



**HAL**  
open science

# Bayesian space-frequency separation of wide-band sound sources by a hierarchical approach

Erliang Zhang, Jérôme Antoni, Bin Dong, Hichem Snoussi

► **To cite this version:**

Erliang Zhang, Jérôme Antoni, Bin Dong, Hichem Snoussi. Bayesian space-frequency separation of wide-band sound sources by a hierarchical approach. *Journal of the Acoustical Society of America*, 2012, 132 (5), pp.3240-3250. 10.1121/1.4754530 . hal-01018733

**HAL Id: hal-01018733**

**<https://hal.science/hal-01018733v1>**

Submitted on 3 Sep 2021

**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.

# Bayesian space-frequency separation of wide-band sound sources by a hierarchical approach

Erliang Zhang

Laboratoire Roberval, UMR 7337, Université de Technologie de Compiègne, 60205 Compiègne cédex, France

Jérôme Antoni<sup>a)</sup> and Bin Dong

Université de Lyon, LVA, INSA-Lyon, 69621 Lyon cédex, France

Hichem Snoussi

Institut Charles Delaunay, UMR 6279, Université de Technologie de Troyes, 10000 Troyes cédex, France

(Received 8 January 2012; revised 13 August 2012; accepted 28 August 2012)

This paper proposes an efficient solution to the separation of uncorrelated wide-band sound sources which overlap each other in both space and frequency domains. The space-frequency separation is solved in a hierarchical way by (1) expanding the sound sources onto a set of spatial basis functions whose coefficients become the unknowns of the problem (backpropagation step) and (2) blindly demixing the coefficients of the spatial basis into uncorrelated components relating to sources of distinct physical origins (separation step). The backpropagation and separation steps are both investigated from a Bayesian perspective. In particular, Markov Chain Monte Carlo sampling is advocated to obtain Bayesian estimates of the separated sources. Separation is guaranteed for sound sources having different power spectra and sufficiently smooth spatial modes with respect to frequency. The validity and efficiency of the proposed separation procedure are demonstrated on laboratory experiments.

© 2012 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4754530>]

PACS number(s): 43.60.Sx, 43.60.Pt, 43.60.Jn, 43.35.Sx [ZHM]

Pages: 3240–3250

## I. INTRODUCTION

The localization and quantification of sound sources play an important role in noise control applications related to acoustic engineering. Sound source quantification has been investigated by several methods such as near-field acoustic holography,<sup>1</sup> Helmholtz equation least squares<sup>2</sup> and inverse numerical acoustics.<sup>3,4</sup> More than the quantification of the total source field, sound source separation aims to discriminate all source distributions from the measured sound field by decomposing it into contributions stemming from distinct physical origins. Much attention has been recently paid to visualize and analyze relative contributions from mutually incoherent sources in the near-field. An early work<sup>5</sup> reports that by placing a number of reference microphones next to the apparent sources the individual contributions from the sources may be estimated from cross-correlation measurements; yet this approach requires prior knowledge of the locations of individual sources. One attempt reported by Nam<sup>6</sup> to avoid specifying the prior locations of sources is to use the pressure signals at source positions estimated by the backward prediction of near-field acoustic holography as coherent signals to the sources. A multi-reference cross-spectral analysis approach is proposed to decompose the measured sound field into a number of partial fields owing to the reference sensors put close to sources, and these partial fields are projected back to any reconstruction surface. It has been addressed in a series of work, with source nonstationarity in the scan-based measurement,<sup>7</sup> with optimally located virtual

references,<sup>8</sup> in the presence of reference signals corrupted by noise and source level variation.<sup>9</sup> Nam and Kim<sup>10</sup> further propose an extended partial field decomposition algorithm that does not require placing reference sensors close to individual sources, where calculation is used rather than measurement technique to decompose a hologram image into images of incoherent sources. However, in the acoustic literature, the separation of sound sources has rarely been addressed in an unsupervised way, that is to find the hidden mixture of sound sources from the unlabeled measurements (without reference). Such an objective has recently motivated considerable research efforts in the signal processing community—see, e.g., Refs. 11–13—where the blind separation of mixtures of sources has been shown to be possible based on the sole assumption on the statistics of the sources, e.g., their mutual independence, their diversity in time or in frequency domain, etc. Regarding separation approaches, the Bayesian method has been intensively adopted (see, for instance, Refs. 13–15) as it provides an ideal framework to incorporate any knowledge available before the experiment in order to update the probability distribution that fully describes the unknown parameters of the problem; it also embodies an internal mechanism to regularize ill-posed problems.<sup>13,16</sup> The available information (even assumption) is incorporated in the form of probability density functions for instance through the use of the maximum entropy principle; the probability represents the current state of knowledge on a parameter based on the available information in the Bayesian context.<sup>17</sup>

Unfortunately, all the blind separation approaches existing in the signal community are not directly transposable to the separation of sound sources: one major difference is that sound sources are signals of both time (or frequency) and

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: [jerome.antoni@insa-lyon.fr](mailto:jerome.antoni@insa-lyon.fr)

space, which yields a huge amount of parameters to be estimated in the mixture and thus makes the problem very underdetermined; another difficulty is that acoustic separation implicitly implies inversion of the wave equation, a problem which is known to be severely ill-posed. In order to circumvent these difficulties, the present work proposes an original parametric formulation for sound source separation in both space and frequency. Not only does it improve the conditioning of the problem by reducing a huge number of parameters to a tractable amount, but it also lessens the separation effort by proceeding in a hierarchical way: first the sources to be separated are expanded onto a spatial basis whose coefficients become the unknowns of the problem (backpropagation step); second, the coefficients pertaining to each source (primitives of the sound field) are blindly estimated (blind separation step). Both steps are investigated in an optimal way from a Bayesian perspective to reduce the ill-conditionness, treat the measurement noise and modeling error, and enhance the spatial resolution of reconstructed sources. Eventually, the diversity of the source power spectra as well as their spatial property are exploited to blindly estimate the unknown variables in the mixture, which is solved by implementing Markov Chain Monte Carlo (MCMC).

The rest of the paper is organized as follows. Section II derives the formulation for space-frequency sound source separation, a hierarchical framework including backpropagation and separation. Section III presents the Bayesian reconstruction of the sound field with robust regularization and basis function. Section IV illustrates the application of a Bayesian inference with MCMC to estimate source parameters. Section V validates the proposed separation methodology on laboratory examples. The conclusion is drawn in Sec. VI.

## II. FORMULATION OF SPACE-FREQUENCY SOUND SOURCE SEPARATION

Sound sources radiate an acoustic field which can be computed through an analytical (Green's function) or an acoustic boundary element model: this is called the forward problem. The acoustic propagation from a source distribution

to a microphone array is depicted in Fig. 1. The recorded temporal pressure signals at the microphones are analyzed by using the Short Time Fourier Transform (STFT), as described hereafter. The total observation time of measurements is divided into  $n_\omega$  overlapped intervals of identical length. The intervals are  $j$ -indexed and serve to generate a set of space-frequency complex vectors by Fourier transforming the data over their durations. These complex vectors are usually referred to as frequency-domain snapshots.<sup>18</sup>

In linear acoustics, the measured pressure can be considered as the sum of the pressures contributed by each sound source plus measurement noise. Specifically, the pressure signal  $p(\mathbf{r}_j; \omega; \varpi)$  measured by the  $j$ th microphone at position  $\mathbf{r}_j$  reads

$$p(\mathbf{r}_j; \omega; \varpi) = \sum_{i=1}^n p_{0,i}(\mathbf{r}_j; \omega; \varpi) + \nu(\mathbf{r}_j; \omega; \varpi), \quad (1)$$

where  $p_{0,i}(\mathbf{r}_j; \omega; \varpi)$  is the sound field radiated by the  $i$ th sound source,  $\nu(\mathbf{r}_j; \omega; \varpi)$  is an output error term which basically accounts both for instrumentation, environment noise, and modeling errors,  $n$  denotes the underlying number of sources,  $\omega$  is the angular frequency variable, and  $\varpi$  labels the frequency-domain snapshot. Under random sound excitation, the  $\varpi$  indexed variables are seen as random draws of snapshots. The contribution  $p_{0,i}(\mathbf{r}_j; \omega; \varpi)$  of the  $i$ th source is obtained by the spatial integration of the source distribution  $s_i(\mathbf{r}'; \omega; \varpi)$  on the source domain ( $\Gamma$ ) with the Green's function  $g(\mathbf{r}_j; \mathbf{r}'; \omega)$ ,

$$p_{0,i}(\mathbf{r}_j; \omega; \varpi) = \int_{\Gamma} s_i(\mathbf{r}'; \omega; \varpi) g(\mathbf{r}_j; \mathbf{r}'; \omega) d\Gamma(\mathbf{r}'). \quad (2)$$

In the following, the backpropagation surface—on which the sources are to be separated—will be assumed to coincide with the source domain ( $\Gamma$ ), although there will be no technical difficulty for this assumption to be relaxed. Furthermore, the backpropagation surface may have any topology and the array any geometry—not necessarily parallel planes as depicted for simplicity in Fig. 1.

