Mirabilii, Daniele; Schlecht, Sebastian J.; Habets, Emanuël A.P.

Generating coherence-constrained multisensor signals using balanced mixing and spectrally smooth filters

Published in:
Journal of the Acoustical Society of America

DOI:
10.1121/10.0003565

Published: 01/03/2021

Document Version
Publisher's PDF, also known as Version of record

Generating coherence-constrained multisensor signals using balanced mixing and spectrally smooth filters

Daniele Mirabilii,[1,a)] Sebastian J. Schlecht,[2,b)] and Emanuël A. P. Habets[1,c)]

[1]*International Audio Laboratories Erlangen, a joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Am Wolfsmantel 33, 91058 Erlangen, Germany*

[2]*Department of Signal Processing and Acoustics and Department of Media, Aalto University, Espoo, Finland*

ABSTRACT:

The spatial properties of a noise field can be described by a spatial coherence function. Synthetic multichannel noise signals exhibiting a specific spatial coherence can be generated by properly mixing a set of uncorrelated, possibly non-stationary, signals. The mixing matrix can be obtained by decomposing the spatial coherence matrix. As proposed in a widely used method, the factorization can be performed by means of a Choleski or eigenvalue decomposition. In this work, the limitations of these two methods are discussed and addressed. In particular, specific properties of the mixing matrix are analyzed, namely, the spectral smoothness and the mix balance. The first quantifies the mixing matrix-filters variation across frequency and the second quantifies the number of input signals that contribute to each output signal. Three methods based on the unitary Procrustes solution are proposed to enhance the spectral smoothness, the mix balance, and both properties jointly. A performance evaluation confirms the improvements of the mixing matrix in terms of objective measures. Furthermore, the evaluation results show that the error between the target and the generated coherence is lowered by increasing the spectral smoothness of the mixing matrix. https://doi.org/10.1121/10.0003565

(Received 4 November 2020; revised 1 February 2021; accepted 2 February 2021; published online 2 March 2021)

[Editor: Kainam Thomas Wong] Pages: 1425–1433

I. INTRODUCTION

A noise field recorded with multiple microphones can be analyzed in terms of spatial properties, such as the signals correlation or coherence. For some cases, these quantities can be approximated by analytical models. Cylindrical and spherical isotropic noise fields, referred to as two-dimensional (2D) (Cook *et al.*, 1955) and three-dimensional (3D) (Cron and Sherman, 1962) diffuse noise fields, respectively, can be thought of as generated by an infinite number of independent acoustic sources uniformly distributed on a cylinder or a sphere (Dal-Degan and Prati, 1988; Elko, 2000). These spatial coherence models are commonly time-invariant, real-valued, and dependent on the sensor positions (Elko, 2001). The 3D-diffuse noise field can be suitable to model the acoustic condition of a multi-talker environment, e.g., a crowded room (Jacobsen and Roisin, 2000).

In our recent work (Mirabilii and Habets, 2018), we analyzed the noise field recorded with closely spaced microphones that were exposed to the airflow generated by a stationary fan. The low-frequency spatial coherence of the aerodynamic noise agreed well with the empirical model identified by Corcos (Corcos, 1964). In Mirabilii *et al.* (2020), the noise field induced by atmospheric wind presented a similar coherence.

Generating synthetic noise allows us to evaluate dereverberation, noise reduction, and sound source localization algorithms in a controlled environment and circumvent the time-consuming collection of real data. This also avoids the inclusion of unwanted acoustic sources in the recordings. Examples of multi-channel dereverberation approaches that utilize synthetic noise signals with specific spatial properties can be found, e.g., in Dietzen *et al.* (2020) and Thüne and Enzner (2017). In Mirabilii and Habets (2019, 2020), synthetic wind noise exhibiting the Corcos model was used to evaluate the performance of two multi-channel wind noise reduction algorithms by adding the noise to the speech signals. In Mirabilii *et al.* (2020), the synthetic noise was used as a training set for a deep neural network approach that aimed at estimating the airflow speed and direction based on the spatial characteristics of the noise field.

In this respect, the method proposed in Habets *et al.* (2008) generates a set of signals that exhibit a predefined spatial coherence, which can be measured from real data or computed using analytical models. The signals are obtained by filtering and summing an equal number of uncorrelated noise instances in the short-time Fourier transform (STFT) domain. The filter weights for generating the correlated output signals from the uncorrelated input signals can be arranged in a so-called mixing matrix, i.e., the output signals are obtained by multiplying the input signals with the

a)Electronic mail: daniele.mirabilii@audiolabs-erlangen.de, ORCID: 0000-0002-9471-8683.
b)ORCID: 0000-0001-8858-4642.
c)ORCID: 0000-0002-2613-8046.

mixing matrix. The mixing matrix is computed by decomposing the spatial coherence matrix using an eigenvalue (EVD) or a Choleski (CHD) decomposition. The same method was used, e.g., in Mirabilii and Habets (2018) to simulate wind noise exhibiting the Corcos model and in Adrian *et al.* (2017) to generate perceptually plausible noise signals based on different statistical properties.

The method proposed in Habets *et al.* (2008), however, presents two main limitations. The first limitation is given by the CHD, which yields an unbalanced mix of the input signals, i.e., the output signals do not contain equal or similar proportions of each input signal. This is due to the fact that the CHD solution yields an upper triangular matrix such that, e.g., the first output signal is always equal to the first input. The balance problem was addressed in Habets *et al.* (2008) only for mixing two signals. The mix balance is relevant for perceptual reasons such as signal density or timbral similarity (Adami *et al.*, 2016), i.e., a balanced mix sounds more natural and plausible. The second limitation is given by the EVD, where the obtained eigenvectors can largely vary from one frequency band to the following, leading to discontinuities of the filters' magnitude response. This results in long filter responses and an increased error between the generated and the target spatial coherence. We refer to this property as spectral variation. In particular, we refer to mixing matrices exhibiting a slow-varying filters' response as spectral smooth in the following, where spectral smoothness denotes a low spectral variation.

In this work, we propose to extend the method in Habets *et al.* (2008) to increase the mix balance and decrease the spectral variation of the mixing matrix obtained with the CHD or the EVD for any arbitrary number of signals and microphone constellations. After defining a target spatial coherence matrix, the latter is decomposed via CHD or EVD. Subsequently, the mix balance, the spectral smoothness, or both properties of the obtained mixing matrix are enhanced using three proposed methods based on the unitary Procrustes solution (Gower and Dijksterhuis, 2004). Mutually uncorrelated input signals are then mixed using the obtained mixing matrix to yield output signals exhibiting the target spatial coherence.

The remainder of the paper is structured as follows. In Sec. II, we formulate the problem and introduce the notation. Section III describes the methods proposed in Habets *et al.* (2008) to generate multi-channel signals with a predefined spatial coherence. Section IV introduces objective measures to quantify the mix balance and the spectral smoothness of the mixing matrix. The limitations of the state-of-the-art (SoA) methods are described in terms of the above-mentioned properties. In Sec. V, three methods are proposed to respectively enhance the spectral smoothness, the mix balance, and both properties simultaneously. In Sec. VI, we evaluate the proposed methods in terms of mix balance and spectral smoothness. Section VII contains some conclusions.

