mathbf{R}}_{CSM}^{[z]}(f)$  is the estimated  $M \times M$  CSM of the measured pressure  $\mathbf{P}_l^{[z]}(f)$  in (7).  $\|\cdot\|_l$  denotes  $l$ -norm, and  $(\cdot)^H$  denotes the complex conjugate transpose. It is noted that  $\mathbf{R}_{CSM}^{[z]}(f)$  is defined as the expectation of the inner product of acoustic pressure vectors in (10). Assume that the measured signals are stochastically stationary, thus  $\mathbf{R}_{CSM}^{[z]}(f)$  can be estimated by averaging the snapshot blocks over time as:

$$\begin{cases} \mathbf{R}_{CSM}^{[z]}(f) = \mathbb{E} \left[ \mathbf{P}_l^{[z]}(f) \left( \mathbf{P}_l^{[z]}(f) \right)^H \right] \\ \hat{\mathbf{R}}_{CSM}^{[z]}(f) \approx \frac{1}{L} \sum_{l=1}^L \left[ \mathbf{P}_l^{[z]}(f) \left( \mathbf{P}_l^{[z]}(f) \right)^H \right] \end{cases} \quad (10)$$

where  $\mathbb{E}[\cdot]$  denotes mathematical expectation. In the following, we omit  $f$  for the simplicity of math symbols.

For non-synchronous measurements, the final goal is to achieve the same results as the simultaneous measurements of  $Z$  sets of arrays at  $Z$  positions. Therefore, the critical problem boils down to a matrix completion from the item-missing CSM  $\hat{\mathbf{R}}_{CSM}^z \in \mathcal{C}^{MZ \times MZ}$  composed by  $\hat{\mathbf{R}}_{CSM}^{[z]} \in \mathcal{C}^{M \times M}$  to the full CSM  $\hat{\mathbf{R}}_{CSM} \in \mathcal{C}^{MZ \times MZ}$ , as shown in Fig. 2.

The matrix completion problem can be solved by the Fast Iterative Shrinkage Thresholding Algorithm (FISTA), as introduced in [24]. Based on the completed CSM, the improved beamforming is derived from (9) as:

$$y_n = \frac{\mathbf{g}_n^H \hat{\mathbf{R}}_{CSM} \mathbf{g}_n}{\|\mathbf{g}_n\|_2^4}, \quad n = 1, \dots, N \quad (11)$$

where  $\mathbf{G} = [\mathbf{G}^{[1]}; \dots; \mathbf{G}^{[z]}; \dots; \mathbf{G}^{[Z]}] = (\mathbf{g}_1, \dots, \mathbf{g}_n, \dots, \mathbf{g}_N) \in \mathcal{C}^{MZ \times N}$  is the full discrete propagation matrix.  $\mathbf{G}^{[z]}$  is derived from (8).

### B. FORWARD MODEL OF POWER PROPAGATION

Based on the assumption of incoherent monopole sources, the full CSM  $\mathbf{R}_{CSM}$  can be expressed as:

$$\begin{aligned} \mathbf{R}_{CSM} &= \mathbf{G} \mathbb{E} \left[ \mathbf{S}_l \left( \mathbf{S}_l \right)^H \right] \mathbf{G}^H + \mathbb{E} \left[ \mathbf{E}_l \left( \mathbf{E}_l \right)^H \right] \\ &= \sum_{n=1}^N x_n \mathbf{g}_n \mathbf{g}_n^H + \mathbb{E} \left[ \mathbf{E}_l \left( \mathbf{E}_l \right)^H \right] \end{aligned} \quad (12)$$

where  $\mathbf{E}_l = (\mathbf{E}_l^{[1]}, \dots, \mathbf{E}_l^{[z]}, \dots, \mathbf{E}_l^{[Z]})^T \in \mathcal{R}^{MZ \times 1}$  denotes the model uncertainty of sequential measurements and noise

interference at  $Z$  array positions, whose item  $\mathbf{E}_l^{[z]}$  is defined in (5).  $\mathbf{x} = \text{diag}(\mathbb{E}[\mathbf{S}_l (\mathbf{S}_l)^H])$  denotes the powers of acoustic sources,  $\mathbf{x} = (x_1, \dots, x_n, \dots, x_N)^T \in \mathcal{R}^{N \times 1}$  with  $x_n$  being the power of a discrete source (grid) at the  $n$ -th grid. And  $\text{diag}(\cdot)$  means diagonal items of a matrix.

Inserting (12) into (11), the beamforming output of non-synchronous measurements can be expressed as:

$$\begin{aligned} y_{n_1} &= \frac{\mathbf{g}_{n_1}^H \mathbf{R}_{CSM} \mathbf{g}_{n_1}}{\|\mathbf{g}_{n_1}\|_2^4} \\ &= \frac{\mathbf{g}_{n_1}^H \sum_{n_2=1}^N x_{n_2} \mathbf{g}_{n_2} \mathbf{g}_{n_2}^H \mathbf{g}_{n_1}}{\|\mathbf{a}_{n_1}\|_2^4} + e_{n_1} \\ &= \sum_{n_2=1}^N x_{n_2} \frac{|\mathbf{g}_{n_1}^H \mathbf{g}_{n_2}|^2}{\|\mathbf{g}_{n_1}\|_2^4} + e_{n_1} \quad (n_1, n_2 = 1, \dots, N) \end{aligned} \quad (13)$$

A linear equation group can summarize the discrete forward model of acoustic power propagation in (13) as:

$$\begin{cases} \mathbf{y} = \mathbf{H} \mathbf{x} + \mathbf{e} \\ h_{n_1, n_2} = \frac{|\mathbf{g}_{n_1}^H \mathbf{g}_{n_2}|^2}{\|\mathbf{g}_{n_1}\|_2^4} \end{cases} \quad (14)$$

where  $\mathbf{e} = (e_1, \dots, e_n, \dots, e_N)^T \in \mathcal{R}^{N \times 1}$  with  $e_n$  related to the  $n$ -th grid.  $\mathbf{e}$  denotes the model uncertainty, including the noise interference at each microphone and the approximation error of the forward model.  $h_{n_1, n_2}$  is the element of power propagation matrix  $\mathbf{H} \in \mathcal{R}^{N \times N}$ .

In (13) and (4), the beamforming output  $y_{n_1}$  at the  $n_1$ -th grid mainly depends on the weighted combination of all discrete power vector  $\mathbf{x}$  and its weight  $h_{n_1, n_2}$ , which comes from each row of the power propagation matrix  $\mathbf{H}$ . In (14), the CBF result  $\mathbf{y}$  of non-synchronous measurements depends on the weighted combinations ( $\mathbf{H} \mathbf{x}$ ) of all discrete grids. Then the source power  $\mathbf{x}$  can be estimated from an inverse problem in (14) when given  $\mathbf{y}$  under model uncertainty  $\mathbf{e}$ .

Compared with the discrete forward model of acoustic signals in (7), the advantages of (14) are 1) it is a system of determined equations, whereas (7) is an underdetermined system of equations, because the number of microphones is much less than the number of scanning grids ( $M \ll N$ ); 2) it directly infers the acoustic source power  $\mathbf{x}$  instead of acoustic signals  $\mathbf{S}_l$  (including phases), and source power is directly used to visualize the acoustic strength distribution (acoustic imaging); 3) it gives more useful constraints on the ill-posed inverse problem such as positivity ( $\mathbf{x} \geq \mathbf{0}$ ) and sparsity ( $K \ll N$ ) of source powers than those in (7).

### C. PROPOSED CONVOLUTION APPROXIMATION

To make an insight on the PSF influence at the non-synchronous measurement beamforming, in this section, we propose a convolution model to approximate the forward model of power propagation in (14). We find out that power propagation matrix  $\mathbf{H}$  seems to be a quasi-Symmetric Block Toeplitz (SBT) matrix in the far-field measurement, so that

the (in)variant convolution kernels (sizes and values) can be derived from this SBT matrix [5], [6]. Thus a 2D-convolution forward model can be used to approximate  $\mathbf{H}\mathbf{x}$  as:

$$\begin{cases} \mathbf{H}\mathbf{x} \approx [\mathbf{h}^* * \mathbf{x}_0]' \\ \mathbf{y} = [\mathbf{h}^* * \mathbf{x}_0]' + \mathbf{e} \end{cases} \quad (15)$$

where  $\mathbf{x}_0$  denotes the ground-truth image of acoustic powers. Here  $\mathbf{x}_0$  is the matrix form of  $\mathbf{x}$ . And  $[\cdot]'$  denotes the vectorization operator transforming a matrix to a vector in column order.  $\mathbf{h}^* \in \mathcal{R}^{A \times B}$  is a 2D-convolution kernel (PSF).  $*$  denotes the valid convolution, whose output is made up of the overlap parts without zero-padded edges, so that the size of the output is the same as the input.  $A \times B$  denotes the size of the PSF. Normally, the PSF with square form ( $A = B$ ) is mostly used, and its size is no larger than that of  $\mathbf{x}_0$ . It is noted that in (15),  $\mathbf{e}$  denotes the model uncertainty, including noise interference, the approximation errors of non-synchronous measurements, as well as the 2D-convolution.