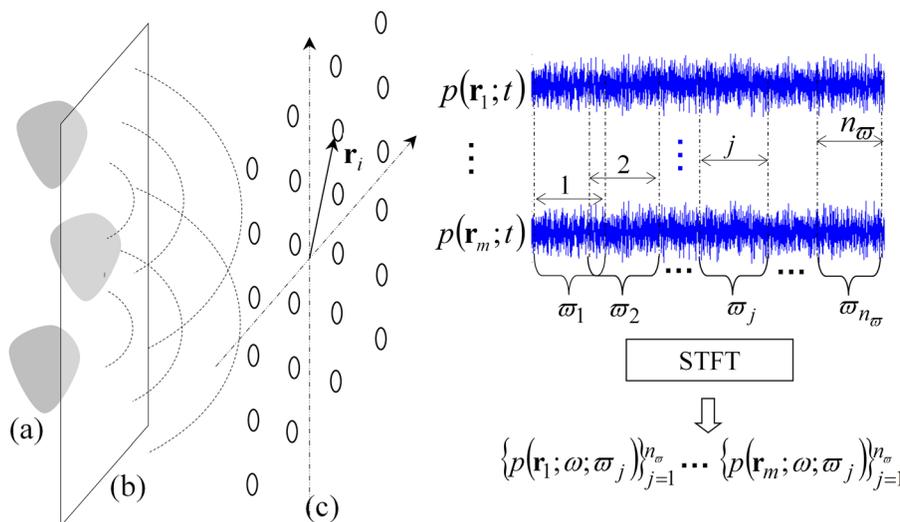


FIG. 1. (Color online) Forward problem [(a): acoustic origin, (b) backpropagation plane, (c) microphone array] and measurement data processing.

In order to satisfy the condition of separability, the sound sources are assumed to possess the following properties.

(A1) Assumption (A1) is mutual decorrelation of sources and consequently of the partial sound fields they radiate,

$$\mathbb{E}_{\varpi}(p_{0,i}(\mathbf{r}; \omega; \varpi)p_{0,j}^*(\mathbf{r}'; \omega; \varpi)) = 0, \quad i \neq j, \forall(\mathbf{r}, \mathbf{r}'),$$

where the superscript asterisk denotes the Hermitian conjugate and  $\mathbb{E}_{\varpi}$  defines the expectation operator over snapshot.

(A2) Assumption (A2) is perfect spatial coherence of sources and consequently of the partial sound fields they radiate,

$$\frac{\|\mathbb{E}_{\varpi}(p_{0,i}(\mathbf{r}; \omega; \varpi)p_{0,i}^*(\mathbf{r}'; \omega; \varpi))\|^2}{\mathbb{E}_{\varpi}\|p_{0,i}(\mathbf{r}; \omega; \varpi)\|^2\mathbb{E}_{\varpi}\|p_{0,i}(\mathbf{r}'; \omega; \varpi)\|^2} = 1, \quad \forall(\mathbf{r}, \mathbf{r}').$$

(A3) Assumption (A3) is stationarity of sources, which implies that the partial sound fields they radiate are uncorrelated over frequencies,

$$\mathbb{E}_{\varpi}(p_{0,i}(\mathbf{r}; \omega; \varpi)p_{0,j}(\mathbf{r}'; \omega'; \varpi)) = 0, \quad \omega \neq \omega', \forall(i, j).$$

Ergodicity will also be assumed for technical reasons, which simply means that expectation  $\mathbb{E}_{\varpi}$  over snapshots converges to a deterministic quantity.

Note that assumptions (A1) and (A2) are actually at the roots of a proper definition of the concept of ‘‘sound sources.’’ In physical terms, they simply mean that sources originating from different physical origins are not coherent (no phase relationship), whereas an individual source is perfectly coherent with itself.<sup>19</sup> Assumption (A3) is less fundamental and could be easily relaxed if necessary, even though it surely corresponds to many practical situations of interest.

Based on these assumptions, it is now possible to model sound source distribution, which is a crucial step towards achieving a successful separation. Each source distribution,  $s_i(\mathbf{r}; \omega; \varpi)$ , is assigned a spatial mode  $\phi_i(\mathbf{r}; \omega)$  multiplied by a quantity  $\epsilon_i(\omega; \varpi)$  describing its random amplitude. Note that the spatial mode is independent of the snapshot variable  $\varpi$  and that the amplitude  $\epsilon_i(\omega; \varpi)$ , called the ‘‘latent source’’ here, is used to reflect the snapshot-dependent properties of the source:

$$s_i(\mathbf{r}; \omega; \varpi) = \underbrace{\phi_i(\mathbf{r}; \omega)}_{\substack{\text{spatial dependence} \\ \text{(deterministic)}}} \times \underbrace{\epsilon_i(\omega; \varpi)}_{\substack{\text{snapshot dependence} \\ \text{(stochastic)}}, \quad (3)$$

for  $i = 1, \dots, n$ .

By construction, mutually uncorrelated latent sources will automatically fulfill assumptions (A1) and (A2) since

$$\mathbb{E}_{\varpi}(\epsilon_i(\omega_k; \varpi)\epsilon_j^*(\omega_l; \varpi)) = \delta_{ij}\delta_{kl}S_{\epsilon_i}(\omega_k), \quad (4)$$

with  $S_{\epsilon_i}(\omega_k)$  the power spectrum of  $\epsilon_i(\omega_k; \varpi)$  and  $\delta_{ij}$  the Kronecker symbol.

In a general setting the spatial modes may not be known *a priori*, but they may be developed on a predefined finite-dimensional set of spatial basis functions,  $\{\mu_k\}_{k=1}^{n_{\mu}}$ , to a reasonable accuracy:

$$\phi_i(\mathbf{r}; \omega) = \sum_{k=1}^{n_{\mu}} a_{ki}(\omega)\mu_k(\mathbf{r}; \omega). \quad (5)$$

[There are many possible choices for the spatial basis  $\{\mu_k(\mathbf{r}; \omega)\}_{k=1}^{n_{\mu}}$ , such as plane waves, splines, etc.; how to design an optimal basis of minimal dimension given an array geometry and a source surface topology is described in Ref. 20.] Now, inserting Eqs. (2)–(5) into Eq. (1) yields

$$p(\mathbf{r}_j; \omega; \varpi) = \sum_{i=1}^n \epsilon_i(\omega; \varpi) \sum_{k=1}^{n_{\mu}} a_{ki}(\omega)h_{jk}(\omega) + \nu(\mathbf{r}_j; \omega; \varpi), \quad j = 1, \dots, m, \quad (6)$$

where  $h_{jk}(\omega) = \int_{\Gamma} \mu_k(\mathbf{r}'; \omega)g(\mathbf{r}_j; \mathbf{r}'; \omega)d\Gamma(\mathbf{r}')$  is the radiation of the spatial basis function  $\mu_k(\mathbf{r}'; \omega)$  to the  $j$ th microphone of the array. Equation (6) can be rewritten in a matrix form,

$$\mathbf{p}(\omega; \varpi) = \mathbf{H}(\omega)\mathbf{A}(\omega)\boldsymbol{\epsilon}(\omega; \varpi) + \mathbf{v}(\omega; \varpi), \quad (7)$$

where the column vectors  $\mathbf{p}(\omega; \varpi) \in \mathbb{C}^m$  and  $\mathbf{v}(\omega; \varpi) \in \mathbb{C}^m$  collect the measured pressures and measurement noise at  $m$  microphones, respectively, the propagation matrix  $\mathbf{H}(\omega) \in \mathbb{C}^{m \times n_{\mu}}$  is of the  $j$ -row and  $k$ -column element formed by  $h_{jk}(\omega)$ ,  $\mathbf{A}(\omega) \in \mathbb{C}^{n_{\mu} \times n}$  is of the  $k$ -row and  $i$ -column element formed by  $a_{ki}(\omega)$ , and the column vector  $\boldsymbol{\epsilon}(\omega; \varpi) \in \mathbb{C}^n$  collects the amplitude variables. Equation (7) expresses the inverse problem in a compact and discrete form: the unknowns of the problem are matrix  $\mathbf{A}(\omega)$  of projection coefficients and vector  $\boldsymbol{\epsilon}(\omega; \varpi)$  of latent sources. Standard back-propagation (i.e., near-field acoustical holography) amounts to estimating the product  $\mathbf{c}(\omega; \varpi) = \mathbf{A}(\omega)\boldsymbol{\epsilon}(\omega; \varpi)$  from measurements  $\mathbf{p}(\omega; \varpi)$ ,<sup>20</sup> from which the global source distribution

$$s(\mathbf{r}; \omega; \varpi) = \sum_{k=1}^{n_{\mu}} \mu_k(\mathbf{r}; \omega)c_k(\omega; \varpi) \quad (8)$$

can be reconstructed, whereas source separation ambitions the further issue of estimating the two factors  $\mathbf{A}(\omega)$  and  $\boldsymbol{\epsilon}(\omega; \varpi)$  from which individual sources

$$s_i(\mathbf{r}; \omega; \varpi) = \sum_{k=1}^{n_{\mu}} \mu_k(\mathbf{r}; \omega)a_{ki}(\omega)\epsilon_i(\omega; \varpi) \quad (9)$$

can be recovered. So far, it is quite obvious that such a separation has no unique solution if no further physical characteristics are taken into account (many couples of different factors  $\mathbf{A}$  and  $\boldsymbol{\epsilon}$  are likely to produce the same product). One property investigated in this paper is the case of sources with distinct power spectra and smooth spatial modes in frequency:

(A4) Assumption (A4) is that there exists non-empty (narrow) frequency bands  $\mathcal{B}$  centered at  $\omega_m$  over which sound sources have distinct power spectra,  $S_{\epsilon_i}(\omega) \neq S_{\epsilon_j}(\omega)$ ,  $i \neq j$ , and  $\phi(\mathbf{r}; \omega) \approx \phi(\mathbf{r}, \omega_m)$ ,  $\omega \in \mathcal{B}$ .