II. PROBLEM FORMULATION

This section introduces the notation and the quantities of interest. In Sec. II A, we define the spatial coherence function, and we give some examples of analytical models. Section II B illustrates how to obtain the mixing matrix given a predefined coherence matrix. In Sec. II C, we explain the non-uniqueness of the matrix decomposition, which allows imposing desired properties on the mixing matrix.

A. Spatial coherence

The complex-valued spatial coherence is a measure of the correlation between the ith and jth microphone signals in the frequency domain, expressed as

$$\Gamma_{ij}(\omega) = \frac{\phi_{ij}(\omega)}{\sqrt{\phi_{ii}(\omega)\phi_{jj}(\omega)}}, \tag{1}$$

where ϕ_{ij} denotes the cross-power spectral density for $i \neq j$ and the auto-spectral density for $i = j$, and ω denotes the angular frequency. Examples of analytical models of Eq. (1) are the spherically isotropic noise field,

$$\Gamma_{ij}(\omega) = \frac{\sin(\omega d_{ij}/c)}{\omega d_{ij}/c}, \tag{2}$$

and the cylindrically isotropic noise field,

$$\Gamma_{ij}(\omega) = J_0(\omega d_{ij}/c), \tag{3}$$

where c denotes the sound velocity in ms^{-1}, d_{ij} denotes the distance between the ith and the jth microphones in m, and J_0 denotes the zero-order Bessel function of the first kind. An additional example is given by the spatial coherence of wind-noise contributions measured with closely spaced microphones (Mirabilii and Habets, 2018; Mirabilii *et al.*, 2020), expressed as

$$\Gamma_{ij}(\omega) = \exp\left(\frac{-\omega d_{ij}[\alpha(\theta_\mathrm{w}) - \imath\cos(\theta_\mathrm{w})]}{U_c}\right), \tag{4}$$

where $\imath = \sqrt{-1}$, $\alpha(\theta_\mathrm{w})$ denotes a decay-rate parameter which depends on the airflow direction θ_w, and U_c denotes the convective turbulence speed. The pair-wise coherence functions, e.g., Eqs. (2), (3), or (4), can be arranged in a matrix

$$\boldsymbol{\Gamma}(\omega) = \begin{bmatrix} \Gamma_{11}(\omega) & \cdots & \Gamma_{1N}(\omega) \\ \vdots & \ddots & \vdots \\ \Gamma_{N1}(\omega) & \cdots & \Gamma_{NN}(\omega) \end{bmatrix} \in \mathbb{C}^{N\times N}, \tag{5}$$

where N is the number of channels, $\boldsymbol{\Gamma}(\omega)$ is Hermitian, i.e., $\Gamma_{ij}(\omega) = \Gamma_{ji}^*(\omega)$, positive semi-definite and $|\Gamma_{ii}(\omega)| = 1$. As Eqs. (2), (3), and (4) are Hermitian symmetric across frequency, i.e., $\Gamma_{ij}(-\omega) = \Gamma_{ij}^*(\omega)$, the corresponding time-domain signals are real-valued.

B. Mixing matrix

We aim at generating N signals exhibiting a predefined coherence $\mathbf{\Gamma}(\omega)$ by filtering and summing N uncorrelated signals. Given N mutually uncorrelated, and possibly non-stationary, signals $\mathbf{X}(\omega) = [X_1(\omega), X_2(\omega), \ldots, X_N(\omega)]$, we obtain N signals $\mathbf{Y}(\omega) = [Y_1(\omega), Y_2(\omega), \ldots, Y_N(\omega)]$ exhibiting the predefined coherence by applying a mixing matrix $\mathbf{C}(\omega) \in \mathbb{C}^{N\times N}$, i.e.,

$$\mathbf{Y}(\omega) = \mathbf{C}^{\mathrm{H}}(\omega)\mathbf{X}(\omega), \tag{6}$$

where $(.)^{\mathrm{H}}$ denotes the Hermitian operator. The mixing matrix can be obtained by decomposing the spatial coherence matrix [Eq. (5)], such that

$$\mathbf{C}^{\mathrm{H}}(\omega)\mathbf{C}(\omega) = \mathbf{\Gamma}(\omega). \tag{7}$$

Because of $\mathbf{\Gamma}(\omega)$ being normalized, every column $\mathbf{C}_i(\omega)$ of the mixing matrix is a unit vector, i.e., $||\mathbf{C}_i(\omega)||_2 = 1$. This guarantees that the output signals $\mathbf{Y}(\omega)$ have equal power when the input signals $\mathbf{X}(\omega)$ have equal power. In the following, we assume homogeneous input signals, i.e., $\mathbf{X}(\omega)$ presents the same auto-power spectral density $\phi_x(\omega)$ for each signal. The covariance matrix of the mutually uncorrelated input signals is given by $\mathbf{\Phi}_x(\omega) = \mathbb{E}\{\mathbf{X}(\omega)\mathbf{X}^{\mathrm{H}}(\omega)\} = \phi_x(\omega)\mathbf{I}$, where $\mathbb{E}\{.\}$ denotes the mathematical expectation and $\mathbf{I} \in \mathbb{R}^{N\times N}$ denotes the identity matrix. It follows from Eq. (6) that the output signals $\mathbf{Y}(\omega)$ yield a covariance matrix $\mathbf{\Phi}_y(\omega) = \mathbb{E}\{\mathbf{Y}(\omega)\mathbf{Y}^{\mathrm{H}}(\omega)\} = \phi_x(\omega)\mathbf{\Gamma}(\omega)$. Therefore, the methods described in the following are limited to the simulation of homogeneous noise fields captured by unit-gain sensors exhibiting a predefined spatial coherence $\mathbf{\Gamma}(\omega)$. A method to simulate non-homogeneous signals exhibiting a predefined covariance matrix can be derived by replacing the desired spatial coherence matrix in Eq. (7) with a desired covariance matrix.

C. Unitary mixing matrix

The matrix decomposition in Eq. (7) is, however, not unique. This non-uniqueness can be exploited to impose desired properties on the mixing matrix. Specifically, given any unitary matrix $\mathbf{U} \in \mathbb{C}^{N\times N}$, i.e., $\mathbf{U}^{\mathrm{H}}\mathbf{U} = \mathbf{I}$, the equality $\mathbf{C}^{\mathrm{H}}\mathbf{C} = \mathbf{C}^{\mathrm{H}}\mathbf{U}^{\mathrm{H}}\mathbf{U}\mathbf{C} = \mathbf{\Gamma}$ holds. Therefore, the unitary transformation,

$$\hat{\mathbf{C}}(\omega) = \mathbf{U}(\omega)\mathbf{C}(\omega), \tag{8}$$

preserves the inner product in Eq. (7), so that $\hat{\mathbf{C}}$ and $\mathbf{C}$ yield the same target coherence $\mathbf{\Gamma}$.

One can modify the mixing matrix to fulfill specific requirements by using $\mathbf{U}$ as a design parameter. In this work, we aim at obtaining a mixing matrix that yields a balanced mix and spectral smoothness. These two properties are introduced in Secs. I A and IV B.

III. STATE OF THE ART

In this section, we summarize the matrix decomposition algorithms presented in Habets *et al.* (2008), where the authors introduce two different methods to decompose Eq. (5), namely, CHD and EVD. Whenever possible, we omit the frequency argument for brevity.