In the forward convolution model of (15), the beamforming result  $\mathbf{y}$  can be seen as the convolution result of acoustic source power image  $\mathbf{x}_0$  and PSF  $\mathbf{h}^*$ . But sometimes, the 2D-convolution kernel is spatially variant for the large-surface scanning, that is, the same acoustic source is blurred by different PSFs in the non-synchronous measurements. To simplify this problem, suppose that the convolution kernel is spatially invariant. For the  $n$ -th source, an average distance  $\bar{r}_n$  is defined to satisfy the condition  $\bar{r}_n/r_{mn} \approx 1$  as:

$$\bar{r}_n = \frac{1}{M} \sum_{m=1}^M r_{m,n} \quad (16)$$

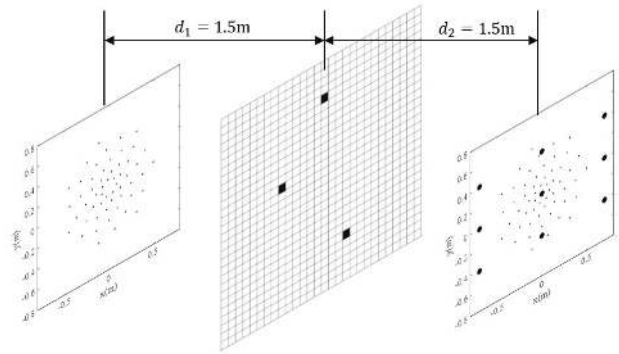
Based on the above assumption, each element  $h_{n_1, n_2} \in \mathbf{H}$  in (14) can be approximated by:

$$\begin{aligned} h_{n_1, n_2} &= \left| \frac{\sum_{m=1}^M \frac{\exp^{-j\frac{2\pi f}{c}(r_{m, n_2} - r_{m, n_1})}}{r_{m, n_1} r_{m, n_2}}}{\sum_{m=1}^M \frac{1}{r_{m, n_1}^2}} \right|^2 \\ &\approx \bar{r}_{n_1}^2 \frac{1}{M^2} \left| \sum_{m=1}^M \exp^{-j\frac{2\pi f}{c}(r_{m, n_2} - r_{m, n_1})} \right|^2 \frac{1}{\bar{r}_{n_2}^2} \end{aligned} \quad (17)$$

Then the power propagation matrix can be separated into an approximated SBT matrix  $\tilde{\mathbf{H}}$  and two diagonal matrices  $\mathbf{D}_1$  and  $\mathbf{D}_2$  as:

$$\begin{aligned} \mathbf{H} &\approx \mathbf{D}_1 \tilde{\mathbf{H}} \mathbf{D}_2 \\ \text{with } \begin{cases} \tilde{h}_{n_1, n_2} = \frac{1}{M^2} \left| \sum_{m=1}^M \exp^{-j\frac{2\pi f}{c}(r_{m, n_2} - r_{m, n_1})} \right|^2 \\ \mathbf{D}_1 = \text{Diag} \left[ \bar{r}_{n_1}^2 \right], & n_1 = 1, \dots, N \\ \mathbf{D}_2 = \text{Diag} \left[ \frac{1}{\bar{r}_{n_2}^2} \right], & n_2 = 1, \dots, N \end{cases} \end{aligned} \quad (18)$$

where  $\text{Diag}[\cdot]$  denotes the operator transforming a vector to a diagonal matrix. For the approximated SBT  $\tilde{\mathbf{H}}$ , its middle row ( $n_1 = \lfloor (N+1)/2 \rfloor$ ) contains more diverse items than the other rows. Thus the PSF of microphone array can be derived from the elements in the middle row of SBT  $\tilde{\mathbf{H}}$ . In (15),



**FIGURE 3. Demonstration of non-synchronous measurements. Prototype microphone array (left), acoustic source plane (middle), non-synchronous measurements of prototype microphone arrays at 9 positions (right).**

suppose  $A = B = \sqrt{N}$ , the invariant convolution kernel  $\mathbf{h}^* = [h_{a,b}^*]_{A \times B}$  is derived from  $\tilde{\mathbf{H}}$  in (18) as:

$$\begin{cases} h_{a,b}^* = \tilde{h}_{n_1, n_2} \\ n_1 = \left\lfloor \frac{N+1}{2} \right\rfloor \\ n_2 = n_1 + \left( \left\lfloor \frac{\sqrt{N}+1}{2} \right\rfloor - a \right) \sqrt{N} + \left\lfloor \frac{\sqrt{N}+1}{2} \right\rfloor - b \end{cases} \quad (19)$$

where  $\lfloor \cdot \rfloor$  denotes the integer part. A real and non-negative symmetric  $\mathbf{h}^*$  matrix is obtained to approximate the spatially invariant PSF of the non-synchronous measurements.

In Fig. 3, the distance between planar acoustic sources and the planar array is 1.5 m. These two planes are parallel to get a 2D acoustic imaging. The sub-figures in the left, middle, and right are respectively the prototype of microphone array with 56 channels, acoustic source plane, and non-synchronous measurements of the prototype array at 9 following positions: the centre of the prototype array are:  $(-0.41, -0.7)$ ,  $(-0.41, 0)$ ,  $(-0.41, 0.7)$ ,  $(0, -0.7)$ ,  $(0, 0)$ ,  $(0, 0.7)$ ,  $(0.41, -0.7)$ ,  $(0.41, 0)$ ,  $(0.41, 0.7)$ . Here the acoustic source plane is divided into 841 grids ( $N = 841$ ) with 29 rows and 29 columns. The aperture of the prototype array is about 1 m. The synthetic aperture of the non-synchronous array is about 2.4 m.

Setting that the analysis frequencies are 800 Hz and 1000 Hz respectively. According to the definition of the spatially invariant 2D-convolution kernel in (18), the PSF comparisons between the prototype array and non-synchronous measurements are demonstrated in Fig. 4. In far-field, the PSF would theoretically be a Dirac delta function under ideal conditions such as large enough array aperture, high enough microphone density, etc. Although the PSF is not ideal in experiments, it can still be measurable by the array response. According to the results in Fig. 4(a) and 4(b), when a prototype array is used to measure the acoustic pressures, the response region of PSF has been flatted over the scanning plane, and the form of PSF is fat and smooth, which results

in significant blur-effect on beamforming result. In contrast, the non-synchronous measurements successfully refine the PSF performance by reducing significantly the PSF volume and sharpening the PSF shape. Moreover, the PSF of the prototype array at 800 Hz in Fig. 4(c) has apparent deterioration from the one at 1000 Hz in Fig. 4(b). On the contrary, the PSF of non-synchronous measurements at 800 Hz is as fine as that in Fig. 4(b).

### III. REGULARIZATION AND BAYESIAN INFERENCE

#### A. REGULARIZATION IN BAYESIAN FRAMEWORK

In this section, two classical regularization methods are developed in the Bayesian framework. For a stochastically stationary process, the Gaussian distribution is regarded as the most common prior to the unknown variables. The Gaussian priors are assigned to model uncertainty in (14) or (15): each component of  $\mathbf{e}$  is approximated as a zero-mean Gaussian distribution with unknown variance  $v_e$  and all the elements of  $\mathbf{e}$  are independent of each other. The rationality of this assumption is detailed in the reference [3], [14], [19], [29]. Therefore the prior distribution of  $\mathbf{e}$  and likelihood function can be expressed as:

$$\begin{cases} \varphi(\mathbf{e} | v_e) = \prod_{i=1}^N \mathcal{N}(0, v_e) = (2\pi v_e)^{-\frac{N}{2}} \exp^{-\frac{\|\mathbf{e}\|_2^2}{2v_e}} \\ \varphi(\mathbf{y} | \mathbf{x}) = \mathcal{N}(\mathbf{H}\mathbf{x}, v_e) = (2\pi v_e)^{-\frac{N}{2}} \exp^{-\frac{\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2}{2v_e}} \end{cases} \quad (20)$$

where  $\mathcal{N}(0, v_e)$  stands for Gaussian independent distribution. To obtain the posterior distribution of the estimated object  $\mathbf{x}$ , proper priors should be assigned to  $\mathbf{x}$ . Here two specific prior distributions are discussed later (seen in Fig. 5).