Indeed, for many types of source distributions the spatial mode  $\phi(\mathbf{r}; \omega)$  can be reasonably considered smooth with respect to frequency  $\omega$ , as well as basis functions  $\mu_k(\mathbf{r}; \omega)$ . Thus matrix  $\mathbf{A}$  is invariant across  $\mathcal{B}$  and Eq. (7) is simplified into

$$\mathbf{p}(\omega; \varpi) \approx \mathbf{H}(\omega)\mathbf{A}\boldsymbol{\epsilon}(\omega; \varpi) + \mathbf{v}(\omega; \varpi), \quad \omega \in \mathcal{B}. \quad (10)$$

Comparing Eq. (10) with the standard formulation of source separation,<sup>12</sup>  $\mathbf{A}$  is understood as a mixing matrix,  $\boldsymbol{\epsilon}$  as sources (latent variables) to be separated, but with a pre-multiplying matrix  $\mathbf{H}(\omega)$ . This equation will then have to be solved for  $\mathbf{A}$  and  $\boldsymbol{\epsilon}(\omega; \varpi)$  over a set of (possibly overlapping) narrow frequency bands  $\{\mathcal{B}\}$ . The resolution of Eq. (10) has its own specific challenges: (a) the propagation matrix  $\mathbf{H}(\omega)$  is usually heavily ill-conditioned due to the presence of evanescent waves,<sup>1</sup> (b) the mixing matrix is of large size ( $n_\mu \times n$ ) due to the large number of microphones. Both of these difficulties plus the fact that the issue involves the joint backpropagation and separation steps make the space-frequency separation of sound sources considerably more intricate than classical blind source separation. To break down these difficulties, a tractable hierarchical scheme is proposed by considering a vector of hidden variables  $\mathbf{c}(\omega; \varpi)$  which represents the projection of the source distribution onto the space of basis functions  $\{\mu_k\}_{k=1}^{n_\mu}$ ; thus Eq. (10) is equivalently transformed into a hierarchical form,

$$\mathbf{p}(\omega; \varpi) = \mathbf{H}(\omega)\mathbf{c}(\omega; \varpi) + \mathbf{v}(\omega; \varpi), \quad (11)$$

and

$$\mathbf{c}(\omega; \varpi) = \mathbf{A}\boldsymbol{\epsilon}(\omega; \varpi) + \mathbf{e}(\omega; \varpi), \quad (12)$$

where  $\mathbf{e}(\omega; \varpi)$  embodies uncertainty in the separation step; it is classically described by a zero-mean complex normal distribution with unknown diagonal covariance matrix  $\mathbf{C}_e$ . The introduction of the hidden variable is the key to break down the original problem of Eq. (10) into two sub-problems: first Eq. (11) involves source reconstruction by estimating the projection coefficients  $\mathbf{c}(\omega; \varpi)$  from the measurements and then Eq. (12) involves blind separation of coefficients  $\mathbf{c}(\omega; \varpi)$  into uncorrelated latent sources mixed by  $\mathbf{A}$ .

In following sections, the hierarchical scheme is developed and each step addressed from a Bayesian perspective

### III. SOURCE RECONSTRUCTION BY A BAYESIAN APPROACH

The reconstruction of sound sources from discrete field measurements is a difficult inverse problem which has been approached by different techniques.<sup>2,21</sup> The source reconstruction—i.e., resolution of Eq. (11)—has been recently investigated by the authors from a Bayesian perspective in Ref. 20. A brief overview is given here since it is at the root of the proposed source separation solution.

Equation (8) indeed shows the expansion of the total sound source onto the set of basis functions used in Eq. (5), where projection coefficients  $c_k$  have been formally intro-

duced in Eq. (11). They are the unknowns to be found from merging prior information and sound measurements. Before the experiment, they are assigned a prior probability distribution of the form

$$\pi(\mathbf{c}) \propto \sigma_\alpha^{-2n_\mu} \exp[-\sigma_\alpha^{-2}\Delta(\mathbf{c})], \quad (13)$$

where  $\propto$  stands for the proportionality sign,  $\Delta(\mathbf{c}) = \int \|\sum_{k=1}^{n_\mu} c_k(\omega; \varpi)\mu_k(\mathbf{r}; \omega)\|^2 \sigma_s^{-2}(\mathbf{r}) d\Gamma(\mathbf{r})$ , and  $\sigma_s^2(\mathbf{r})$  is an aperture function that reflects the belief on the acoustic spatial origins, which is used to enhance the resolution of reconstructed sources. In other words, before the experiment is carried out, the actual source distribution is assumed to be a realization of a Gaussian random field with covariance function  $\mathbb{E}_\varpi(s(\mathbf{r}; \omega; \varpi)s^*(\mathbf{r}'; \omega; \varpi)) = \delta(\mathbf{r} - \mathbf{r}')\sigma_s^2(\mathbf{r})$  in the Gibbs ensemble sense, with  $\delta$  the Dirac delta—see Ref. 20 for details. The next step is to update the probability distribution of the unknown coefficients  $\mathbf{c}$  after the experiment has returned a set of measured pressures; in this respect, the probability density function of the additive noise, i.e., the difference between theoretical and measured pressures, is required.

By virtue of the Central Limit Theorem<sup>22</sup> applied to the Fourier transform, the measurement noise in frequency domain has a probability distribution rapidly converging to a zero-mean complex Gaussian distribution. In addition, it is physically sensible to assume that measurement noises at different microphones are spatially uncorrelated with identical unknown variance  $\sigma_\nu^2$ . Based on the additive output error model in Eq. (11), the likelihood function of the unknowns then reads

$$\pi(\mathbf{p}|\mathbf{c}) \propto \sigma_\nu^{-2m} \exp[-\Delta(\mathbf{p} - \mathbf{H}\mathbf{c})], \quad (14)$$

with  $\Delta(\mathbf{p} - \mathbf{H}\mathbf{c}) = \|\mathbf{p}(\omega; \varpi) - \mathbf{H}(\omega)\mathbf{c}(\omega; \varpi)\|_{\sigma_\nu^2}^2$  (where the notational convention  $\|\mathbf{v}\|_{\boldsymbol{\Omega}}^2 = \mathbf{v}^*\boldsymbol{\Omega}^{-1}\mathbf{v}$  is used).

Through Bayes' theorem, the posterior distribution of the unknowns is finally returned as

$$\pi(\mathbf{c}|\mathbf{p}) = \frac{\pi(\mathbf{p}|\mathbf{c})\pi(\mathbf{c})}{\pi(\mathbf{p})} = \mathcal{N}_c[\hat{\mathbf{c}}(\omega; \varpi), \mathbf{C}_c], \quad (15)$$

that is, the projection coefficients follow *a posteriori* a Gaussian law with mean and covariance matrix equal to

$$\begin{aligned} \hat{c}_k(\omega; \varpi) &= \frac{\lambda_k}{\lambda_k^2 + \hat{\sigma}_\nu^2/\hat{\sigma}_\alpha^2} \mathbf{U}_k^* \mathbf{p}(\omega; \varpi), \\ C_c(k, k') &= \frac{\hat{\sigma}_\nu^2}{\lambda_k^2 + \hat{\sigma}_\nu^2/\hat{\sigma}_\alpha^2} \delta_{kk'}, \end{aligned} \quad (16)$$

respectively, where  $\mathbf{U}_k$  is the  $k$ th eigenvector of the Gramian matrix  $[\int_\Gamma g(\mathbf{r}_j; \mathbf{r}'; \omega)g^*(\mathbf{r}_j; \mathbf{r}'; \omega)\sigma_s^2(\mathbf{r}')d\Gamma(\mathbf{r}')]$ ,  $\delta_{kk'}$  is the Kronecker symbol, and  $\hat{\sigma}_\alpha^2$  and  $\hat{\sigma}_\nu^2$  are optimal values of the source and noise power estimated by evidence optimization.<sup>20,23</sup> Following similar lines, one may also construct a spatial basis  $\{\mu_k\}_{k=1}^{n_\mu}$  which is optimal for a given source topology and array geometry and has a minimal dimension equal to number of microphones ( $n_\mu = m$ )—see Ref. 20.

To make a seamless connection between the reconstruction and separation steps, let us denote the projection coefficients  $\mathbf{c}(\omega; \varpi)$  as the current estimate  $\hat{\mathbf{c}}(\omega; \varpi)$  plus some discrepancy  $\delta\mathbf{c}(\omega; \varpi)$ , i.e.,  $\mathbf{c}(\omega; \varpi) = \hat{\mathbf{c}}(\omega; \varpi) + \delta\mathbf{c}(\omega; \varpi)$  with  $\delta\mathbf{c}(\omega; \varpi) \sim \mathcal{N}_c(\mathbf{0}, \mathbf{C}_c)$ . According to the previous discussion, the separation step in Eq. (12) is thus rewritten as

$$\hat{\mathbf{c}}(\omega; \varpi) = \mathbf{A}\boldsymbol{\epsilon}(\omega; \varpi) + \tilde{\mathbf{e}}(\omega; \varpi), \quad (17)$$

with  $\tilde{\mathbf{e}}(\omega; \varpi) = \mathbf{e}(\omega; \varpi) - \delta\mathbf{c}(\omega; \varpi)$ .