The CHD yields an upper triangle matrix $\mathbf{C}_{\mathrm{CHD}}(\omega)$, which solves Eq. (7) if $\mathbf{\Gamma}(\omega)$ is positive definite. The positive-definite property holds, e.g., for Eqs. (2), (3), and (4) at $\omega \neq 0$. A fully coherent matrix, i.e., $\Gamma_{ij}(\omega) = 1$ for all i and j, is rank-deficient and thus positive semi-definite. Therefore, the CHD is not applicable to a fully coherent matrix, for example, the one obtained using Eq. (4) for $\omega = 0$. In the latter case, the mixing matrix must be computed otherwise. In addition, the spatial response of a point source, which yields a rank-1 spatial coherence matrix irrespective of the source incident angle (Bienvenu and Kopp, 1981; Epain and Jin, 2016; Karasalo, 1986), cannot be generated with the CHD.

In contrast to the CHD, the EVD is applicable to positive semi-definite and rank-deficient matrices. The mixing matrix given by the EVD is computed as follows. First, the spatial coherence is decomposed as

$$\mathbf{\Gamma} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^{\mathrm{H}} = \mathbf{V}\sqrt{\mathbf{\Lambda}}\sqrt{\mathbf{\Lambda}}\mathbf{V}^{\mathrm{H}}, \tag{9}$$

where $\mathbf{\Lambda}$, $\mathbf{V} \in \mathbb{C}^{N\times N}$ are, respectively, the diagonal matrix whose elements are the eigenvalues of $\mathbf{\Gamma}$ and the matrix whose columns are the corresponding eigenvectors. Then, the mixing matrix is given by

$$\mathbf{C}_{\mathrm{EVD}}(\omega) = \sqrt{\mathbf{\Lambda}(\omega)}\mathbf{V}^{\mathrm{H}}(\omega). \tag{10}$$

IV. DESIGN CRITERIA AND SOA LIMITATIONS

In Secs. IV A and IV B, we propose novel requirements on the mixing matrix and we introduce the related objective measures. In particular, Sec. IV A introduces the concept of coherence error and spectral smoothness and describes their relation. Section IV B introduces the concept of mix balance. Subsequently, in Sec. IV C, we analyze the CHD and the EVD in terms of coherence error, spectral smoothness, and mix balance.

A. Design criterion: Spectral smoothness

The overall accuracy of the mixing matrix filter is given by the mean squared error (MSE),

$$\xi = \frac{1}{2\pi}\int_{-\pi}^{\pi} ||\mathbf{\Gamma}(\omega_n) - \mathbf{C}^{\mathrm{H}}(\omega_n)\mathbf{C}(\omega_n)||_{\mathrm{F}}^2 \, d\omega, \tag{11}$$

where $||.||_{\mathrm{F}}$ denotes the Frobenius norm of a matrix and ω_n is the angular frequency, ω, normalized by the sampling frequency. We refer to Eq. (11) as coherence error in the following. A discrete-time mixing filter can be designed by solving the mixing matrix at discrete frequency points ω_k, yielding $\mathbf{C}(\omega_k)$ at frequency instances $k = 0, \ldots, K-1$. An inverse discrete Fourier transform (DFT) $\mathcal{F}^{-1}\{\mathbf{C}(\omega_k)\}$ with length K yields a matrix of filters $\mathbf{C}(n)$ where $n = 0, \ldots, K-1$ denotes the time index.

The discrete solution can yield large errors in Eq. (11), especially between the frequency points ω_k. To illustrate this, let us assume the target spatial coherence matrix $\boldsymbol{\Gamma}(\omega) \equiv \mathbf{I}$. A possible mixing matrix solution is $\mathbf{C}(\omega_k) = \mathbf{U}_k$, where $\mathbf{U}_k$ are different unitary matrices for each k. If we, however, compute from $\mathbf{C}(n)$ the continuous frequency response via a discrete-time Fourier transform, $\mathbf{C}(\omega)$ for $\omega \neq \omega_k$ are not necessarily unitary and consequently $\boldsymbol{\Gamma}(\omega) \neq \mathbf{I}$. As a result, although $\mathbf{C}(\omega_k)$ are accurate, the MSE ξ in Eq. (11) can become large. In practice, we evaluate the MSE in Eq. (11) with a zero-padded DFT of length $K' \gg K$. In the given example, the generated coherence is inaccurate due to the discontinuities along frequency of the mixing matrix, i.e., due to a low spectral smoothness.

Therefore, we propose to choose the mixing matrices at adjacent frequency bands such that they are as similar as possible, i.e., we aim at minimizing

$$\epsilon = \frac{1}{K}\sum_k ||\mathbf{C}(\omega_k) - \mathbf{C}(\omega_{k-1})||_{\mathrm{F}}^2, \tag{12}$$

where $\omega_{-1} = \omega_{K-1}$. We refer to Eq. (12) as the spectral variation in the following. In particular, we refer to fast-varying mixing matrices along frequency as non-smooth matrices and to slowly varying mixing matrices as smooth matrices. In Sec. VI, we demonstrate that the spectral smoothness is a sufficient requirement for a low coherence error [Eq. (11)]. Additionally, smooth mixing matrices result in short filters $\mathbf{C}(n)$, which can be further truncated for efficient processing.

In Fig. 1, we show the difference between non-smooth and smooth mixing matrices in terms of coherence error and spectral smoothness, for $\boldsymbol{\Gamma}(\omega) \equiv \mathbf{I}$, $K = 1024$ and $N = 4$. The non-smooth solution is generated from random unitary matrices at each frequency instance ω_k. For the smooth solution, we applied the method used in Schlecht and Habets (2015), originally derived to achieve continuous feedback matrix modulation. We imposed smooth transitions, i.e., similar unitary matrices across frequency, between $\mathbf{C}(\omega_0) = \mathbf{I}$ and $\mathbf{C}(\omega_{K/2}) = \mathbf{H}/\sqrt{N}$, where $\mathbf{H}$ denotes a complex Hadamard matrix, e.g., a DFT matrix of size $N \times N$. The coherence error, Eq. (11), was computed by applying a DFT of length $K' = \lceil \pi \cdot K \rceil = 3217$ to $\mathbf{C}(n)$, i.e., $\mathcal{F}_{K'}\{\mathbf{C}(n)\}$, where π was chosen to obtain a sufficiently large non-integer multiple of K. The non-smooth solution presents long and discontinuous filters [Fig. 1(a)] and a large coherence error [Fig. 1(c)], whereas the smooth solution presents shorter filters [Fig. 1(b)] associated with a lower coherence error [Fig. 1(d)].