Firstly, a Gaussian prior is attributed to  $\mathbf{x}$ : each component of  $\mathbf{x}$  is approximated by a zero-mean Gaussian independent distribution with unknown squared variance  $v_x$ . Due to the Bayesian posterior law, the posterior probability can be obtained as:

$$\begin{cases} \varphi(\mathbf{x} | \mathbf{y}) \propto \varphi(\mathbf{y} | \mathbf{x}) \varphi(\mathbf{x}) \\ -\ln[\varphi(\mathbf{y} | \mathbf{x}) \varphi(\mathbf{x})] = \frac{1}{2v_e} \left( \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{v_e}{v_x} \|\mathbf{x}\|_2^2 \right) \\ \quad + \frac{N}{2} \ln(2\pi v_e) + \frac{N}{2} \ln(2\pi v_x) \end{cases} \quad (21)$$

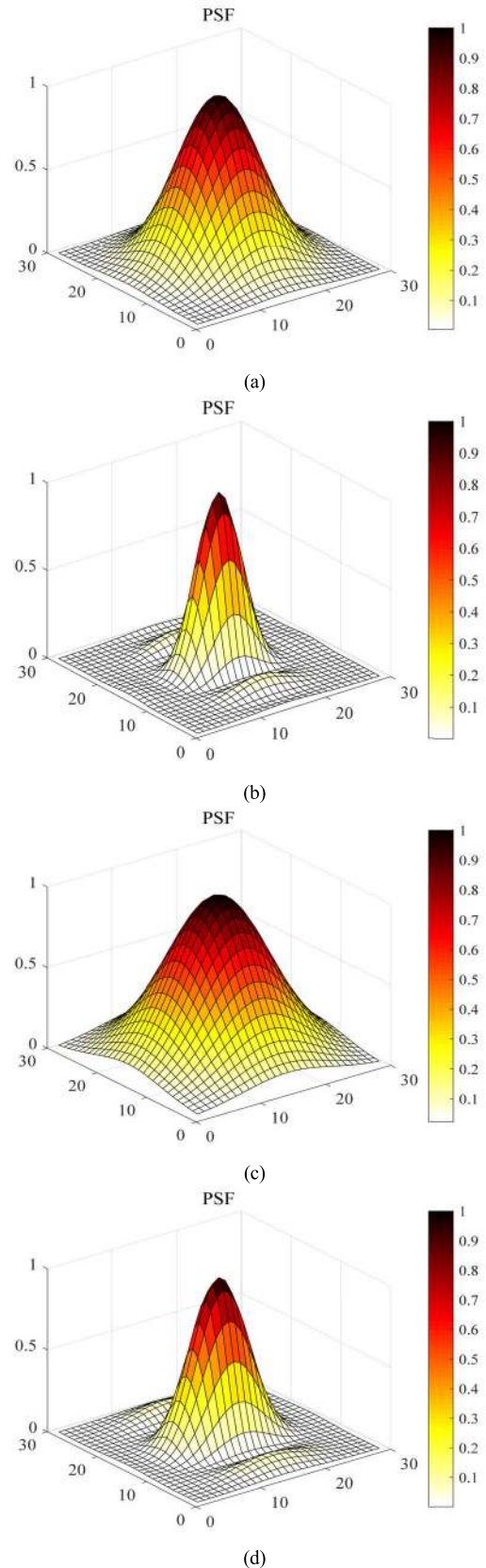
where  $\ln(\cdot)$  denotes the natural logarithmic operator. According to (21), the problem of maximizing a posterior can be transformed into minimizing the cost function of  $\mathbf{x}$  as:

$$J_G(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \frac{v_e}{v_x} \|\mathbf{x}\|_2^2 \quad (22)$$

Then  $\mathbf{x}$  is obtained from the following optimization problem:

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}} \left( \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_2^2 \right) \quad (23)$$

where  $\lambda = v_e/v_x$ ,  $\lambda > 0$  denotes the regularization parameter, and it controls the trade-off between the smoothed solution and data-fitting error. The equation (23) is exactly the



**FIGURE 4.** PSF Comparisons. (a) PSF of the prototype microphone array at 1000 Hz. (b) PSF of non-synchronous measurements at 1000 Hz. (c) PSF of prototype array at 800 Hz. (d) PSF of non-synchronous measurements at 800 Hz. Prototype aperture is 1 m, the synthetic aperture is 2.4 m.

Tikhonov regularization formula, or called  $L_2$  regularization. Without the Bayesian framework,  $\lambda$  has to be carefully tuned by conventional  $L_2$  regularization for optimal performance. Owing to the Bayesian inference,  $\lambda$  is inherently determined by the Signal-to-Noise Ratio (SNR). In general, the  $L_2$  regularization is solved by ridge regression or time-consuming convex optimization [26], [27], and it is robust to noise interference, but  $L_2$  norm prefers to smooth estimation of  $x$  and its spatial resolution still needs improving. This fact can be explained by Bayesian inference that the Gaussian independent prior related to  $L_2$  regularization is not a sparse distribution, since there is a long tail and slim body for sparse distribution (seen in Fig. 5).

Another commonly used sparse prior is the Laplace prior assigned to  $x$ : each component of  $x$  is approximated as a zero-mean Laplace distribution with unknown scale parameter  $b_x$  and all the elements of  $x$  are independent of each other. Then the posterior probability can be expressed as:

$$\begin{cases} \wp(x|y) \propto \wp(y|x) \wp(x) \\ -\ln[\wp(y|x) \wp(x)] = \frac{1}{2v_e} \left( \|y - Hx\|_2^2 + \frac{2v_e}{b_x} \|x\|_1 \right) \\ \quad + \frac{1}{2} \ln(2\pi v_e) + N \ln(2b_x) \end{cases} \quad (24)$$

The cost function under Laplace prior is defined as:

$$J_L(x) = \|y - Hx\|_2^2 + \frac{2v_e}{b_x} \|x\|_1 \quad (25)$$

Then  $x$  is obtained from the following optimization problem:

$$\hat{x} = \operatorname{argmin}_x \left( \|y - Hx\|_2^2 + \eta \|x\|_1 \right) \quad (26)$$

where  $\eta = 2v_e/b_x$ ,  $\eta > 0$  denotes the regularization parameter corresponding to  $L_1$  regularization and its function is similar to  $\lambda$  in  $L_2$  regularization in (23). The  $L_1$  regularization is commonly solved by the Least Absolute Shrinkage and Select Operator (LASSO) [28], [29]. Owing to the sparse prior, the  $L_1$  regularization can obtain better spatial resolution than  $L_2$  regularization does. But due to its non-differentiable at  $x = \mathbf{0}$ ,  $L_1$  regularization inevitably causes the deconvolution artifacts and cannot guarantee a fast convergence in strong noise interference.

In the inverse problem of acoustic localization in (23) and (26), the physical meaning of  $x$  is the power distribution on the scanning plane. Generally, the number of acoustic sources is much less than the number of scanning grids ( $K \ll N$ ). So that the power distribution of acoustic sources is sparsely distributing along the scanning plane. This sparsity is an essential constraint for solving the inverse problem of (14). From the perspective of the Bayesian framework, Probability Density Function (PDF) is the most intuitive representation of the sparsity of prior information. Different prior distributions will bring varying degrees of sparsity constraints. Moreover, the Bayesian framework reveals that the regularization parameters  $\lambda$  and  $\eta$  can be determined by the SNR.

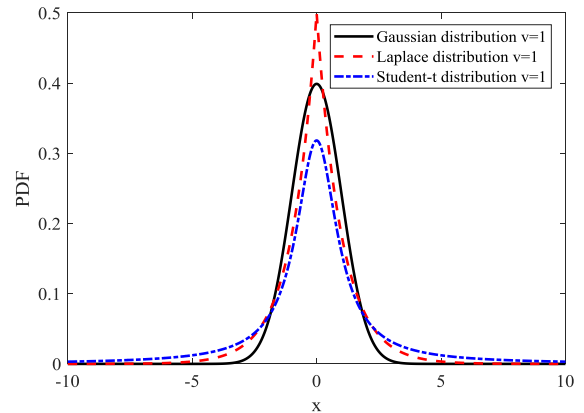


FIGURE 5. PDF comparison of gaussian distribution, laplace distribution, and student-t distribution.

In this paper, a Student-t distribution is introduced to balance the sparsity and computation complexity compared with the Gaussian distribution and Laplace distribution. As expressed in (27),  $x$  is one of the elements within  $x$ .  $\wp_1(x)$ ,  $\wp_2(x)$ , and  $\wp_3(x)$  are the PDFs of Gaussian distribution, Laplace distribution, and Student-t distribution, respectively. Assume the mean values of these distributions are 0, and the shapes of these distributions are controlled by scale parameter  $v$ , as shown in Fig. 5. Student-t distribution has the longest tail, and more sparsity than the Laplace distribution. In fact, the Gaussian distribution has less sparsity due to its short tail. Moreover, the background noise can also be modeled by Student-t distribution, which can be adaptively fit for Gaussian and non-Gaussian noise interference such as Cauchy or Laplace noise.

$$\begin{cases} \wp_1(x) = \frac{1}{\sqrt{2\pi}v} \exp^{-\frac{x^2}{2v}} \\ \wp_2(x) = \frac{1}{2v} \exp^{-\frac{|x|}{v}} \\ \wp_3(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{v\pi}\Gamma\left(\frac{v}{2}\right)} \left(1 + \frac{x^2}{v}\right)^{-\frac{v+1}{2}} \end{cases} \quad (27)$$

where  $\Gamma(\cdot)$  denotes the Gamma function. Combining the (27) and corresponding PDFs in Fig. 5, most of the high probability values concentrate on zero-mean, and very few non-zero values are distributing over the tail of PDF. In general, a sparse distribution should have a slim body and a long tail such as Student-t and Laplace distributions. According to Fig. 5, the dynamic range of Laplace distribution is wider than the Gaussian distribution and narrower than the Student-t. Meanwhile, the tail of Laplace distribution is longer than the tail of Gaussian distribution and shorter than the Student-t. Hence, the solution of  $L_1$  regularization is more sparse than the  $L_2$  regularization. But  $L_1$  regularization is more complex to calculate than the  $L_2$  regularization. Moreover, the Student-t distribution has a great potential to coordinate the solution sparsity and calculation complexity. Therefore, the following part will explore the advantages of Student-t distribution.