The next section addresses the blind separation of Eq. (17) by means of a full Bayesian approach.

## IV. SOURCE SEPARATION BY A FULL BAYESIAN APPROACH

### A. Specification of priors and likelihood function

Assumptions (A1)–(A2) imply that the projection of the sound sources onto the basis functions can be separated into uncorrelated components originating from sound sources of distinct physical origins. The separation can be achieved based on different assumptions on the statistical properties of sources such as independence, nonstationarity, sparsity.<sup>12</sup> The present work considers stationary sources [assumption (A3)] and exploits the diversity of the sources spectra in the frequency band of interest [assumption (A4)]. Before the experiment is undertaken, these properties have to be encoded by means of a prior probability distribution.

#### 1. A prior probability distribution on latent sources

The latent sources are modeled as stationary with different covariance matrix at different frequency lines. Based on only the stationary second-order statistics, the prior probability distribution of the latent sources reads (up to a constant)

$$\pi(\boldsymbol{\epsilon}|\mathbf{C}_\epsilon) \propto \prod_{k=1}^{n_\omega} |\mathbf{C}_\epsilon^{-1}(\omega_k)|^{n_\varpi} \exp[-\Delta(\boldsymbol{\epsilon})], \quad (18)$$

where  $\Delta(\boldsymbol{\epsilon}) = \sum_{k=1}^{n_\omega} \sum_{j=1}^{n_\varpi} \|\boldsymbol{\epsilon}(\omega_k; \varpi_j)\|_{\mathbf{C}_\epsilon(\omega_k)}^2$ , the covariance matrix  $\mathbf{C}_\epsilon(\omega_k) = \text{diag}(\sigma_{\epsilon_{1k}}^2, \dots, \sigma_{\epsilon_{nk}}^2)$  with  $\sigma_{\epsilon_{ik}}^2 = \sigma_{\epsilon_i}^2(\omega_k)$ ,  $n_\omega$  is the number of frequency lines in band  $\mathcal{B}$ , and  $n_\varpi$  is the number of snapshots. Conjugate prior probability distributions are assigned to the inverse of source variances by a Gamma distribution,  $\sigma_{\epsilon_{ik}}^{-2} \sim \text{Gamma}(\alpha_\epsilon, \beta_\epsilon)$ ,  $\forall(k, i)$ , with known hyper-parameters  $(\alpha_\epsilon, \beta_\epsilon)$ . A conjugate prior is in the same probability distribution family as the posterior,<sup>14</sup> which is advocated in the present work as it can make efficient the MCMC sampling of the posterior probability distribution.

#### 2. A prior probability distribution on mixing matrix

Similarly, the Bayesian approach can incorporate any prior on the mixing matrix  $\mathbf{A}$ , a fundamental step towards regularization of the inverse problem. Since sound sources stem from distinct physical origins, they are likely to display

distinct spatial modes. Such a property implies a weak correlation,

$$\begin{aligned} \rho(\mathbf{a}_i, \mathbf{a}_j) &= \int_{\Gamma} \phi_i^*(\mathbf{r}; \omega_m) \phi_j(\mathbf{r}; \omega_m) d\Gamma(\mathbf{r}) \\ &= \mathbf{a}_i^* \mathbf{R}_{\mu\mu} \mathbf{a}_j, \end{aligned} \quad (19)$$

between modes of different sound sources at center frequency  $\omega_m$  in a band  $\mathcal{B}$ ,  $i \neq j$ , where  $\phi(\mathbf{r})$  is defined in Eq. (5),  $\mathbf{a}_i$  and  $\mathbf{a}_j$  are the  $i$ th and  $j$ th columns of the mixing matrix, and  $[\mathbf{R}_{\mu\mu}]_{kl} = \int_{\Gamma} \mu_k^*(\mathbf{r}; \omega) \mu_l(\mathbf{r}; \omega) d\Gamma(\mathbf{r})$ . The more distinct the sound sources in space, the smaller their spatial correlation. The complex correlation variables  $\{\rho(\mathbf{a}_i, \mathbf{a}_j)\}_{i \neq j}$  are considered as identically normally distributed with zero mean and unknown variance  $w$ . The following prior is therefore attributed to the columns of the mixing matrix by means of the probability distribution function of the correlation variable,

$$\pi_\rho(\{\mathbf{a}_i\}_{i=1}^n | w) \propto \frac{1}{w^{n(n-1)}} \exp \left[ \frac{-1}{w} \sum_{\substack{i,j=1 \\ i \neq j}}^n \|\rho(\mathbf{a}_i, \mathbf{a}_j)\|^2 \right], \quad (20)$$

which can impose a soft spatial constraint on the sound sources, and favor their separability in space (the higher the prior probability, the more distinct the sound sources in space). In order not to overload the notation, it is used that  $\pi(\mathbf{A}|w) = \pi_\rho(\{\mathbf{a}_i\}_{i=1}^n | w)$ .

The scaling coefficient  $w$  has to be tuned to find an optimal weighting for the prior in the Bayesian formulation. To maintain the conjugacy property, a Gamma distribution is adopted,  $w^{-1} \sim \text{Gamma}(\alpha_w, \beta_w)$ .

### 3. Likelihood function

Once prior probability distributions have been assigned to variables  $\boldsymbol{\epsilon}$  and  $\mathbf{A}$ , it still remains to specify the probability distribution of coefficients  $\mathbf{c}$  conditioned on  $\boldsymbol{\epsilon}$  and  $\mathbf{A}$ : this will “close the loop” in our effort to estimate the latter factors from the observation of their product only.

Let us consider the random variables  $\delta\mathbf{c}(\omega; \varpi)$  and  $\mathbf{e}(\omega; \varpi)$  in the hierarchical steps of 19 to be independent, and further consider  $\mathbf{C}_\epsilon$  to be proportional to  $\mathbf{C}_c$  in favor of the subsequent Bayesian learning by exploring the whole probability space of parameters through MCMC sampling. As a result,  $\mathbf{C}_\epsilon = \tau \mathbf{C}_c$ . The likelihood function of the unknown parameters is then written as,

$$\pi(\mathbf{c}|\mathbf{A}, \boldsymbol{\epsilon}, \tau) \propto |\mathbf{C}_\epsilon^{-1}|^{n_\omega n_\varpi} \exp[-\Delta(\mathbf{A}, \boldsymbol{\epsilon})], \quad (21)$$

with  $\Delta(\mathbf{A}, \boldsymbol{\epsilon}) = \sum_{k=1}^{n_\omega} \sum_{j=1}^{n_\varpi} \|\hat{\mathbf{c}}(\omega_k; \varpi_j) - \mathbf{A}\boldsymbol{\epsilon}(\omega_k; \varpi_j)\|_{\mathbf{C}_\epsilon}^2$ .

An improper prior density  $\pi(\tau) = 1/\tau$  [so-called because the area of the distribution  $\pi(\tau) = 1/\tau$  does not equal one for  $\tau$  taking values between zero and infinity] is used, but any inverse-gamma prior for  $\tau$  would also maintain conjugacy. Conditioning on  $\tau$  is important, because it guarantees a unimodal full posterior.

## B. Markov Chain Monte Carlo implementation

The posterior probability distribution of all the unknown parameters is next obtained from Bayes' theorem,

$$\pi(\mathbf{A}, \boldsymbol{\epsilon}, \mathbf{C}_\epsilon, \tau, w | \hat{\mathbf{c}}) \propto \pi(\hat{\mathbf{c}} | \mathbf{A}, \boldsymbol{\epsilon}, \tau) \pi(\boldsymbol{\epsilon} | \mathbf{C}_\epsilon) \times \pi(\mathbf{A} | w) \pi(\mathbf{C}_\epsilon) \pi(\tau) \pi(w), \quad (22)$$

where all the distributions on the right hand side of the equation have been described above. The posterior  $\pi(\mathbf{A}, \boldsymbol{\epsilon}, \mathbf{C}_\epsilon, \tau, w | \hat{\mathbf{c}})$  plays a crucial role in the proposed approach because it assigns probabilities to possible values of factors  $\boldsymbol{\epsilon}$  and  $\mathbf{A}$  after observing the projection coefficients  $\hat{\mathbf{c}}$ . In particular, values with the highest probability will provide a valid solution to the source separation problem. How to numerically evaluate the full posterior probability distribution—and not only its mode—is described here after. Since the conditional posterior densities of all the unknowns are tractable distributions that do not require rejection sampling, the Gibbs sampling method<sup>24</sup> is particularly recommended here for parameter inference. The Gibbs sampler is basically an iterative simulation scheme for generating samples that converge asymptotically to the target distribution  $\pi(\mathbf{A}, \boldsymbol{\epsilon}, \mathbf{C}_\epsilon, \tau, w | \hat{\mathbf{c}})$ . Each component is visited and updated by a sample drawn from its conditional distribution. The expressions of all the conditional distributions are listed as follows:

- (a)  $\boldsymbol{\epsilon}(\omega_k; \boldsymbol{\tau}_j) | \hat{\mathbf{c}}, \mathbf{A}, \tau, \mathbf{C}_\epsilon(\omega_k) \sim \mathcal{N}_c(\hat{\boldsymbol{\epsilon}}(\omega_k; \boldsymbol{\tau}_j), \hat{\mathbf{C}}_\epsilon(\omega_k))$  with

$$\begin{aligned} \hat{\mathbf{C}}_\epsilon(\omega_k) &= (\mathbf{A}^* \mathbf{C}_\epsilon^{-1} \mathbf{A} + \mathbf{C}_\epsilon^{-1}(\omega_k))^{-1}, \\ \hat{\boldsymbol{\epsilon}}(\omega_k; \boldsymbol{\tau}_j) &= \hat{\mathbf{C}}_\epsilon(\omega_k) (\mathbf{A}^* \mathbf{C}_\epsilon^{-1} \hat{\mathbf{c}}(\omega_k; \boldsymbol{\tau}_j)). \end{aligned} \quad (23)$$

- (b)  $\sigma_{\epsilon_i}^{-2}(\omega_k) \sim \text{Gamma}(\hat{\alpha}_\epsilon, \hat{\beta}_\epsilon)$  with

$$\hat{\alpha}_\epsilon = \alpha_\epsilon + n_\tau, \quad \hat{\beta}_\epsilon = \beta_\epsilon + \sum_{j=1}^{n_\tau} \|\epsilon_i(\omega_k; \boldsymbol{\tau}_j)\|^2, \quad (24)$$

for all  $i = 1, \dots, n$  and  $k = 1, \dots, n_\omega$ .