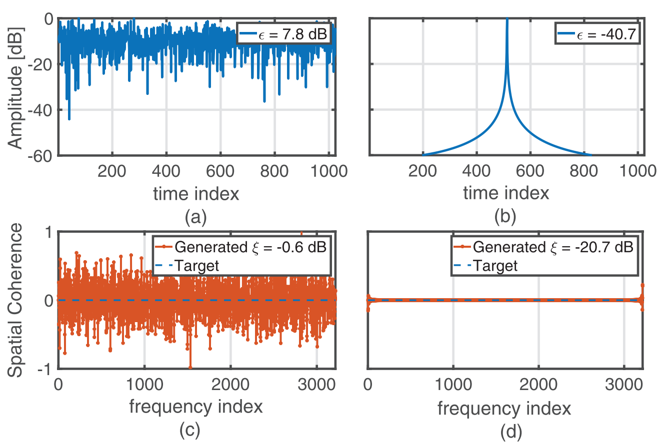

FIG. 1. (Color online) Filter response $C_{12}(n)$ obtained from (a) non-smooth solution with random unitary matrices and (b) smooth solution with unitary matrices varying slowly along frequency. The generated coherence (7) with $\mathcal{F}_{K'}\{C_{12}(n)\}$ is (c) and (d), respectively.

B. Design criterion: Mix balance

We require the mixing matrix to yield a balanced mix, i.e., similar contributions of the input signals at each output signal. As a matter of example, a target coherence $\boldsymbol{\Gamma}(\omega) = \mathbf{I}$ can occur, e.g., in Eq. (4) at sufficiently high frequency. For this case, an unbalanced mix can be given by $\mathbf{C} = \mathbf{I}$, where the output is the exact copy of the input in Eq. (6). An alternative mixing matrix can be given by $\mathbf{C} = \mathbf{H}/\sqrt{N}$. While $\mathbf{I}$ presents zeros outside the diagonal and thus each input signal contributes to only one output signal, $\mathbf{H}/\sqrt{N}$ is a dense matrix, i.e., most of its elements are non-zero, and therefore yields a balanced mix. To objectively evaluate the mix balance, one can compute the entry-wise l_1-norm normalized by the number of channels N of the mixing matrix

$$\beta = \frac{1}{KN\sqrt{N}}\sum_k ||\mathbf{C}(\omega_k)||_1, \tag{13}$$

where $||\mathbf{C}(\omega_k)||_1 = \sum_{i=1}^N \sum_{j=1}^N |C_{ij}(\omega_k)|$, and $C_{ij}(\omega_k)$ denotes the entries of $\mathbf{C}(\omega_k)$. The l_1-norm is an indicator of how dense the mixing matrix is and thus reflects the mix balance obtained with Eq. (6). In the above-mentioned example, in fact, the Hadamard matrix presents an l_1-norm that is higher than the identity matrix, i.e., $||\mathbf{I}||_1 = N < ||\mathbf{H}/\sqrt{N}||_1 = N\sqrt{N}$.

C. SOA limitations

As the CHD yields an upper triangle matrix, the resulting mix is unbalanced. In particular, the first output signal corresponds to the first input signal, while the Nth output signal is a linear combination of all the input signals. Therefore, the input signals do not similarly contribute to each output signal. The mixing matrix obtained with the EVD generally yields a mix that is more balanced than the CHD, but it does not guarantee similar contributions of the input signals when $N > 2$. The EVD, however, can yield an eigenvectors basis $\mathbf{V}(\omega_k)$ that is dissimilar at adjacent frequency instances. This leads to large variations across frequency and thus to pronounced discontinuities in the filter

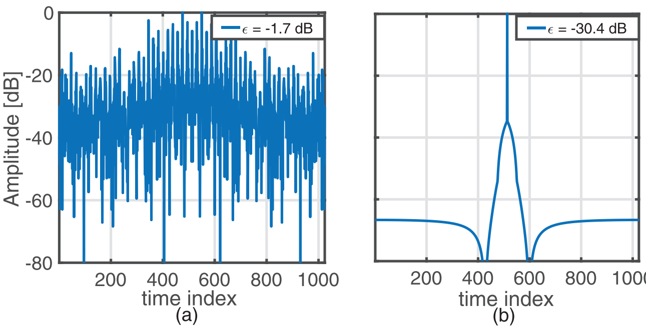

FIG. 2. (Color online) Filter responses $C_{12}(n)$ obtained with the EVD (a) and the CHD (b). The spectral variation [Eq. (12)] is shown in the legend.

responses compared to the mixing matrix obtained with the CHD.

To visualize the difference between the filters obtained with the CHD and the EVD in the time domain, we show the filter responses computed via the inverse DFT in Fig. 2. The spatial coherence was [Eq. (2)] for a uniform linear array with $N = 4$, $K = 1024$, $d_{i(i+1)} = 20$ cm, $F_s = 16\,\text{kHz}$. The CHD yielded a filter response that is significantly shorter than the filter response of the EVD. In Fig. 3, we compared the CHD and the EVD in terms of spectral smoothness and mix balance, using Eqs. (12) and (13), respectively. It is noticeable how the mixing matrix obtained with the EVD was discontinuous across the frequencies, whereas the CHD approached zero in terms of the difference between mixing matrices of adjacent bins. Moreover, the CHD resulted in an unbalanced mix compared to the EVD. Additionally, we can evaluate the MSE in Eq. (11) by applying a DFT of length $K' \gg K$ to $\mathbf{C}(n)$, as explained in Sec. IV A. In Fig. 4, the result is shown for $K' = \lceil \pi \cdot K \rceil$. The obtained coherence matrix $\mathbf{C}^{\mathrm{H}}(\omega_k)\mathbf{C}(\omega_k)$ was compared to the target coherence [Eq. (4)] computed with the increased frequency resolution K'. The coherence obtained with the EVD in Fig. 4(a) presented a high variance noise-like trend at in-between frequencies $\omega \neq \omega_k$ caused by the non-smoothness of the filters. The coherence obtained with the CHD in Fig. 4(b) agreed better with the target coherence, as the noise at in-between frequencies presented a reduced variance.

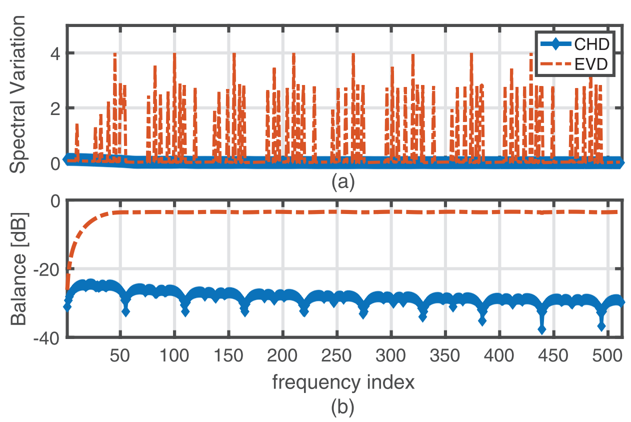

FIG. 3. (Color online) Frequency bin-wise spectral variation (a) and mix balance in dB (b) of the mixing matrix obtained with the CHD (blue lines) and the EVD (red dotted line).

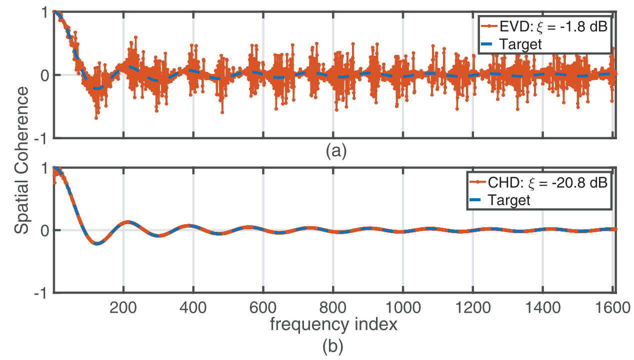

FIG. 4. (Color online) Target coherence $\Gamma_{12}(\omega_k)$ (dotted blue lines) compared to the coherence [Eq. (7)] obtained with the mixing matrix $\mathcal{F}_{K'}\{C_{12}(n)\}$ given by (a) the EVD and (b) the CHD (marked red lines). The coherence error [Eq. (11)] is shown in the legend.