### B. PROPOSED SPARSITY-ENFORCING PRIOR

To enforce the solution sparsity, there are several kinds [3], [5], [15], [25] of priors such as Double Exponential, Generalized Gaussian, and some other mixture models with heavy tails. In this paper, the Student-t distribution is investigated to promote sparse solutions for two reasons. Firstly, as mentioned above, the tail of Student-t distribution is longer than Gaussian distribution and Laplace distribution; secondly, thanks to its Infinite Gaussian Scaled Mixture (IGSM) property, the Student-t can be expressed as a hierarchical Gaussian-Gamma framework as:

$$St(x|lv) = \int_0^\infty \mathcal{N}(x|0, 1/u) \mathcal{G}(u|1/\alpha, 1/\beta) du \quad (28)$$

where  $St(\cdot)$  and  $\mathcal{G}(\cdot)$  denote Student-t distribution and Gamma distribution, respectively.  $1/u$  is the variance of Gaussian distribution.  $1/\alpha$  is the shape parameter and  $1/\beta$  is the scale parameter of Gamma distribution.

Owing to the IGSM property in (28), the following hierarchical prior model is developed as:

$$\begin{cases} \wp(\mathbf{x}|\mathbf{V}_x) = \mathcal{N}(\mathbf{x}|\mathbf{0}, \mathbf{V}_x) \\ \wp(\mathbf{V}_x) = \prod_{j=1}^N \wp(V_{x_{j,j}}) = \prod_{j=1}^N \mathcal{IG}(V_{x_{j,j}}|\alpha_{x_0}, \beta_{x_0}) \\ \wp(\mathbf{e}|\mathbf{V}_e) = \mathcal{N}(\mathbf{e}|\mathbf{0}, \mathbf{V}_e) \\ \wp(\mathbf{V}_e) = \prod_{i=1}^N \wp(V_{e_{i,i}}) = \prod_{i=1}^N \mathcal{IG}(V_{e_{i,i}}|\alpha_{e_0}, \beta_{e_0}) \\ \mathcal{IG}(v|\alpha, \beta) = \beta^\alpha v^{-\alpha-1} \exp^{-\beta/v} / \Gamma(\alpha) \end{cases} \quad (29)$$

where  $\mathbf{V}_x$  and  $\mathbf{V}_e$  denote unknown covariance matrices of  $\mathbf{x}$  and  $\mathbf{e}$  respectively, and  $V_{x_{j,j}}$  is one of the diagonal items within  $\mathbf{V}_x$ , so does  $V_{e_{i,i}}$  to  $\mathbf{V}_e$ . And  $\mathcal{IG}(v|\alpha, \beta)$  denotes inverse Gamma distribution with shape parameter  $\alpha$  and scale parameter  $\beta$ . Compared with (20), the model uncertainty  $\mathbf{e}$  has been modeled by Student-t distribution, which is related to non-Gaussian noise interference such as Cauchy and Laplace noise. According to (29), the Bayesian hierarchical diagram is depicted in Fig.6 as:

Based on the proposed Bayesian hierarchical framework, the joint posterior of all the variables  $\mathbf{x}$ ,  $\mathbf{V}_x$  and  $\mathbf{V}_e$  is derived as:

$$\begin{aligned} \wp(\mathbf{x}, \mathbf{V}_x, \mathbf{V}_e|\mathbf{y}) \\ \propto \wp(\mathbf{y}|\mathbf{x}, \mathbf{V}_x, \mathbf{V}_e) \wp(\mathbf{x}|\mathbf{V}_x) \wp(\mathbf{V}_x|\alpha_{x_0}, \beta_{x_0}) \wp(\mathbf{V}_e|\alpha_{e_0}, \beta_{e_0}) \end{aligned} \quad (30)$$

where the noise interference is supposed to be independent of acoustic sources, thus  $\mathbf{V}_e$  is independent of  $\mathbf{x}$  and  $\mathbf{V}_x$ . And the JMAP inference is implemented to solve (30) in the following.

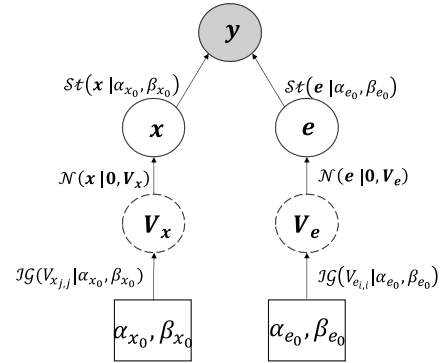


FIGURE 6. Bayesian hierarchical diagram:  $\mathbf{y}$  means beamforming map based on non-synchronous measurements,  $\mathbf{x}$  means ground-truth of acoustic source power, and  $\mathbf{e}$  means model uncertainty, including array noise interference and approximation error. Circle with padding: Observed data; Hollow solid circle: Unknown variables; Hollow dotted circle: Hidden variables; Square: Hyper-parameters.

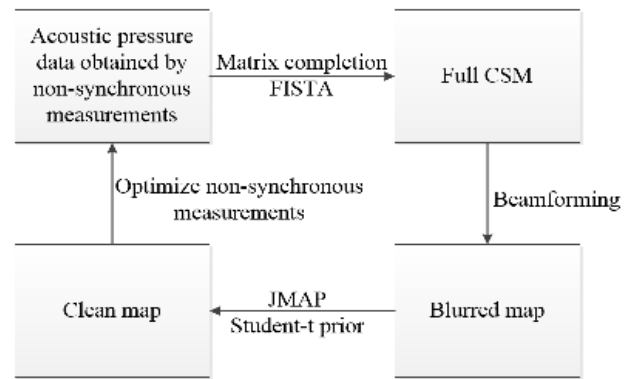


FIGURE 7. Flowchart of the proposed Bi-NAM method.

### C. JMAP INFERENCE

Inserting the (29) and (14) into (30), the JMAP can be transformed into minimization of the cost function as:

$$\begin{aligned} J_{JMAP}(\mathbf{x}, \mathbf{V}_x, \mathbf{V}_e) \\ = -\ln[p(\mathbf{x}, \mathbf{V}_x, \mathbf{V}_e, \mathbf{y})] \\ = \frac{1}{2} \sum_{i=1}^N \ln(V_{e_{i,i}}) \\ + \frac{1}{2} (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{V}_e^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}) \\ + \frac{1}{2} \sum_{j=1}^N \ln(V_{x_{j,j}}) + \frac{1}{2} \mathbf{x}^T \mathbf{V}_x^{-1} \mathbf{x} \\ + \sum_{j=1}^N \left[ (\alpha_{x_0} + 1) \ln(V_{x_{j,j}}) + \frac{\beta_{x_0}}{V_{x_{j,j}}} \right] \\ + \sum_{i=1}^N \left[ (\alpha_{e_0} + 1) \ln(V_{e_{i,i}}) + \frac{\beta_{e_0}}{V_{e_{i,i}}} \right] \end{aligned} \quad (31)$$



The estimation of all the unknown variables  $\mathbf{x}$ ,  $\mathbf{V}_x$ ,  $\mathbf{V}_e$  can be achieved by solving the alternate optimization as:

$$\begin{cases} \hat{\mathbf{x}} = \operatorname{argmin}_{(\mathbf{x})} J_{JMAP}(\mathbf{x}, \hat{\mathbf{V}}_x, \hat{\mathbf{V}}_e) \\ \hat{\mathbf{V}}_x = \operatorname{argmin}_{(\mathbf{V}_x)} J_{JMAP}(\hat{\mathbf{x}}, \mathbf{V}_x, \hat{\mathbf{V}}_e) \\ \hat{\mathbf{V}}_e = \operatorname{argmin}_{(\mathbf{V}_e)} J_{JMAP}(\hat{\mathbf{x}}, \hat{\mathbf{V}}_x, \mathbf{V}_e) \end{cases} \quad (32)$$

For this optimization problem, there are the theoretical solution for some unknown variables under the assumption that other variables are fixed. Firstly, when  $\mathbf{V}_x$  and  $\mathbf{V}_e$  are fixed, the cost function for  $\mathbf{x}$  is:

$$J_x(\mathbf{x}) = \frac{1}{2}(\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{V}_e^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x}) + \frac{1}{2}\mathbf{x}^T \mathbf{V}_x^{-1}\mathbf{x} \quad (33)$$