- (c)  $\tau^{-1} | \hat{\mathbf{c}}, \mathbf{A} \sim \text{Gamma}(\hat{\alpha}_\tau, \hat{\beta}_\tau)$  with  $\hat{\alpha}_\tau = mn_\omega n_\tau + 1$  and

$$\hat{\beta}_\tau = \sum_{k=1}^{n_\omega} \sum_{j=1}^{n_\tau} \|\hat{\mathbf{c}}(\omega_k; \boldsymbol{\tau}_j) - \mathbf{A}\boldsymbol{\epsilon}(\omega_k; \boldsymbol{\tau}_j)\|_{\mathbf{C}_\epsilon}^2. \quad (25)$$

- (d) The  $i$ th column of the mixing matrix,  $\mathbf{a}_i | \hat{\mathbf{c}}, \{\mathbf{a}_j\}_{j=1, j \neq i}^n, \boldsymbol{\epsilon}, \tau, w \sim \mathcal{N}_c(\hat{\mathbf{a}}_i, \hat{\mathbf{C}}_{\mathbf{a}_i})$  where

$$\begin{aligned} \hat{\mathbf{C}}_{\mathbf{a}_i} &= (\mathbf{R}_{\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i} \mathbf{C}_\epsilon^{-1} + w^{-1} \mathbf{R}_{\mu\mu}^* \mathbf{R}_{\mathbf{a}\mathbf{a}} \mathbf{R}_{\mu\mu})^{-1}, \\ \hat{\mathbf{a}}_i &= \hat{\mathbf{C}}_{\mathbf{a}_i} \mathbf{C}_\epsilon^{-1} \mathbf{R}_{\boldsymbol{\epsilon}_i}, \end{aligned} \quad (26)$$

for all  $i = 1, \dots, n$  with  $\mathbf{R}_{\mathbf{a}\mathbf{a}} = \sum_{l=1, l \neq i}^n \mathbf{a}_l \mathbf{a}_l^*$ ,  $\mathbf{R}_{\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i} = \sum_{k=1}^{n_\omega} \sum_{j=1}^{n_\tau} \|\epsilon_i(\omega_k; \boldsymbol{\tau}_j)\|^2$  and  $\mathbf{R}_{\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i} = \sum_{k=1}^{n_\omega} \sum_{j=1}^{n_\tau} (\hat{\mathbf{c}}(\omega_k; \boldsymbol{\tau}_j) - \sum_{l=1, l \neq i}^n \mathbf{a}_l \epsilon_l(\omega_k; \boldsymbol{\tau}_j)) \epsilon_i^*(\omega_k; \boldsymbol{\tau}_j)$ .

- (e)  $w | \mathbf{A} \sim \text{Gamma}(\hat{\alpha}_w, \hat{\beta}_w)$  with  $\hat{\alpha}_w = \alpha_w + n(n-1)$ ,

$$\hat{\beta}_w = \beta_w + \sum_{i,j=1; i \neq j}^n \|\rho(\mathbf{a}_i, \mathbf{a}_j)\|^2. \quad (27)$$

In summary, Gibbs sampling basically initializes firstly  $\mathbf{A}$ ,  $\mathbf{C}_\epsilon$ ,  $\tau$ ,  $w$  following their prior distributions, and then executes in a cyclic manner the steps from (a) to (e). A good starting point of MCMC sampling is crucial to shorten the burn-in period of the Markov chain, especially for our case with many unknown parameters. The burn-in period is the time required by the Markov chain for its convergence to the target probability density function. The mixing matrix is initialized with the Second Order Blind Identification (SOBI) method.<sup>25</sup> [SOBI performs separation by exploiting the decorrelation properties of the source signals: it basically relies only on finding  $\mathbf{A}$  which jointly diagonalizes the set of covariance matrices  $\mathbf{R}_{\boldsymbol{\epsilon}\boldsymbol{\epsilon}}(\omega_k) = \sum_{j=1}^{n_\tau} \boldsymbol{\epsilon}(\omega_k; \boldsymbol{\tau}_j) \boldsymbol{\epsilon}^*(\omega_k; \boldsymbol{\tau}_j)$  for all  $k = 1, \dots, n_\omega$  in band  $\mathcal{B}$ .]

## C. Bandwidth

The choice of bandwidth of  $\mathcal{B}$  plays an important role in maintaining the performance of the algorithm. For instance, if the range of  $\mathcal{B}$  is chosen large, this increases the diversity of the spectra of the sources which plays in favor of separation; yet at the same time the mixing matrix  $\mathbf{A}$  is less likely to remain constant in the frequency band, which jeopardizing the model of Eq. (10). A reasonable trade-off will be made by the following criterion. By considering the sound sources originating from distinct space origins, the separation quality can be indicated by the correlation coefficients between the spatial modes of two different sound sources [as computed in Eq. (19)]. Following Eq. (20), a correlation index is defined as

$$\rho = \sum_{i,j=1; i \neq j}^n \|\rho(\mathbf{a}_i, \mathbf{a}_j)\|^2, \quad (28)$$

which may decrease when the width of  $\mathcal{B}$  increase from zero to a certain level. The error caused by choosing a finite width of  $\mathcal{B}$  can be reflected by the normalized residual

$$\gamma = \frac{\sum_{l=1}^m \sum_{k=1}^{n_\omega} \sum_{j=1}^{n_\tau} \|\hat{c}_l(\omega_k; \boldsymbol{\tau}_j) - \hat{\mathbf{c}}_l(\omega_k; \boldsymbol{\tau}_j)\|^2}{\sum_{l=1}^m \sum_{k=1}^{n_\omega} \sum_{j=1}^{n_\tau} \|\hat{c}_l(\omega_k; \boldsymbol{\tau}_j)\|^2}, \quad (29)$$

where  $\hat{\mathbf{c}} = \hat{\mathbf{A}} \hat{\boldsymbol{\epsilon}}$  is computed with the estimates of the mixing matrix and sources returned by separation. It will increase with the chosen bandwidth. Both index will serve to determine an appropriate bandwidth in a less heuristic way.

## D. Indeterminacy

An indeterminacy problem always exists with blind source separation. To see this, let us remember that estimates of  $\mathbf{A}$  and  $\boldsymbol{\epsilon}$  have to be found such that  $\hat{\mathbf{c}} \simeq \mathbf{A}\boldsymbol{\epsilon}$ . It is immediate that such a factorization is not unique, since  $\hat{\mathbf{c}} \simeq (\mathbf{A}\mathbf{P}^{-1})(\mathbf{P}\boldsymbol{\epsilon})$  will also return a valid solution for any invertible matrix  $\mathbf{P}$  that keeps diagonal the correlation matrix of  $\mathbf{P}\boldsymbol{\epsilon}$ , as required from assumptions (A1) and (A2). This may be solved by imposing a constraint on the mixing matrix. For instance, the first row of  $\mathbf{A}$  are constrained to be unit. To keep invariant the Markovian properties of MCMC due to the scaling

constraint on the columns of  $\mathbf{A}$ , the following adjustments are made,  $\forall k = 1, \dots, n_\omega$ ,

$$\begin{aligned} \mathbf{A} &\leftarrow \mathbf{A}\mathbf{P}^{-1}, \\ \mathbf{C}_\epsilon(\omega_k) &\leftarrow \mathbf{P}\mathbf{C}_\epsilon(\omega_k)\mathbf{P}^*, \end{aligned} \quad (30)$$

where the  $l$ th element of the diagonal matrix  $\mathbf{P}$  is in the first element of the  $l$ th column of  $\mathbf{A}$ .