In conclusion, the CHD yields a smooth mixing matrix with the drawback of an unbalanced mix, while the EVD yields a more balanced mix at the cost of discontinuous filters across the frequencies.

V. PROPOSED METHOD

The objective of this work is to obtain a mixing matrix from the decomposition of a target coherence matrix. In addition to what was proposed in Habets *et al.* (2008), we require that the mixing matrix is smooth across the frequencies and yields a balanced mix. These are conflicting requirements, since achieving mix balance could disrupt the spectral smoothness. The main idea is to fulfill these requirements by multiplying the mixing matrix $\mathbf{C}_{\mathrm{CHD}}$ or $\mathbf{C}_{\mathrm{EVD}}$ with an unitary matrix as explained in Sec. II C. In Sec. V A, we propose a method to enhance the spectral smoothness. In Sec. V B, we propose a method to enhance the mix balance. Finally, in Sec. V C, we propose a method to jointly enhance both properties.

A. Spectral smoothness

To minimize Eq. (12), we make use of the orthogonal (unitary) Procrustes problem (Gower and Dijksterhuis, 2004), which finds the unitary matrix that most closely maps one matrix to another. The problem can be formally expressed as

$$\hat{\mathbf{U}}_{\mathrm{f}}(\omega_k) = \min_{\mathbf{U}_{\mathrm{f}}} ||\mathbf{U}_{\mathrm{f}}\mathbf{C}(\omega_k) - \mathbf{C}(\omega_{k-1})||_{\mathrm{F}}, \tag{14}$$

subject to $\mathbf{U}_{\mathrm{f}}^{\mathrm{H}}\mathbf{U}_{\mathrm{f}} = \mathbf{I}$ with $\mathbf{U}_{\mathrm{f}} \in \mathbb{C}^{N\times N}$. The solution of Eq. (14) is given by the orthogonal polar factor of the matrix $\mathbf{R} = \mathbf{C}(\omega_{k-1})\mathbf{C}(\omega_k)^{\mathrm{H}}$, which is obtained from the singular value decomposition (SVD) of $\mathbf{R}$ as follows:

$$\mathbf{R} = \mathbf{W}\boldsymbol{\Sigma}\mathbf{Z}^{\mathrm{H}}, \tag{15}$$

where $\boldsymbol{\Sigma} = \mathrm{diag}[\sigma_1, \sigma_2, \ldots, \sigma_N]$ denotes the diagonal matrix containing the singular values σ_i of $\mathbf{R}$, and $\mathbf{W}$, $\mathbf{Z}$ are unitary matrices whose columns are the left-singular and right-singular vectors, respectively. Finally, the closed-form solution is given by

$$\hat{\mathbf{U}}_{\mathrm{f}}(\omega_k) = \mathbf{W}\mathbf{Z}^{\mathrm{H}}. \tag{16}$$

The spectral smoothness can be propagated from any initial frequency index $\hat{k}$ to the whole frequency range by applying Eq. (14) iteratively, i.e., we first solve Eq. (14) for $\hat{k}+1$, then for $\hat{k}+2$, etc., and vice versa in the negative direction. The mixing matrix is updated between each solution. In practice, we commonly choose $\hat{k} = K/2$ (for even K). The mixing matrix at the kth frequency bin is then computed as

$$\hat{\mathbf{C}}_{\mathrm{f}}(\omega_k) = \hat{\mathbf{U}}_{\mathrm{f}}(\omega_k)\mathbf{C}(\omega_k). \tag{17}$$

The resulting smooth mixing matrix is unique up to multiplication with a broadband unitary matrix, i.e., for any two matrices $\hat{\mathbf{C}}_{\mathrm{f}1}$ and $\hat{\mathbf{C}}_{\mathrm{f}2}$, there is a unitary matrix $\mathbf{U}$ such that $\hat{\mathbf{C}}_{\mathrm{f}2}(\omega_k) = \mathbf{U}\hat{\mathbf{C}}_{\mathrm{f}1}(\omega_k)$ for all k. The choice of $\mathbf{U}$ does not affect the spectral smoothness.

This method can be applied, e.g., to the mixing matrices obtained with the EVD to attenuate the discontinuity along the frequencies shown in Fig. 3(a) and the coherence error shown in Fig. 4(a).

B. Mix balance

To maximize the mix balance [Eq. (13)], we propose the following iterative algorithm. The maximum value of the l_1-norm is attained by the unitary Hadamard matrix $\mathbf{H}$. Therefore, we are interested in finding the unitary matrix that most closely maps the mixing matrix to $\mathbf{H}$ in terms of element-wise absolute values. This can be achieved by minimizing the following cost function:

$$\hat{\mathbf{U}}_{\mathrm{b}}(\omega_k) = \min_{\mathbf{U}_{\mathrm{b}}} |||\mathbf{U}_{\mathrm{b}}\mathbf{C}(\omega_k)| - |\mathbf{H}|||_{\mathrm{F}}, \tag{18}$$

subject to $\mathbf{U}_{\mathrm{b}}^{\mathrm{H}}\mathbf{U}_{\mathrm{b}} = \mathbf{I}$ with $\mathbf{U}_{\mathrm{b}} \in \mathbb{C}^{N\times N}$, where $|.|$ denotes the element-wise absolute values of a matrix. The minimization in Eq. (18) is equivalent to

$$\hat{\mathbf{U}}_{\mathrm{b}}(\omega_k) = \min_{\mathbf{U}_{\mathrm{b}}} ||\mathbf{U}_{\mathrm{b}}\mathbf{C}(\omega_k) - \mathbf{P}||_{\mathrm{F}}, \tag{19}$$

where $\mathbf{P}$ is the optimal set of phases, such that $|P_{ij}| = 1$, where P_{ij} denotes the ith row and jth column of $\mathbf{P}$. A similar method was used in Schlecht and Habets (2018) for designing the spatial configuration of feedback delay networks. We propose to solve Eq. (19) iteratively by using the unitary Procrustes solution (Gower and Dijksterhuis, 2004). At the mth iteration, the unitary matrix is given by

$$\hat{\mathbf{U}}_{\mathrm{b}}^{(m)}(\omega_k) = \min_{\mathbf{U}_{\mathrm{b}}} ||\mathbf{U}_{\mathrm{b}}\mathbf{C}(\omega_k) - \hat{\mathbf{P}}^{(m-1)}||_{\mathrm{F}}, \tag{20}$$

where $\hat{\mathbf{P}}$ is the set of phases updated as

$$\hat{\mathbf{P}}^{(m)} = \angle\left(\hat{\mathbf{U}}_{\mathrm{b}}^{(m)}(\omega_k)\mathbf{C}(\omega_k)\right), \tag{21}$$

where $\angle(.)$ denotes the element-wise phase of a matrix.