The cost function (33) can be solved by a gradient descent approach. Note that the initialization of  $\mathbf{x}^{(0)}$  can be the analytical solution by putting the derivative of (33) equal to zero. This initialization makes alternate iteration to converge much faster than that initialized by random process. Thus the initialization and iteration can be expressed as:

$$\begin{cases} \mathbf{x}^{(0)} = [\mathbf{H}^T \mathbf{V}_e^{-1} \mathbf{H} + \mathbf{V}_x^{-1}]^{-1} \\ \mathbf{x}^{[c+1]} = \mathbf{x}^{[c]} - \rho \frac{\partial J_x(\mathbf{x}^{[c]})}{\partial \mathbf{x}^{[c]}} \end{cases} \quad (34)$$

where  $\rho$  denotes the step size of gradient decent,  $c$  denotes the number of iterations. Then assume  $\mathbf{x}$  being fixed, the cost function for  $\mathbf{V}_e$  and  $\mathbf{V}_x$  can be alternatively expressed as:

$$\begin{cases} J_{V_x}(\mathbf{V}_x) = \frac{1}{2} \sum_{j=1}^N \ln(V_{x_{j,j}}) + \frac{1}{2} \mathbf{x}^T \mathbf{V}_x^{-1} \mathbf{x} \\ \quad + \sum_{j=1}^N \left[ (\alpha_{x_0} + 1) \ln(V_{x_{j,j}}) + \frac{\beta_{x_0}}{V_{x_{j,j}}} \right] \\ J_{V_e}(\mathbf{V}_e) = \frac{1}{2} \sum_{i=1}^N \ln(V_{e_{i,i}}) \\ \quad + \frac{1}{2} (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{V}_e^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}) \\ \quad + \sum_{i=1}^N \left[ (\alpha_{e_0} + 1) \ln(V_{e_{i,i}}) + \frac{\beta_{e_0}}{V_{e_{i,i}}} \right] \end{cases} \quad (35)$$

Similarly, putting the derivative of  $\mathbf{V}_x$  and  $\mathbf{V}_e$  equal to zero, the JMAP estimation of  $\mathbf{V}_x$  and  $\mathbf{V}_e$  can be expressed as:

$$\begin{cases} [V_{x_{j,j}}]_{JMAP} = \frac{\beta_{x_0} + \frac{x_j^2}{2}}{\alpha_{x_0} + \frac{3}{2}} \\ [V_{e_{i,i}}]_{JMAP} = \frac{\beta_{e_0} + \frac{1}{2}(y_i - \mathbf{H}_i \mathbf{x})^2}{\alpha_{e_0} + \frac{3}{2}} \end{cases} \quad (36)$$

where  $\mathbf{H}_i$  denotes the  $i$ -th row of matrix  $\mathbf{H}$ . Here  $\alpha_{x_0}$ ,  $\beta_{x_0}$ ,  $\alpha_{e_0}$  and  $\beta_{e_0}$  are initialized by a random number between 0 and 1. These alternate iterations are summarized by the Bayesian JMAP Inference Algorithm in Table 1:

where  $\varepsilon$  denotes a small value to stop the alternate iteration. Bayesian JMAP inference finally achieves reasonable estimations of object  $\mathbf{x}$  and hidden variables  $\mathbf{V}_x$ ,  $\mathbf{V}_e$ .

TABLE 1. Bayesian Jmap inference algorithm.

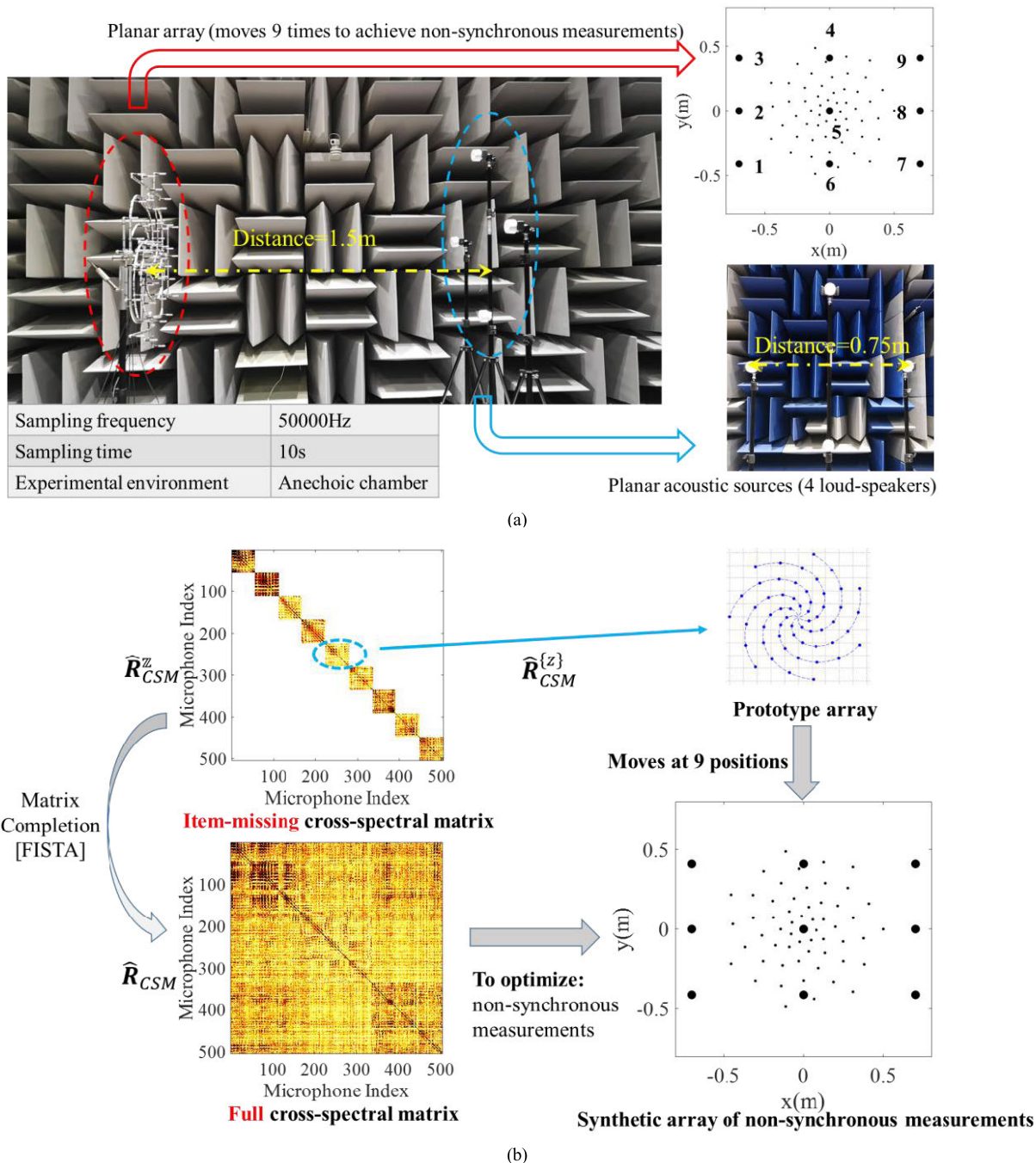
Bayesian JMAP inference algorithm	
1:	<b>Initialization</b>
2:	$\alpha_{x_0}, \beta_{x_0}, \alpha_{e_0}, \beta_{e_0} \rightarrow$ random between 0 and 1.
3:	$\mathbf{V}_x^{(0)} = \operatorname{Diag} \left[ \frac{\beta_{x_0}}{\alpha_{x_0} - 1} \right]$
4:	$\mathbf{V}_e^{(0)} = \operatorname{Diag} \left[ \frac{\beta_{e_0}}{\alpha_{e_0} - 1} \right]$
5:	$\mathbf{x}^{(0)} = (\mathbf{H}^T [\mathbf{V}_e^{(0)}]^{-1} \mathbf{H} + \mathbf{V}_x^{(0)})^{-1} \mathbf{H}^T [\mathbf{V}_e^{(0)}]^{-1} \mathbf{y}$
6:	<b>Alternate Iteration</b>
7:	For $c=1:C$ ( $C$ denotes the maximum iteration)
8:	$\beta_{x_j}^{(c)} = \beta_{x_0} + \frac{x_j^{(c)2}}{2}$
9:	$\alpha_{x_j}^{(c)} = \alpha_{x_0} + \frac{3}{2}$
10:	$[V_{x_{j,j}}]_{JMAP}^{(c)} = \frac{\beta_{x_j}^{(c)}}{\alpha_{x_j}^{(c)}}$
11:	$\beta_{e_i}^{(c)} = \beta_{e_0} + \frac{1}{2}(y_i - \mathbf{H}_i \mathbf{x}^{(c)})^2$
12:	$\alpha_{e_i}^{(c)} = \alpha_{e_0} + \frac{3}{2}$
13:	$[V_{e_{i,i}}]_{JMAP}^{(c)} = \frac{\beta_{e_i}^{(c)}}{\alpha_{e_i}^{(c)}}$
14:	$\frac{\partial J_x(\mathbf{x}^{(c)})}{\partial \mathbf{x}^{(c)}} = [\mathbf{V}_x^{(c)}]^{-1} \mathbf{x}^{(c)} - \mathbf{H}^T [\mathbf{V}_e^{(c)}]^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}^{(c)})$
15:	$\mathbf{x}^{[c+1]} = \mathbf{x}^{[c]} - \rho \frac{\partial J_x(\mathbf{x}^{[c]})}{\partial \mathbf{x}^{[c]}}$
16:	If $\ \mathbf{x}^{[c+1]} - \mathbf{x}^{[c]}\  \leq \varepsilon$ , then break;
17:	End For
18:	Output: $\mathbf{x}^{[c+1]}, \mathbf{V}_x^{[c+1]}, \mathbf{V}_e^{[c+1]}$

The whole framework is shown in Fig. 7. The proposed Bi-NAM via the JMAP algorithm is implemented to obtain high-resolution imaging at low frequencies. Here the clean map from our proposed method can be feedback to optimize the non-synchronous measurements. But this point will not be discussed due to limited pages.

#### IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

##### A. EXPERIMENT SETTING

In this section, experiments are carried out to validate the proposed Bi-NAM method. In order to remove the effects

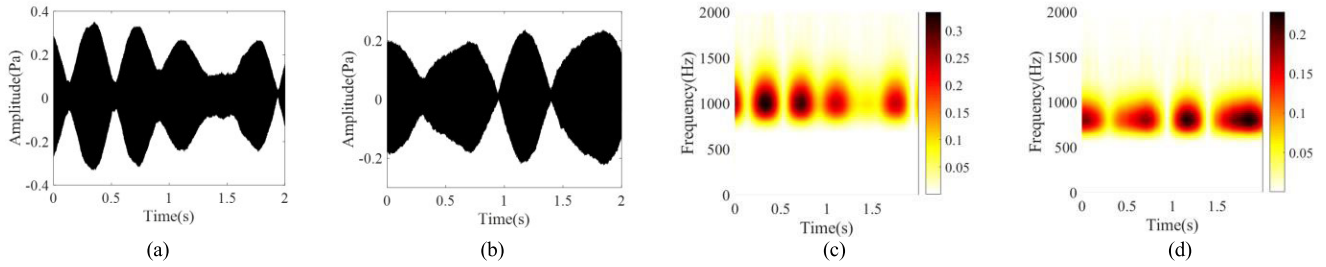


**FIGURE 8.** Experiment setup: (a) Planar spiral array with 56 channels and acoustic monopole sources generated by 4 loud-speakers. (b) Non-synchronous measurements by sequentially moving prototype array at 9 positions, and procedure from item-missing CSM of prototype array to full CSM of the synthetic array.

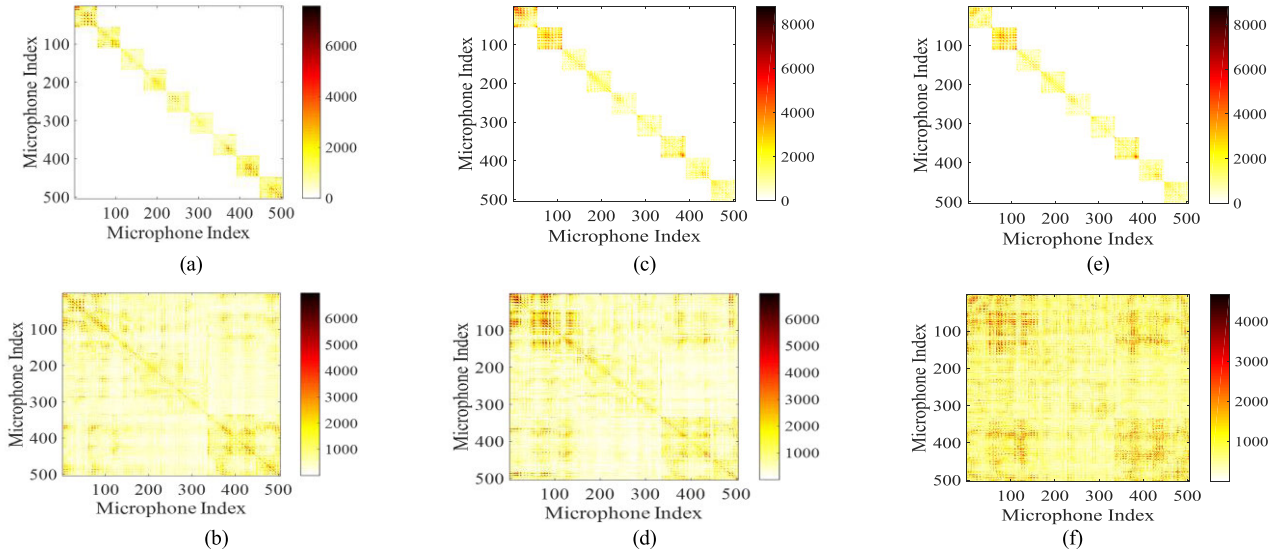
of uncontrollable factors, the experiments are performed in an anechoic chamber, as shown in Fig. 8(a). The distance between planar acoustic sources and the planar array is 1.5 m, and the distance between diagonal loud-speakers is about 0.75 m. The prototype array is moved sequentially 9 times during non-synchronous measurements, and array centre positions at each move are shown in Fig. 8(b) with coordinates  $(-0.41, -0.7), (-0.41, 0), (-0.41, 0.7), (0, 0.7),$

$(0, 0), (0, -0.7), (0.41, -0.7), (0.41, 0), (0.41, 0.7)$ . It is noted that the movement order of array does not substantially affect the imaging results. However, it is a convenient choice to move in sequence according to the convenience in the experiment.

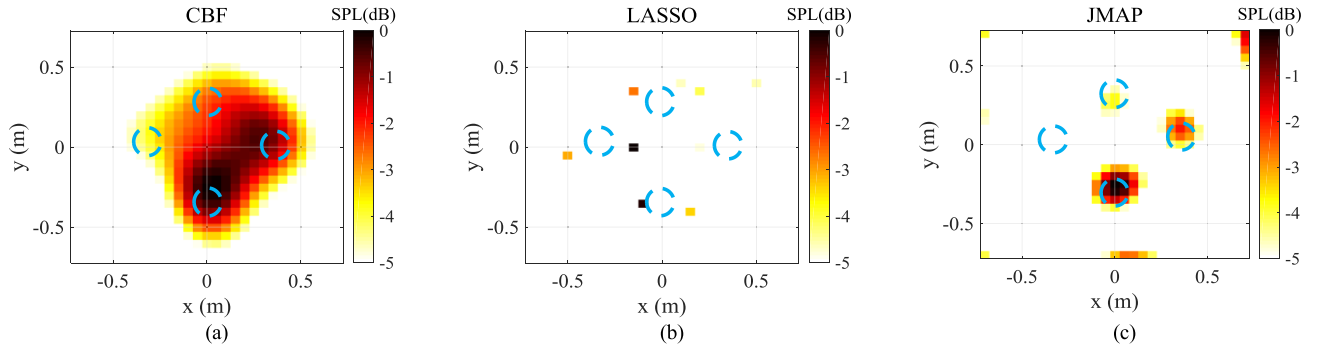
In Fig. 9, narrow-band signals with center frequencies at 1000 Hz and 800 Hz are played respectively by 4 loud-speakers to simulate acoustic monopole sources. At each



**FIGURE 9.** Experimental data: (a) Acoustic signals with narrow-band 400 Hz and center frequency 1000 Hz; (b) Acoustic signals with narrow-band 400 Hz and center frequency 800 Hz; (c) Wavelet transform of (a); (d) Wavelet transform of (b).



**FIGURE 10.** CSM completion (a) Item-missing CSM at 1000 Hz; (b) Full CSM at 1000 Hz; (c) Item-missing CSM at 800 Hz; (d) Full CSM at 800 Hz. (e) Difference between (a) and (c); (f) Difference between (b) and (d).

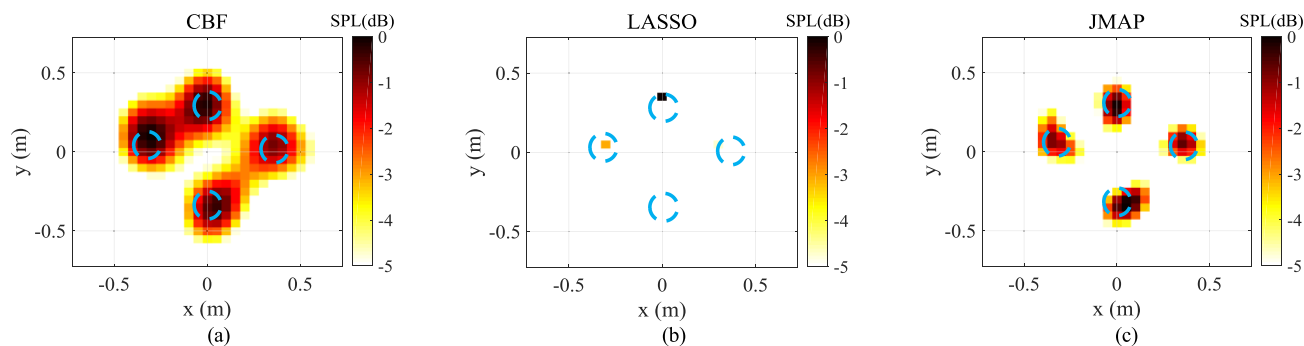


**FIGURE 11.** Real-data results of 4 monopole sources with narrow-band 400 Hz, center frequency 800 Hz, without noise interference in acoustic anechoic chamber obtained by (a) CBF, (b) LASSO, (c) Proposed Bi-NAM based on JMAP. Blue dashed circles stand for the positions of four loudspeakers.