### E. Acceleration strategy

A good initialization of the numerous elements in the mixing matrix is a key step to accelerate MCMC sampling in a large frequency band. It was mentioned previously that this can be achieved by first running the SOBI algorithm. Here a alternative acceleration strategy is presented based on the smooth property of source spatial modes across consecutive frequency bands. Namely, if the  $i$ th spatial mode in band  $\mathcal{B}^{(k+1)}$  is not too different from that in the previous band,  $\mathcal{B}^{(k)}$ , then its coefficients onto basis  $\{\mu_k(\mathbf{r}; \omega)\}_{k=1}^{n_\mu}$  [see Eq. (5)] are likely to follow a random-walk model,

$$a_{il}^{(k+1)} = a_{il}^{(k)} + f_{il}, \quad (31)$$

where  $f_{il}$  represents a zero-mean random variable. It has been found in applications involving sound sources with smooth spatial modes (e.g., as generated by loudspeakers) that this initialization strategy works better than SOBI, especially when the mixing matrix has a large dimension.

### F. Sound source discrimination

Once the backpropagation/separation procedure is terminated, one can easily reconstruct each sound source from the estimated mixing matrix and latent sources. Two principal indicators are defined to discriminate the sound source distributions. One is the power radiated by each source to the microphones, defined as  $R_{s_i}(\omega_k) = \mathbb{E}_{\varpi} \{ \int_{\Gamma} \|p_{0,i}(\mathbf{r}; \omega_k; \varpi)\|^2 \times d\Gamma(\mathbf{r}) \}$  at frequency  $\omega_k$ ,  $\forall i = 1, \dots, n$ . In practice,

$$R_{s_i}(\omega_k) \approx \frac{1}{mn_{\varpi}} \sum_{j=1}^{n_m} \|\mathbf{H}(\omega_k)\hat{\mathbf{a}}_i\hat{\epsilon}_i(\omega_k; \varpi_j)\|^2, \quad (32)$$

with  $\hat{\mathbf{a}}_i$  the  $i$ th column of  $\hat{\mathbf{A}}$  and  $\hat{\epsilon}_i(\omega_k; \varpi_j)$  the  $i$ th estimated latent source at frequency  $\omega_k$  for the  $j$ th snapshot. By definition, this can be used to rank sound sources, for instance in an attempt to identify the most noisy ones. The other indicator is the acoustic map of the source magnitude on the backpropagation surface, defined as  $M_{s_i}(\mathbf{r}, \omega_k) = \mathbb{E}_{\varpi} \{ \|s_i(\mathbf{r}; \omega_k)\| \}$  at frequency  $\omega_k$ ,  $\forall i = 1, \dots, n$ . In practice,

$$M_{s_i}(\mathbf{r}, \omega_k) \approx \frac{1}{n_{\varpi}} \sum_{j=1}^{n_m} \left\| \sum_{l=1}^{n_\mu} \mu_l(\mathbf{r}; \omega_k)\hat{a}_{li}\hat{\epsilon}_i(\omega_k; \varpi_j) \right\|, \quad (33)$$

with  $\mu_l(\mathbf{r}; \omega_k)$  the  $l$ th basis function evaluated in a user-defined spatial grid and  $\hat{a}_{li}$  the  $l$ th row and  $i$ th column element of  $\hat{\mathbf{A}}$ . This indicator is found useful to localize the distribution of the  $i$ th source. Other indicators may as well be defined on similar lines, such as the acoustic intensity for instance.

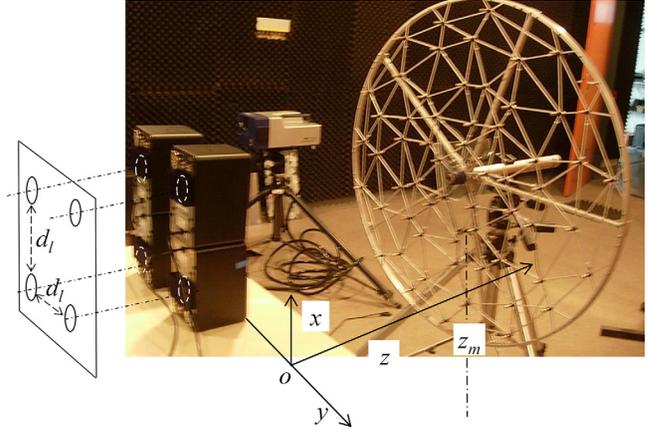


FIG. 2. (Color online) Experimental plant with microphone array.

## V. EXPERIMENTAL APPLICATIONS

In this section, the proposed methodology is demonstrated on laboratory experiments to separate the sound sources emitted by loudspeakers. The experiments are performed in a semi-anechoic room with a planar array of 60 randomly spaced microphones, as shown in Fig. 2. The mean spacing between microphones is  $d_m = 0.1$  m. The loudspeakers are driven by independent Gaussian noises with distinct spectra. The separation procedure is here verified on two experimental configurations with different source numbers and array distances. In each test case, additional experiments are systematically carried out where only one loudspeaker is switched on, one after the other, thus allowing the assessment of the partial radiations to the microphones from each source; for instance, with the  $i$ th source switched on only,

$$R_{s_i}(\omega_k) = \frac{1}{mn_{\varpi}} \sum_{l=1}^m \sum_{j=1}^{n_{\varpi}} \|p_i(\mathbf{r}_l; \omega_k; \varpi_j)\|^2, \quad (34)$$

which serves as a reference point to validate Eq. (32). In addition, the recorded data will serve to validate Eq. (33) after backpropagating individual sources according to the reconstruction step addressed in Sec. III.

The backpropagation/separation capacity of the algorithm is restricted to some extent by the experimental

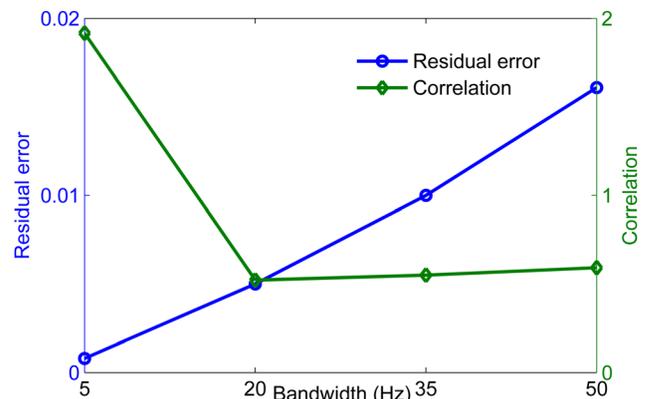


FIG. 3. (Color online) Normalized residual in Eq. (29) and correlation indicator in Eq. (28).

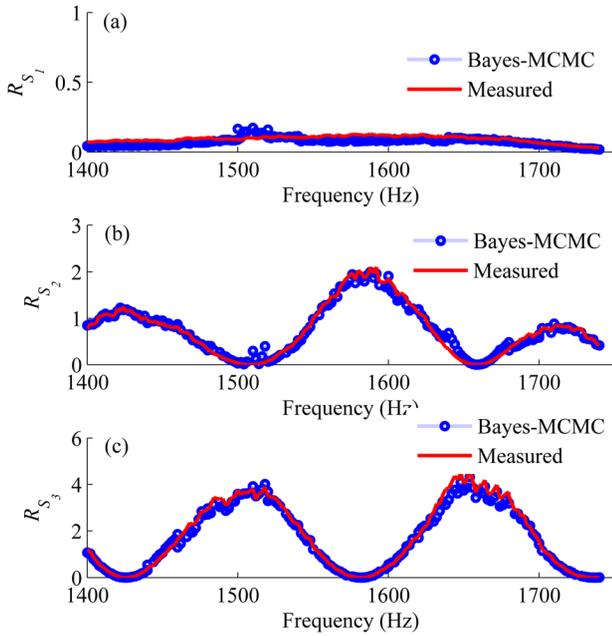


FIG. 4. (Color online) Radiated power,  $\{\mathbf{R}_{s_i}(\omega)\}_{i=1}^3$ , of each individual source.

configuration, such as the array shape, size and placement, microphone spacing, size of the source object, acquisition hardware, and so on. For instance, the spacing between the microphones determines the half-wavelength of the maximum frequency, and the size of the array roughly determines the half-wavelength of the minimum frequency.

### A. Case with three sources

Three loudspeakers are used with a center-distance of  $d_l = 0.12$  m. The array is situated in front of the loudspeakers with a distance of  $z_m = 0.7$  m. The acoustical signals are recorded with a sampling frequency  $F_s = 4096$  Hz.  $n_{\omega} = 420$  and  $n_{\omega} = 11.5 \times 10^4$  MCMC samples are kept after the burn-in phase, from which the mean values of  $\mathbf{A}$  and  $\epsilon$  are estimated. After a series of bandwidth of  $\mathcal{B}$  is investigated, as shown in Fig. 3, a width of 20 Hz is chosen for the source separation. The power radiated to the microphones by each source computed using Eq. (32) is shown in Fig. 4. The separation results agree well with their counterparts computed using Eq. (34) when only one source is switched on, which demonstrates the validity of the proposed backpropagation/separation procedure.