The proof of convergence is given by the Procrustes solution being the global minimum of the difference [Eq. (20)] in the Frobenius-norm sense, similar to Eq. (14). Therefore,

$$||\hat{\mathbf{U}}_{\mathrm{b}}^{(m)}\mathbf{C} - \hat{\mathbf{P}}^{(m-1)}||_{\mathrm{F}} \le ||\hat{\mathbf{U}}_{\mathrm{b}}^{(m-1)}\mathbf{C} - \hat{\mathbf{P}}^{(m-1)}||_{\mathrm{F}}, \tag{22}$$

and

$$||\hat{\mathbf{U}}_{\mathrm{b}}^{(m)}\mathbf{C} - \hat{\mathbf{P}}^{(m)}||_{\mathrm{F}} \le ||\hat{\mathbf{U}}_{\mathrm{b}}^{(m-1)}\mathbf{C} - \hat{\mathbf{P}}^{(m-1)}||_{\mathrm{F}}. \tag{23}$$

The balanced mixing matrix is finally obtained as

$$\hat{\mathbf{C}}_{\mathrm{b}}(\omega_k) = \hat{\mathbf{U}}_{\mathrm{b}}(\omega_k)\mathbf{C}(\omega_k). \tag{24}$$

The solution of Eq. (18) is a local minimum and represents a proxy measure for the mix balance, whose results are evaluated in terms of Eq. (13) in Sec. VI. The balancing step can be skipped, e.g., when the input signals are given by realizations of zero-mean random Gaussian processes, since a balanced mix of noise-like input signals does not exhibit substantial perceptual differences compared to an unbalanced mix.

C. Joint spectral smoothness and mix balance

One can use the method proposed in Secs. V A or V B to individually increase the spectral smoothness or the mix balance of the obtained mixing matrix, respectively. However, one might be interested in jointly enhancing both properties at the same time. Solving simultaneously Eqs. (14) and (19) to achieve smoothness and balance by finding a single unitary matrix (i.e., $\hat{\mathbf{U}}_{\mathrm{b}} = \hat{\mathbf{U}}_{\mathrm{f}}$) could lead to a cumbersome nonconvex optimization problem. Moreover, applying a cascade of Eqs. (14) and (19) replaces the obtained balanced-mixing matrix with the smooth-mixing matrix, neglecting the benefit

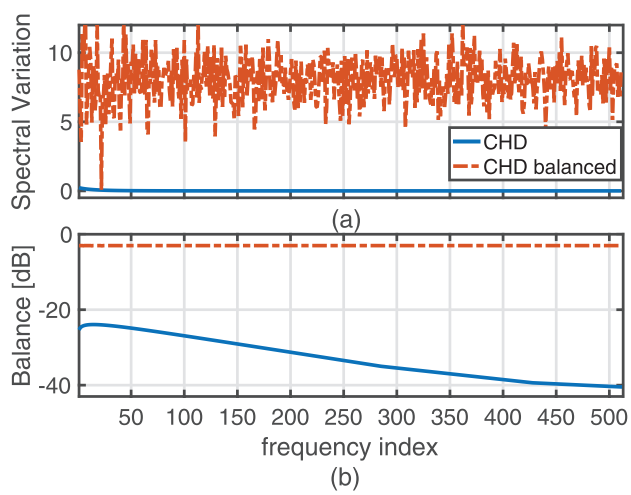

FIG. 5. (Color online) Frequency bin-wise spectral variation (a) and mix balance (b) of the mixing matrix obtained with the CHD (solid blue lines), compared to the mixing matrix obtained with CHD and Eq. (24) (dashed red lines).

of the balance algorithm. Conversely, a cascade of Eqs. (14) and (19) could disrupt the spectral smoothness.

As a matter of example, let us decompose, e.g., Eq. (4) using the CHD. As shown in Fig. 3, the CHD yields smooth filters but an unbalanced mix. Therefore, we can induce balance by applying Eq. (24) to $\mathbf{C}$ and obtain $\hat{\mathbf{C}}_{\mathrm{b}}$. The balance of $\mathbf{C}$ and $\hat{\mathbf{C}}_{\mathrm{b}}$ is shown in Fig. 5(b). On one hand, the method proposed in Sec. V B yields a mixing matrix with maximum balance. On the other hand, while $\mathbf{C}$ was smooth, $\hat{\mathbf{C}}_{\mathrm{b}}$ results as highly discontinuous after applying Eq. (24), in Fig. 5(a).

Therefore, we propose a method to jointly increase the mix balance and the spectral smoothness of the mixing matrix. This is performed by alternating the spectral smoothness [Eq. (14)] and a smoothness-preserving version of Eq. (19). First, we apply Eq. (19) to the mixing matrix at any initial frequency $\omega_{\hat{k}}$, i.e., we compute $\hat{\mathbf{C}}_{\mathrm{b}}(\omega_{\hat{k}})$ to maximize the balance of the initial state. Then, we solve Eq. (17) obtaining $\hat{\mathbf{C}}_{\mathrm{f}}(\omega_k)$ at the neighbour frequency band. Subsequently, we solve

$$\hat{\mathbf{U}}_{\mathrm{b}} = \min_{\mathbf{U}_{\mathrm{b}}} ||\mathbf{U}_{\mathrm{b}}\hat{\mathbf{C}}_{\mathrm{f}}(\omega_k) - \mathbf{P}_{\mathrm{f}}||_{\mathrm{F}}, \tag{25}$$

by iterating Eq. (20) one time, where $\hat{\mathbf{P}}_{\mathrm{f}}$ is deterministic and defined as the phases of $\hat{\mathbf{C}}_{\mathrm{f}}(\omega_k)$, i.e.,

$$\hat{\mathbf{P}}_{\mathrm{f}} = \angle\Big(\hat{\mathbf{C}}_{\mathrm{f}}(\omega_k)\Big). \tag{26}$$

After solving Eq. (20) using the unitary Procrustes solution, we obtain the smoothness-preserving balanced matrix,

$$\hat{\mathbf{C}}_{\mathrm{fb}}(\omega_k) = \hat{\mathbf{U}}_{\mathrm{b}}\hat{\mathbf{C}}_{\mathrm{f}}(\omega_k). \tag{27}$$

The choice of iterating one time is justified by the fact that the algorithm in Eq. (20) converges rapidly if the mixing matrix is already dense. This is true, e.g., when we impose spectral smoothness [Eq. (18)] between the current and the previous frequency bin of the mixing matrix, to which we previously imposed balance. In other words, imposing smoothness propagates the balance.

In Fig. 6, we show the trade-off between the spectral smoothness and the mix balance averaged over the frequency bins at each iteration of Eq. (20). By smoothness we refer to Eq. (12), where a small value is desired. By balance we refer to Eq. (13), where a large value is desired. The measures on the abscissa origin (marked with triangles) correspond to the matrix obtained solely with the CHD. The following values (circles) correspond to the matrix obtained with the CHD where we imposed smoothness using Eq. (17) from the maximum to the minimum frequency and by applying Eq. (19) to the initial state. It is noticeable how inducing smoothness between a balanced matrix and the neighbour matrix has the side effect of propagating the balance. The third point (squares), corresponds to the measures of the matrix obtained with a single iteration, i.e., using Eq. (27). From the third point on, the axis corresponds to the mth iteration in Eq. (20). The best trade-off is attained at the first iteration [Eq. (27)], where the balance is close to the maximum while the Frobenius norm of the difference between adjacent matrices is close to the minimum attained by Eq. (17), i.e., the spectral smoothness is reasonably high. By iterating more than once, the balance algorithm increasingly degrades the spectral smoothness without further enhancing the balance. In addition, using a single iteration to obtain Eq. (27) is beneficial in terms of computational efficiency. Audio examples of the SoA and the proposed methods can be found at the AudioLabs website (AudioLabs, 2020).