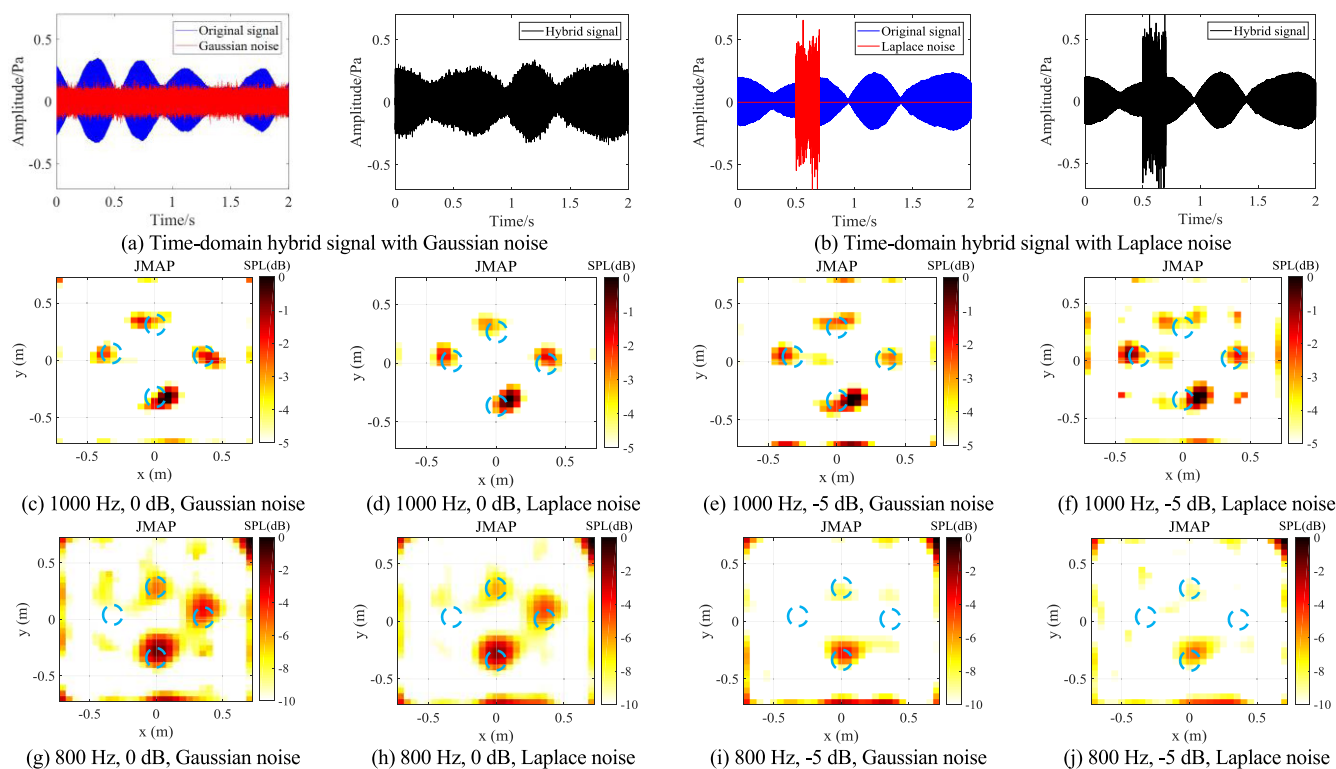
position of prototype array, the signals are sampled for 10 seconds with a sampling frequency of 50000 Hz. Then the measured data are processed by the CBF, LASSO, and proposed Bi-MAN based on JMAP at 1000 Hz and 800 Hz, respectively, in Fig. 10-12. The hybrid data with the Gaussian and Laplace noise interferences are also tested in Fig. 13. In reality, the difficulties of source setup are that the used loud-speakers are not identical sources, and they cannot synchronously emit signals. Moreover, the emitted sounds are not strictly stationary and stable.

**B. RESULTS AND DISCUSSIONS**

Firstly, the CSM completion is shown in Fig. 10. Colorbar represents the relative sound power level. Without noise interference, the differences between full CSMs from 1000 Hz to 800Hz cannot be neglectable. This is because that the tested signals are not strictly stationary and synchronous. And these differences might lead to the localization result of the proposed Bi-NAM at 800 Hz in Fig. 11(c) is not better than that at 1000 Hz in Fig. 12(c).



**FIGURE 12.** Real-data results of 4 monopole sources with narrow-band 400 Hz, center frequency 1000 Hz, without noise interference in acoustic anechoic chamber obtained by (a) CBF, (b) LASSO, (c) Proposed Bi-NAM based on JMAP. Blue dashed circles denote the positions of four loudspeakers.



**FIGURE 13.** Hybrid-data results by proposed Bi-NAM: (a) Real data + Simulated Gaussian noise; (b) Real data + Simulated Laplace noise; (c-f) 1000 Hz, at SNR 0 dB and -5 dB respectively; (g-j) 800 Hz, at SNR 0 dB and -5 dB respectively. Blue dashed circles stand for the positions of four loudspeakers.

In Fig. 11, it shows the acoustic imaging results at 800 Hz without noise interference. The colormap dynamic range is 5 dB. The blue bold dotted line stands for the positions of four acoustic sources. The CBF and LASSO can hardly demonstrate any valid information about acoustic sources. The proposed Bi-NAM method can localize each source accurately and break the limitation caused by the PSF from the blurred CBF result. In Fig. 11(c), the intensity of the left acoustic source is much lower than those of the other 3 sources, so that it is invisible in the 5 dB dynamic range. The reason is that the acoustic intensity is unstable and nonstationary due to loudspeaker quality.

In Fig. 12, it shows the acoustic imaging results at 1000Hz without noise interference. The CBF is not able to distinguish clearly 4 monopoles due to the fat and flat PSF at 1000Hz in Fig. 4. The LASSO can provide a high-resolution result, but the reconstructed distributions are too sparse that the over-sparsity results will lose the continuous distribution. When the dynamic range is increased, the massive artifacts will inevitably cause frequent false-alarm detection. It is difficult to tell whether these points are sources or sidelobes, and not easy to confirm the number of the monopole sources. The proposed Bi-NAM method can achieve high-resolution imaging, and adjacent sources can be separated clearly. The

**TABLE 2.** Calculation time of LASSO and proposed Bi-NAM.

	800Hz	1000Hz
LASSO	3.18 s	1.82 s
Bi-NAM	4.68 s	3.09 s

**TABLE 3.** Method comparisons.

PSF	800Hz		1000Hz	
	Prototype quite flat, quite fat	NSM nearly sharp, nearly slim	Prototype flat & fat	NSM sharp & slim
CBF	failed	failed	failed	over-blurred
LASSO	failed	too sparse, failed	failed	too sparse
Bi-NAM	OK, but very low- resolution	75% success, high- resolution	OK, but low- resolution	100% success, high- resolution

Note: **Prototype:** prototype array; **NSM:** non-synchronous measurements of prototype array; **75% success:** 3 of 4 monopoles are localized successfully.

distribution of acoustic sources can also be vividly displayed. In addition to the acoustic imaging results, the calculation efficiency is also a significant indicator to evaluate the performance of the algorithm. In this paper, all programs are run in MATLAB R2017b software on Intel(R) Core(TM) i7-7700HQ CPU @ 2.80GHz system with 16 GB-RAM. The iteration termination parameter is set as  $\epsilon = 0.1$ , and the calculation time of the LASSO and the proposed Bi-NAM based on JMAP is given in Table 2, in which we can conclude that the proposed Bi-NAM method doesn't bring extra computation burden.

In Fig.13(a) and (b), the hybrid data are generated by adding the Gaussian and Laplace noise interferences to the real data in Fig.9. The SNR are set 0 dB and -5 dB, respectively. Then the proposed Bi-NAM method can achieve reasonable and consistent localization results at 1000 Hz in Fig.13(c), (d), (e) and (f), as well as at 800 Hz in Fig.13(g), (h), (i) and (j). But due to the non-stationarity of four loudspeakers, it would worsen the phase continuity and CSM completion of the non-synchronous measurements, and bring the apparent sidelobes and artifacts around corners to the localization results.

In brief, the proposed Bi-NAM approach has been confirmed to achieve high-resolution acoustic imaging at 1000Hz and 800Hz at low SNR, respectively, even under the Laplace noise interference. The comparisons of the PSF and aforementioned methods are summarized in Table 3.

## V. CONCLUSION

This paper presents that the PSF of microphone array is a critical factor that causes quite blurred imaging for acoustic localization and visualization at low-frequencies. In fact, the PSF has become an insightful way from array topology

to localization resolution. In order to obtain high-resolution imaging at low-frequency, what is worth breaking is not only to improve the rationality of sparse constraints in the inverse algorithm, but also to reduce the PSF influence by optimizing the prototype array, or making the synthetic array of non-synchronous measurements.