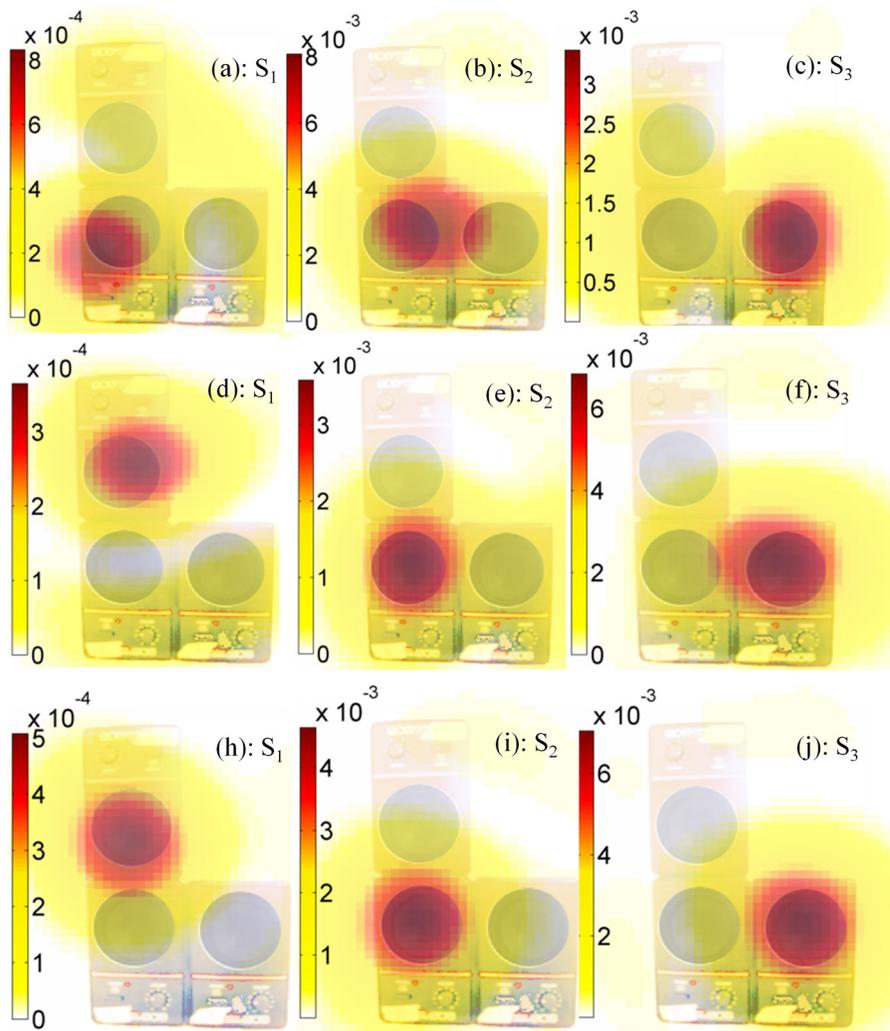


FIG. 5. (Color online) Maps of separated sound sources,  $\{M_{s_i}(\mathbf{r}, \omega)\}_{i=1}^3$ , at  $\omega = 1450$  Hz [(a)–(c): separation with SOBI; (d)–(f): Bayesian separation; (h)–(j): references].

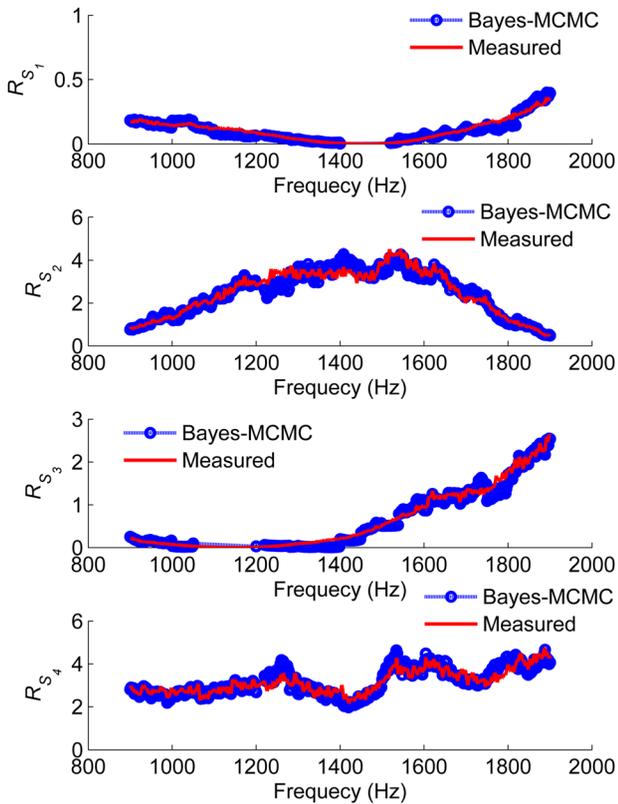


FIG. 6. (Color online) Radiated power,  $\{\mathbf{R}_{s_i}(\omega)\}_{i=1}^4$ , of each individual source.

In addition, the separation results are visualized by the acoustic maps of the reconstructed sources on the backpropagation plane. They are computed with Eq. (33) using the estimated separation parameters returned by SOBI and the established Bayesian estimation with MCMC method. Figure 5 compares the results at  $\omega = 1450$  Hz with the reconstructed counterparts when only one source is switched on. The separated sources are well localized with a quite accurate estimation of their varying amplitude levels, as can be seen by comparison with the references. The essential

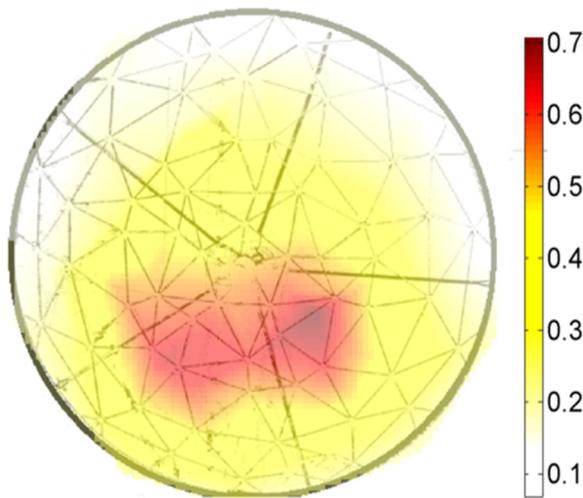


FIG. 7. (Color online) Magnitude of the interpolated sound field averaged over snapshots, at 2100 Hz, in the microphone plane.

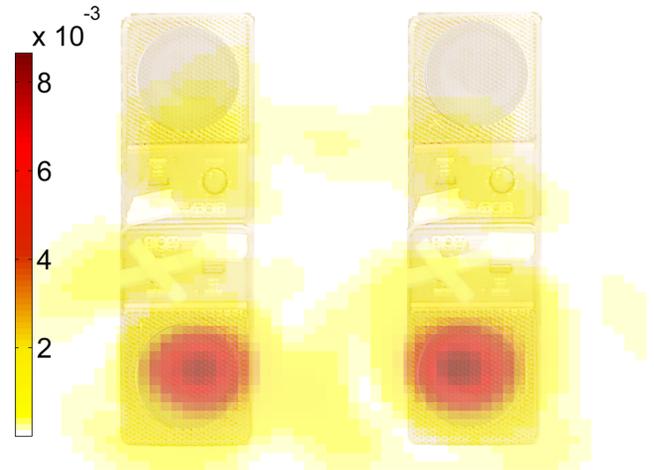


FIG. 8. (Color online) Reconstructed total sound distribution, at 2100 Hz, in the backpropagation plane.

reasons why the proposed Bayesian separation demonstrates supremacy over the standard SOBI algorithm is because it is rooted in a probabilistic approach that explicitly accounts for (1) measurement and modeling errors through the specification of a likelihood function and (2) for structural information about sources to be separated through the specification of prior probability distributions. The resulting posterior probability distribution is then explored by means of a powerful MCMC algorithm capable of exploring the whole space of the parameters, thus avoiding the estimate to be trapped in a local mode. Yet SOBI offers initial estimates, which is quite important to shorten MCMC running time.

### B. Case with four sources

In this case, four loudspeakers with an equal center-distance of  $d_l = 0.24$  m are assembled to generate a mixture of sound sources. The data are recorded with a higher

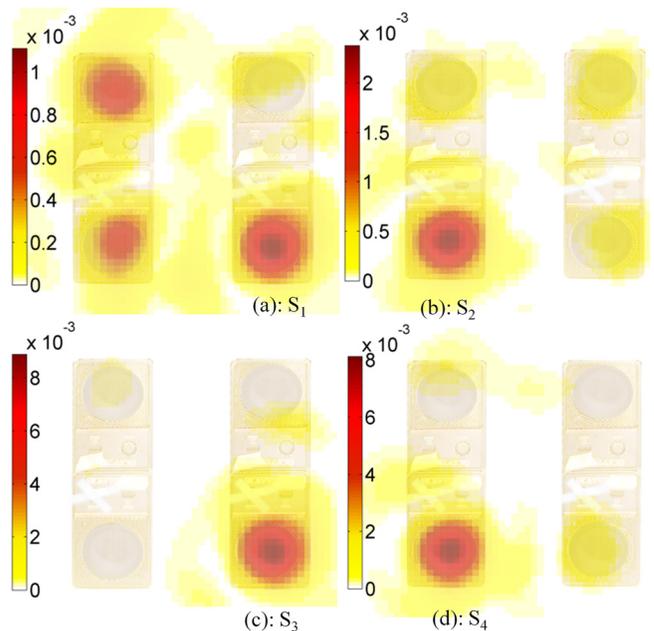


FIG. 9. (Color online) Map  $\{M_{s_i}(\mathbf{r}, \omega)\}_{i=1}^4$  of separated sources with SOBI, at  $\omega = 2100$  Hz.