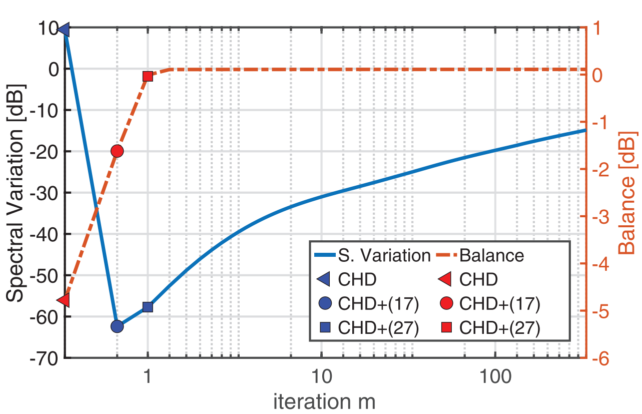

FIG. 6. (Color online) Spectral variation (in solid blue) and mix balance (in dashed red) against the iterations of Eq. (20).

VI. PERFORMANCE EVALUATION

In this section we evaluated the mixing matrix obtained with the proposed methods, i.e., Eqs. (17), (24), and (27), in terms of mix balance, spectral smoothness, and coherence error. We compared the results with the original mixing matrix obtained with CHD and EVD to assess the improvements. In the following, the measures are expressed in dB. We expect $\beta = 0\,\mathrm{dB}$ [Eq. (13)] for maximum balance or close to 0 dB for well-balanced matrices, while ξ [Eq. (11)] and ϵ [Eq. (12)] as small as possible for smooth matrices.

We simulated the spatial response of a uniform linear array with $N = 4$, $K = 1024$, $F_s = 16\,\mathrm{kHz}$, and a variable inter-microphone distance. The coherence matrix was defined with the models in Eqs. (2), (3), and (4). For Eqs. (2) and (3), we computed the performance measures for an inter-microphone distance of $d_{i(i+1)} = \{10, 20, 40\}$ cm. For Eq. (4), the distance was kept fixed at 0.5 cm, while the stream direction was $\theta_{\mathrm{w}} \in \{0, \pi/4, \pi/2\}$ rad and the convective turbulence speed was $U_c \in \{3, 6, 12\}\,\mathrm{ms}^{-1}$.

The results are shown in Table I, where we averaged over the microphone distance for Eqs. (2) and (3), and over the stream speed and direction for Eq. (4). The mixing matrix obtained with the CHD (Habets *et al.*, 2008), presented overall a low spectral variation and a low coherence error, but it was poorly balanced. Vice versa, the EVD (Habets *et al.*, 2008) presented an increased mix balance at the cost of a high spectral variation compared to the CHD. To decrease the spectral variation and the coherence error, we applied Eq. (17) to both CHD and EVD from the maximum to the minimum frequency. The obtained mixing matrices yielded overall the minimum spectral variation and

TABLE I. Mix balance (β), spectral variation (ϵ) and coherence error (ξ) of different mixing matrices. All measures are expressed in dB. The coherence matrix was defined with the spherical [Eq. (2)] and cylindrical [Eq. (3)] isotropic diffuse model, using an increasing inter-microphone distance, and with the Corcos model [Eq. (4)], using different stream direction θ_w and convective turbulence speed U_c. The results were then averaged over the considered set of microphone spacing and the set of direction/speed, respectively.

		Spherical			Cylindrical			Corcos		
		β (13)	ϵ (12)	ξ (11)	β (13)	ϵ (12)	ξ (11)	β (13)	ϵ (12)	ξ (11)
	CHD	−5.72	−32.92	−17.25	−4.92	−25.97	−16.73	−5.62	−28.15	−17.56
	EVD	−0.54	−1.22	−2.12	−0.57	3.15	−0.56	−0.32	−1.14	−5.14
Smoothness	CHD + [Eq. (17)]	−5.43	**−62.92**	**−21.51**	−4.72	**−54.29**	**−19.97**	−5.33	**−46.24**	**−18.66**
	EVD + [Eq. (17)]	−0.48	−62.43	−20.33	−0.44	−53.96	−18.78	−0.31	−45.41	−18.31
Balance	CHD + [Eq. (24)]	**0.00**	6.01	0.12	**0.00**	6.03	0.12	**0.00**	6.01	0.11
	EVD + [Eq. (24)]	**0.00**	6.01	0.12	**0.00**	6.02	0.09	**0.00**	6.01	0.12
Joint	CHD + [Eq. (27)]	−0.01	−58.62	−19.42	−0.03	−52.12	−19.76	−0.03	−42.76	−18.28
	EVD + [Eq. (27)]	−0.01	−55.9	−18.92	−0.02	−51.45	−18.17	−0.04	−44.82	−18.15

coherence error without any significant impact on the mix balance. To enhance the mix balance, we applied Eq. (24) to the matrices obtained with the CHD and the EVD. The algorithm was iterated until convergence. Although the l_1-norm was overall maximized, the balance algorithm significantly increased the spectral variation and the coherence error yielding the worst performing results, as shown in Fig. 5. Finally, to jointly increase the mix balance and decrease the spectral variation along with the coherence error, we applied Eq. (27) to the CHD and EVD. The spectral smoothness was imposed from the maximum to the minimum frequency, where the matrix balance of the initial state was maximized using Eq. (19). The proposed smoothness-preserving balance algorithm described in Sec. V C consisted in a single iteration of the absolute unitary Procrustes solution [Eq. (20)] and resulted in a good trade-off between a balanced and a smooth matrix. The l_1-norm of the obtained matrix was in fact close to the maximum, while the spectral smoothness and the coherence error presented values of Eqs. (12) and (11) relatively close to the cohesion-algorithm results, as shown in Fig. 6. Optionally, one could increase the mix balance at the cost of increasing the spectral variation by iterating more than once. As hypothesized in Sec. IV A, increasing the spectral smoothness [Eq. (12)] of the mixing matrix was sufficient to decrease the coherence error [Eq. (11)] for every spatial coherence model.

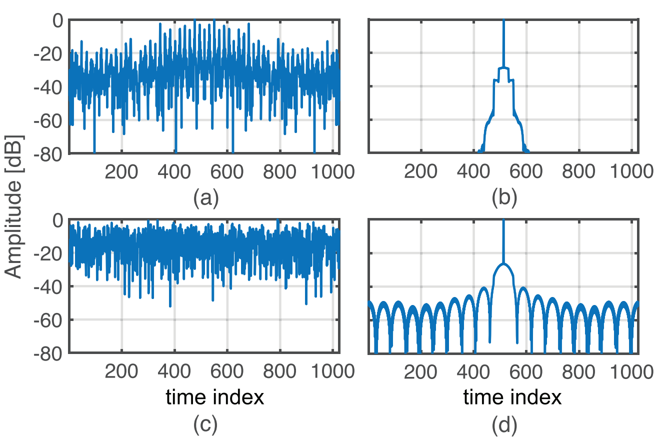

FIG. 7. (Color online) Filter responses $C_{12}(n)$ of the mixing matrix obtained with the (a) EVD, (b) EVD + Eq. (17), (c) EVD + Eq. (24), and (d) EVD + Eq. (27).