The proposed Bi-NAM method has taken the above two aspects into account comprehensively. First, the non-synchronous measurements are implemented to extend the array synthetic aperture and enhance the microphone density. In this way, the PSF of the prototype array, which is flat and fat, has been successfully refined as the synthetic PSF of the non-synchronous measurements, which is sharp and slim. Next, the Bayesian JMAP inference is developed to solve the very ill-posed inverse problem caused by non-synchronous measures, in which, the item-missing CSM needs element-completion and necessary approximation. Student-t prior is employed to promote sparsity for high resolution and low computation. All the variables can be alternatively estimated via a Bayesian hierarchical framework. Finally, localization results of experiment data at 800 Hz and 1000 Hz validate the proposed Bi-NAM method. Our proposal is capable of building up a refined PSF and inferring high-resolution acoustic imaging at both low-frequency and low SNR, even under Laplace noise interference.

In the future, the proposed Bi-NAM method can be extended to 3D acoustic source localization with high resolution. The optimal topological design of non-synchronous measurements is worthy of studying for easier implementation and better performance. In the fast calculation, it is an interesting topic for the Bayesian JMAP algorithm combined with the shift-invariant convolution model.

## ACKNOWLEDGMENT

The authors would like to thank the Zhejiang Institute of Metrology (ZJIM) and Zhejiang SHANGFENG Special Blower Industrial Co., Ltd., for offering the acoustic anechoic-chamber and some of experimental data, respectively.

## REFERENCES

- [1] P. Chiariotti, M. Martarelli, and P. Castellini, "Acoustic beamforming for noise source localization—Reviews, methodology and applications," *Mech. Syst. Signal Process.*, vol. 120, pp. 422–448, Apr. 2019.
- [2] J. C. Chen, K. Yao, and R. E. Hudson, "Source localization and beamforming," *IEEE Signal Process. Mag.*, vol. 19, no. 2, pp. 30–39, Mar. 2002.
- [3] Q. Leclère, A. Pereira, C. Bailly, J. Antoni, and C. Picard, "A unified formalism for acoustic imaging based on microphone array measurements," *Int. J. Aeroacoustics*, vol. 16, nos. 4–5, pp. 431–456, Jul. 2017.
- [4] R. Merino-Martínez, "A review of acoustic imaging methods using phased microphone arrays," *CEAS Aeronaut. J.*, vol. 10, no. 1, pp. 197–230, 2019.
- [5] N. Chu, "Bayesian approach in acoustic source localization and imaging," Ph.D. dissertation, Lab. Signaux Syst., Univ. Paris-Sud (XI), Paris, France, 2013.
- [6] N. Chu, N. Gac, P. José, and A. Mohammad-Djafari, "2D convolution model using (in)variant kernels for fast acoustic imaging," in *Proc. Conf. Berlin Beamforming Conf.*, 2014, p. 15.

[7] J. Li, P. Stoica, and Z. Wang, "On robust capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.

[8] P. Sijtsma, R. Merino-Martinez, A. M. Malgoezar, and M. Snellen, "High-resolution CLEAN-SC: Theory and experimental validation," *Int. J. Aeroacoustics*, vol. 16, nos. 4–5, pp. 274–298, Jul. 2017.

[9] R. Dougherty, "Extensions of DAMAS and benefits and limitations of deconvolution in beamforming," in *Proc. 11th AIAA/CEAS Aeroacoustics Conf.*, May 2005, p. 2961.

[10] T. F. Brooks and W. M. Humphreys, "A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays," *J. Sound Vibrat.*, vol. 294, nos. 4–5, pp. 856–879, Jul. 2006.

[11] P. Simard and J. Antoni, "Acoustic source identification: Experimenting the  $\ell_1$  minimization approach," *Appl. Acoust.*, vol. 74, no. 7, pp. 974–986, 2013.

[12] G. F. Edelmann and C. F. Gaumont, "Beamforming using compressive sensing," *J. Acoust. Soc. Amer.*, vol. 130, no. 4, pp. EL232–EL237, Oct. 2011.

[13] A. Xenaki, P. Gerstoft, and K. Mosegaard, "Compressive beamforming," *J. Acoust. Soc. Amer.*, vol. 136, no. 1, pp. 71–260, Jul. 2014.

[14] M. Q. Pham, B. Oudompheng, J. I. Mars, and B. J. S. P. Nicolas, "A Noise-Robust Method with Smoothed  $\ell_1/\ell_2$  Regularization for Sparse Moving-Source Mapping," *Signal Process.*, vol. 135, pp. 96–106, Jun. 2017.

[15] N. Chu, A. Mohammad-Djafari, and J. Picheral, "Robust Bayesian super-resolution approach via sparsity enforcing a priori for near-field aeroacoustic source imaging," *J. Sound Vibrat.*, vol. 332, no. 18, pp. 4369–4389, Sep. 2013.

[16] Y. Yang, Z. Chu, and G. Ping, "Alternating direction method of multipliers for weighted atomic norm minimization in two-dimensional grid-free compressive beamforming," *J. Acoust. Soc. Amer.*, vol. 144, no. 5, pp. EL361–EL366, Nov. 2018.

[17] Z. Chu, C. Chen, Y. Yang, and L. Shen, "Two-dimensional total variation norm constrained deconvolution beamforming algorithm for acoustic source identification," *IEEE Access*, vol. 6, pp. 43743–43748, 2018.

[18] G. Ping, Z. Chu, Y. Yang, and X. Chen, "Iteratively reweighted spherical equivalent source method for acoustic source identification," *IEEE Access*, vol. 7, pp. 51513–51521, 2019.

[19] C.-X. Bi, Y. Liu, Y.-B. Zhang, and L. Xu, "Sound field reconstruction using inverse boundary element method and sparse regularization," *J. Acoust. Soc. Amer.*, vol. 145, no. 5, pp. 3154–3162, May 2019.

[20] A. Pereira, J. Antoni, and Q. Leclère, "Empirical Bayesian regularization of the inverse acoustic problem," *Appl. Acoust.*, vol. 97, pp. 11–29, Oct. 2015.

[21] N. Chu, J. Picheral, A. M. Djafari, and N. Gac, "A robust super-resolution approach with sparsity constraint in acoustic imaging," *Appl. Acoust.*, vol. 76, pp. 197–208, Feb. 2014.

[22] M. Gockenbach, *Linear Inverse Problems and Tikhonov Regularization*. Miami, FL USA: The Mathematical Association of America Press, 2016.

[23] A. Buccini, M. Donatelli, and L. Reichel, "Iterated tikhonov regularization with a general penalty term," *Numer. Linear Algebra with Appl.*, vol. 24, no. 4, p. e2089, Aug. 2017.

[24] L. Yu, J. Antoni, Q. Leclere, and W. Jiang, "Acoustical source reconstruction from non-synchronous sequential measurements by fast iterative shrinkage thresholding algorithm," *J. Sound Vibrat.*, vol. 408, pp. 351–367, Nov. 2017.

[25] A. Mohammad-Djafari and M. Dumitru, "Bayesian sparse solutions to linear inverse problems with non-stationary noise with Student-t priors," *Digit. Signal Process.*, vol. 47, pp. 128–156, Dec. 2015.

[26] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[27] M. Grant, S. Boyd, and Y. Ye. (2009). *CVX: MATLAB Software for Disciplined Convex Programming*. CVX Research Inc. [Online]. Available: <http://cvxr.com/cvx/>

[28] T. T. Wu and K. Lange, "Coordinate descent algorithms for lasso penalized regression," *Ann. Appl. Statist.*, vol. 2, no. 1, pp. 224–244, Mar. 2008.

[29] R. Tibshirani, "Regression shrinkage and selection via the lasso: A retrospective," *J. Roy. Stat. Soc., B (Stat. Methodol.)*, vol. 73, no. 3, pp. 273–282, Jun. 2011.



Polytechnique Fédérale de Lausanne, Switzerland.

**NING CHU** received the bachelor's degree in information engineering from the National University of Defense Technology, in 2006, and the master's and Ph.D. degrees in automatic, signal image processing from the University of Paris Sud, France, in 2010 and 2014, respectively. He is currently a Lecturer with Zhejiang University. His research interests are acoustic source imaging, and Bayesian inference in fault diagnosis. He received the positions of scientific collaborator in the École



**YUE NING** received the bachelor's degree in process equipment and controlling engineering from Zhejiang University, Hangzhou, China, in 2019, where he is currently pursuing the master's degree. His research interests are array signal processing and Bayesian inference.



**LIANG YU** received the bachelor's degree in communication engineering from the Hubei University of Technology, Wuhan, China, in 2007, the double M.S. degree in communication and information system from Shanghai University, and Waseda University, Fukuoka/Tokyo, Japan, in 2011, and the Ph.D. degree in acoustics from the INSA de Lyon, Lyon, France, in 2015. He is currently a Research Assistant with Shanghai Jiao Tong University. His research interests are acoustics and signal processing.



**QIAN HUANG** received the bachelor's degree from the College of Energy Engineering, Zhejiang University, Hangzhou, China, in 2018, where he is currently pursuing the master's degree. His current research interest includes acoustic source localization.



**DAZHUAN WU** received the bachelor's degree in machinery and equipment of chemical engineering and the Ph.D. degree in chemical process equipment from Zhejiang University, Hangzhou, China, in 1999 and 2004, respectively. He was a Visiting Scholar from Kyoto University, Japan. He is currently a Professor with Zhejiang University. His research interests are flow-induced vibration and noise control, Pumps, and fans and propulsion technology.

...