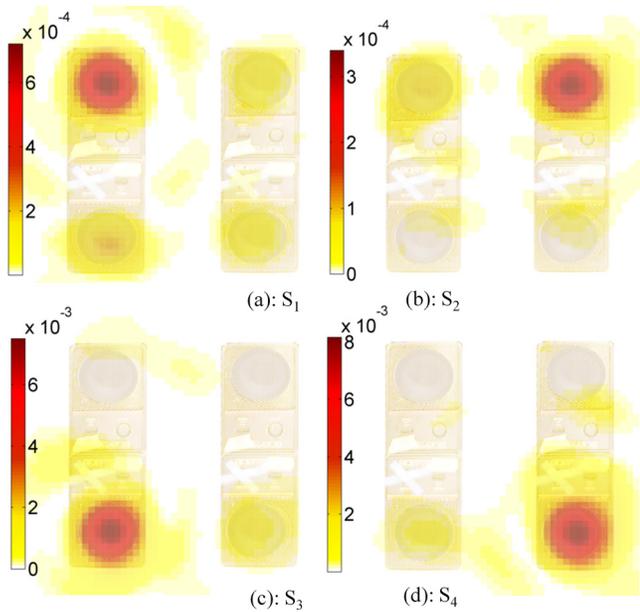


FIG. 10. (Color online) Map  $\{M_{S_i}(\mathbf{r}, \omega)\}_{i=1}^4$  of separated sources with the proposed Bayesian approach, at  $\omega = 2100$  Hz.

sampling frequency of 16 384 Hz. A nearer array distance is considered:  $z_m = 0.1$  m. The width of  $\mathcal{B}$  is chosen as 35 Hz to achieve a fair diversity of the source spectra while still maintaining the assumption of constant spatial modes—see Eq. (10).  $n_{\omega} = 480$  and  $n_{\omega} = 16$ . In the same way as in the previous case, the power radiated to the microphones from each source is computed and shown in Fig. 6. One can see a rather good agreement of the separation results with the reference data (measurements with only one loudspeaker is switched on) except for source 1 in band [1400,1500] Hz and for source 3 in [1050,1200] Hz, since they are too weak there. The measured sound field interpolated in the array plane is shown in Fig. 7. The total source distribution reconstructed in the backpropagation plane according to the meth-

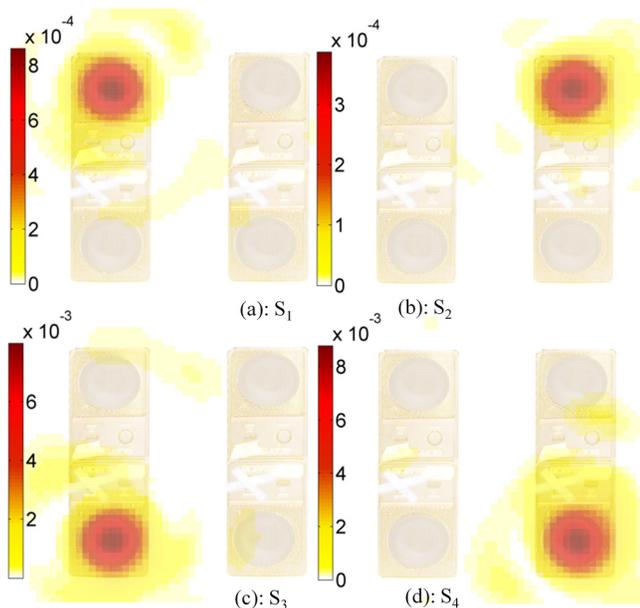


FIG. 11. (Color online) Map  $\{M_{S_i}(\mathbf{r}, \omega)\}_{i=1}^4$  of actual individual sources reconstructed from measurements when only one loudspeaker is switched on.

odology of Sec. III is shown in Fig. 8. The overlapping sources before separation clearly do not lend themselves to be identified—in terms of number and of localization—as they are located close to each other or have a large ratio of intensity levels (see the separation and backpropagation results in Figs. 9–11). This situation illustrates very well one of the motivations to develop a separation procedure in order to obtain a deeper insight into spatial distributions of underlying sources.

## VI. CONCLUSIONS

An efficient space-frequency source separation has been proposed and validated on laboratory experiments. The separation task is dealt with successfully by virtue of a hierarchical framework: the reconstruction step is solved with a minimal reconstruction error and the separation step is treated by a full Bayesian approach in conjunction with a MCMC algorithm. The method applies to mutually uncorrelated and stationary sound sources with different power spectra. The continuity in frequency of the spatial distribution of the sources is also assumed and exploited to accelerate the unsupervised learning in a large frequency band. Following similar lines, the method could be easily extended to exploit other types of diversities, such as source nonstationarity for instance.

## ACKNOWLEDGMENT

The authors would like to thank the reviewers for their valuable remarks.

- <sup>1</sup>E. G. Williams, *Fourier Acoustics: Sound Radiation and Near-Field Acoustical Holography* (Academic Press, San Diego, 1999), Chaps. 3, 5, 7.
- <sup>2</sup>S. F. Wu, “On reconstruction of acoustic pressure fields using helmholtz equation least squares method,” *J. Acoust. Soc. Am.* **107**, 2511–2522 (2000).
- <sup>3</sup>W. A. Verones, “Digital holographic reconstruction of sources with arbitrarily shaped surfaces,” *J. Acoust. Soc. Am.* **85**, 588–598 (1989).
- <sup>4</sup>Z. Zang, N. Vlahopoulos, S. T. Raveendra, T. Allen, and K. Y. Zhang, “A computational acoustic field reconstruction process based on an indirect boundary element formulation,” *J. Acoust. Soc. Am.* **108**, 2167–2178 (2000).
- <sup>5</sup>M. A. Tomlinson, “Partial source discrimination in near field acoustic holography,” *J. Acoust. Soc. Am.* **57**, 243–261 (1999).
- <sup>6</sup>K. U. Nam and Y. H. Kim, “Visualization of multiple incoherent sources by the backward prediction of near-field acoustic holography,” *J. Acoust. Soc. Am.* **109**, 1808–1816 (2001).
- <sup>7</sup>H. S. Kwon, Y. J. Kim, and J. S. Bolton, “Compensation for source nonstationarity in multireference, scan-based near-field acoustical holography,” *J. Acoust. Soc. Am.* **113**, 360–368 (2003).
- <sup>8</sup>Y. J. Kim, J. S. Bolton, and H. S. Kwon, “Partial sound field decomposition in multireference near-field acoustical holography by using optimally located virtual references,” *J. Acoust. Soc. Am.* **115**, 1641–1652 (2004).
- <sup>9</sup>M. H. Lee and J. S. Bolton, “Scan-based near-field acoustical holography and partial field decomposition in the presence of noise and source level variation,” *J. Acoust. Soc. Am.* **119**, 382–393 (2006).
- <sup>10</sup>K. U. Nam and Y. H. Kim, “A partial field decomposition algorithm and its examples for near-field acoustic holography,” *J. Acoust. Soc. Am.* **116**, 172–185 (2004).
- <sup>11</sup>J. F. Cardoso, “Blind signal separation: Statistical principles,” *Proc. IEEE* **90**, 2009–2026 (1998).
- <sup>12</sup>P. Comon and C. Jutten, *Handbook of Blind Source Separation, Independent Component Analysis and Applications* (Academic Press, Oxford, 2010), pp. 1–800.
- <sup>13</sup>A. Mohammad-Djafari, *Inverse Problems in Vision and 3D Tomography* (John Wiley and Sons, New York, 2010), pp. 1–467.

- <sup>14</sup>D. B. Rowe, *Multivariate Bayesian Statistics: Models for Source Separation and Signal Unmixing* (Taylor and Francis, Boca Raton, FL, 2003), Chap. 10.
- <sup>15</sup>T. Cemgil, C. Févotte, and S. J. Godsill, "Variational and stochastic inference for bayesian source separation," *Digital Signal Process.* **17**, 891–913 (2007).
- <sup>16</sup>A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation* (Society of Industrial and Applied Mathematics, Philadelphia, 2005), pp. 1–96.
- <sup>17</sup>D. MacKay, *Information Theory, Inference, and Learning Algorithms* (Cambridge University Press, Cambridge, 2003), Chap. 4.
- <sup>18</sup>H. L. V. Trees, *Optimum Array Processing* (John Wiley and Sons, New York, 2002), Chap. 5, pp. 332–415.
- <sup>19</sup>F. J. Fahy, *Sound Intensity* (Taylor and Francis, London, 1995), Chap. 4, pp. 38–85.
- <sup>20</sup>J. Antoni, "A Bayesian approach to sound source reconstruction: Optimal basis, regularization, and focusing," *J. Acoust. Soc. Am.* **131**, 2873–2890 (2012).
- <sup>21</sup>J. Hald, "Basic theory and properties of statistically optimized near-field acoustical holography," *J. Acoust. Soc. Am.* **125**, 2105–2120 (2009).
- <sup>22</sup>D. R. Brillinger, *Time Series: Data Analysis and Theory* (Society of Industrial and Applied Mathematics, Philadelphia, 2001), Chap. 4.
- <sup>23</sup>I. T. Nabney, *Netlab Algorithms for Pattern Recognition* (Springer Verlag, London, 2004), Chap. 9, pp. 325–366.
- <sup>24</sup>S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721–741 (1984).
- <sup>25</sup>A. Belouchrani, K. Abed-Meraim, J. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *Am. J. Math. Manage. Sci.* **45**, 434–444 (1997).