In Fig. 7, we show the filter responses of the mixing matrix obtained with the above-mentioned methods. The filter responses were computed by applying the inverse DFT to the mixing matrix. The spatial coherence model was Eq. (2) with $N = 2$, $d = 4$ cm and $K = 1024$. In Fig. 7(a), the mixing matrix obtained with the EVD presented a long and discontinuous filter response's magnitude caused by the low spectral smoothness. The smoothness-inducing algorithm [Eq. (17)] applied to the EVD-based mixing matrix yielded a shorter and smoother filter, as shown in Fig. 7(b). The balance-inducing algorithm [Eq. (24)] significantly disrupted the spectral smoothness, yielding a highly discontinuous filter, as shown in Fig. 7(c). Finally, the smoothness-preserving balance algorithm [Eq. (24)] yielded a filter with a response comparable to Fig. 7(b), shown in Fig. 7(d).

VII. CONCLUSION

In this work, we proposed three methods to enhance specific properties of the mixing matrix obtained by decomposing a spatial coherence matrix. In addition to attaining a low error between the target and the generated coherence, our goal was to obtain a mixing matrix that yields spectral smoothness, i.e., small variations of the filters' magnitude response across frequency, and a balanced mix, i.e., similar contributions of the input signals at each output signal. We introduced these additional requirements by means of objective measures and demonstrated that the spectral smoothness is a sufficient requirement for a low coherence error. We developed a method to enhance the spectral smoothness, the mix balance, and both properties jointly. An evaluation of the proposed methods showed that (a) enhancing the smoothness of the mixing matrix lowers the coherence error with no significant impact on the mix balance, (b) enhancing the balance disrupts significantly the smoothness and the coherence error, and (c) jointly enhancing the smoothness and the balance yields results, which were close to the individually enhanced properties. The third method attained the best trade-off in terms of spectral smoothness and mix balance.

Adami, A., Disch, S., Steba, G., and Herre, J. (**2016**). "Assessing applause density perception using synthesized layered applause signals," in *Proceedings of the Conference on Digital Audio Effects*, September 5–9, Brno, Czech Republic, pp. 183–189.

Adrian, J.-A., Gerkmann, T., van de Par, S., and Bitzer, J. (**2017**). "Synthesis of perceptually plausible multichannel noise signals controlled by real world statistical noise properties," J. Audio Eng. Soc. **65**(11), 914–928.

AudioLabs (**2020**). https://www.audiolabs-erlangen.de/resources/2020-JASA-CCR (Last viewed January 30, 2021).

Bienvenu, G., and Kopp, L. (**1981**). "Adaptive high resolution spatial discrimination of passive sources," in *Underwater Acoustics and Signal Processing*, edited by L. Bjørnø (Springer Netherlands, Dordrecht), pp. 509–515.

Cook, R. K., Waterhouse, R. V., Berendt, R. D., Edelman, S., and Thompson, M. C. (**1955**). "Measurement of correlation coefficients in reverberant sound fields," J. Acoust. Soc. Am. **27**(6), 1072–1077.

Corcos, G. (**1964**). "The structure of the turbulent pressure field in boundary-layer flows," J. Fluid Mech. **18**(3), 353–378.

Cron, B. F., and Sherman, C. H. (**1962**). "Spatial-correlation functions for various noise models," J. Acoust. Soc. Am. **34**(11), 1732–1736.

Dal-Degan, N., and Prati, C. (**1988**). "Acoustic noise analysis and speech enhancement techniques for mobile radio applications," Signal Process. **15**, 43–56.

Dietzen, T., Doclo, S., Moonen, M., and van Waterschoot, T. (**2020**). "Integrated sidelobe cancellation and linear prediction Kalman filter for joint multi-microphone speech dereverberation, interfering speech cancellation, and noise reduction," IEEE Trans. Audio, Speech, Lang. Process. **28**, 740–754.

Elko, G. W. (**2000**). "Superdirectional microphone arrays," in *Acoustic Signal Processing for Telecommunication*, edited by S. L. Gay and J. Benesty (Kluwer Academic, Hingham, MA), Chap. 10, pp. 181–237.

Elko, G. W. (**2001**). "Spatial coherence functions," in *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. Brandstein and D. Ward (Springer-Verlag, New York), Chap. 4, pp. 61–85.

Epain, N., and Jin, C. T. (**2016**). "Spherical harmonic signal covariance and sound field diffuseness," IEEE Trans. Audio Speech Lang. Process. **24**(10), 1796–1807.

Jacobsen, F., and Roisin, T. (**2000**). "The coherence of reverberant sound fields," J. Acoust. Soc. Am. **108**, 204–210.

Gower, J. C., and Dijksterhuis, G. B. (**2004**). *Procrustes Problems*, Vol.30 (Oxford University Press, Oxford, UK).

Habets, E. A., Cohen, I., and Gannot, S. (**2008**). "Generating nonstationary multisensor signals under a spatial coherence constraint," J. Acoust. Soc. Am. **124**(5), 2911–2917.

Karasalo, I. (**1986**). "Estimating the covariance matrix by signal subspace averaging," IEEE Trans. Acoust. Speech Signal Process. **34**(1), 8–12.

Mirabilii, D., and Habets, E. A. P. (**2018**). "Simulating multi-channel wind noise based on the Corcos model," in *Proceedings of the International Workshop on Acoustic Signal Enhancement (IWAENC)*, September 17–20, Tokyo, Japan.

Mirabilii, D., and Habets, E. (**2019**). "Multi-channel wind noise reduction using the Corcos model," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 12–17, Brighton, UK, pp. 646–650.

Mirabilii, D., and Habets, E. (**2020**). "Spatial coherence-aware multi-channel wind noise reduction," IEEE Trans. Audio Speech Lang. Process. **28**, 1974–1987.

Mirabilii, D., Lakshminarayana, K. K., Mack, W., and Habets, E. A. (**2020**). "Data-driven wind speed estimation using multiple microphones," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 4–8, Barcelona, Spain, pp. 576–580.

Schlecht, S. J., and Habets, E. A. P. (**2015**). "Time-varying feedback matrices in feedback delay networks and their application in artificial reverberation," J. Acoust. Soc. Am. **138**, 1389–1398.

Schlecht, S. J., and Habets, E. A. P. (**2018**). "Sign-agnostic matrix design for spatial artificial reverberation with feedback delay networks," in *Proceedings of the AES International Conference on Spatial Reproduction-Aesthetics and Science*, August 7–9, Tokyo, Japan.

Thüne, P., and Enzner, G. (**2017**). "Maximum-likelihood approach with Bayesian refinement for multichannel-wiener postfiltering," IEEE Trans. Signal Process. **65**(13), 3399–3413.