Generating nonstationary multisensor signals under a spatial coherence constraint

Emanuël A. P. Habets[a)]
School of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel, and Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel

Israel Cohen[b)]
Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel

Sharon Gannot[c)]
School of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel

(Received 27 June 2008; revised 20 August 2008; accepted 20 August 2008)

Noise fields encountered in real-life scenarios can often be approximated as spherical or cylindrical noise fields. The characteristics of the noise field can be described by a spatial coherence function. For simulation purposes, researchers in the signal processing community often require sensor signals that exhibit a specific spatial coherence function. In addition, they often require a specific type of noise such as temporally correlated noise, babble speech that comprises a mixture of mutually independent speech fragments, or factory noise. Existing algorithms are unable to generate sensor signals such as babble speech and factory noise observed in an arbitrary noise field. In this paper an efficient algorithm is developed that generates multisensor signals under a predefined spatial coherence constraint. The benefit of the developed algorithm is twofold. Firstly, there are no restrictions on the spatial coherence function. Secondly, to generate M sensor signals the algorithm requires only M mutually independent noise signals. The performance evaluation shows that the developed algorithm is able to generate a more accurate spatial coherence between the generated sensor signals compared to the so-called image method that is frequently used in the signal processing community. © *2008 Acoustical Society of America.* [DOI: 10.1121/1.2987429]

PACS number(s): 43.50.Ed, 43.60.Cg [EJS] Pages: 2911–2917

I. INTRODUCTION

A spherical noise field has been shown to be a reasonable model for a number of practical noise fields that can be found in, for example, an office or a car.[1] Cylindrical noise fields are especially useful when, for example, the ceiling and floor in an enclosure are covered with a highly absorbing material.[1] Spherical and cylindrical noise fields are also known as three-dimensional (3D) and two-dimensional (2D) diffuse noise fields, respectively. Researchers in the signal processing community often require sensor signals that result from these noise fields for simulation purposes, e.g., for (superdirective) beamforming,[2–5] adaptive noise cancellation,[6,7] and source localization. In some cases it might even be desired to generate sensor signals that exhibit a specific spatial coherence function, e.g., based on a specific measurement or condition.

It is often assumed that the noise field is (i) spatially homogeneous, i.e., the physical properties of the sound do not depend on the absolute position of the sensor, (ii) isotropic, i.e., the physical properties of the sound are the same in any direction of measurement, and (iii) time invariant.[1,8] The sensor signals acquired in 2D and 3D diffuse noise fields can be generated using a number of independent noise sources that are uniformly spaced on a cylinder or sphere, respectively.[6,8,9] Recently, we have developed an efficient algorithm to generate sensor signals acquired in noise fields that satisfy the above assumptions.[10]

In many cases noise comprises a mixture of independent speech fragments, also known as babble speech, or factory noise. Babble speech can be used to model the background noise encountered in a multitalker environment such as a restaurant or cafeteria. In such a case the short-term power spectral densities (PSDs) of the sensor signals vary in space and time. This relaxes some of the prior assumptions. The algorithm in Ref. 10 can be used to generate stationary and nonstationary signals. However, the algorithm requires a large number of independent noise signals. Furthermore, the spatial coherence depends on the positions of the noise sources. Therefore, it is difficult to obtain an arbitrary spatial coherence. In this paper, we develop an efficient algorithm that generates nonstationary sensor signals under a predefined spatial coherence constraint. The benefit of the developed algorithm is twofold. Firstly, there are no restrictions on the spatial coherence function, allowing the use of an arbitrary or measured spatial coherence function. For example, generated sensor signals with slightly different spatial coherence functions can be used to analyze the robustness and the sensitivity of acoustic signal processing algorithms. Secondly, to generate M sensor signals the developed algo-

[a)]Electronic mail: e.habets@ieee.org. URL: http://home.tiscali.nl/ehabets
[b)]Electronic mail: icohen@ee.technion.ac.il. URL: http://www.ee.technion.ac.il/Sites/People/IsraelCohen
[c)]Electronic mail: gannot@eng.biu.ac.il. URL: http://www.eng.biu.ac.il/~gannot

rithm only requires M mutually independent noise signals, which is a number that is significantly smaller than the number of noise signals that is used by other algorithms.[9,10] The developed algorithm generates the sensor signals in two steps. Firstly, we generate a set of mutually independent noise signals. The cardinality of this set is equal to the number of sensors. Secondly, the noise signals are filtered and mixed such that the obtained sensor signals exhibit a predefined spatial coherence. The filtering and mixing can be performed in the frequency domain or in the time-frequency domain such that these operations reduce to an instantaneous mixing.

The remainder of this paper is organized as follows: In Sec. II, we formulate the problem of generating sensor signals under a predefined spatial coherence. The sensor signals are expressed as instantaneous mixing of mutually independent noise signals. In Sec. III, we generate the mutually independent noise signals and in Sec. IV we compute the instantaneous mixing matrix. In Sec. V we summarize the algorithm and discuss its computational complexity. In Sec. VI, we evaluate the performance in different noise fields by comparing the spatial coherence of the generated sensor signals with the theoretical spatial coherence.

II. PROBLEM FORMULATION

Our objective is to generate M sensor signals with a predefined spatial coherence. The M sensor positions, relative to the first sensor position, are stacked into a matrix $\mathbf{P}$, such that

$$\mathbf{P} = \begin{bmatrix} 0 & x_2 & \cdots & x_M \\ 0 & y_2 & \cdots & y_M \\ 0 & z_2 & \cdots & z_M \end{bmatrix}. \tag{1}$$

The Euclidian distance d_{pq} between the pth and the qth sensor is given by

$$d_{pq} = \|\mathbf{P}_p - \mathbf{P}_q\|_2, \tag{2}$$

where $\mathbf{P}_p$ denotes the pth column of the matrix $\mathbf{P}$.

Let us denote the PSD of the pth sensor signal by $\Phi_{pp}(\omega)$, where ω denotes the angular frequency. The cross-PSD between the pth and the qth sensor is denoted by $\Phi_{pq}(\omega)$.

The assumption that the noise field is homogeneous can be formulated as

$$\Phi_{pp}(\omega) = \Phi(\omega), \quad \forall\, p \in \{1, \ldots, M\}. \tag{3}$$

The spatial coherence between the pth and the qth sensor is defined as[8]

$$\gamma_{pq}(\omega) = \frac{\Phi_{pq}(\omega)}{\sqrt{\Phi_{pp}(\omega)\Phi_{qq}(\omega)}}. \tag{4}$$

In a spherically isotropic noise field the spatial coherence function is given by[8,11]

$$\gamma_{pq}(\omega) = \frac{\sin(\omega d_{pq}/c)}{\omega d_{pq}/c}, \tag{5}$$

where d_{pq} denotes the distance between the pth and qth sensors, and c denotes the sound velocity in ms^{-1}. Another well-known noise field is a cylindrically isotropic noise field. The spatial coherence function is then given by[12]

$$\gamma_{pq}(\omega) = J_0(\omega d_{pq}/c), \tag{6}$$

where $J_0(\cdot)$ is a zero-order Bessel function of the first kind.

We propose to filter and mix M mutually independent noise signals to generate M sensor signals that exhibit a predefined spatial constraint. Since the spatial coherence is defined in the frequency domain it is preferred to work in this domain. The filtering and mixing can efficiently be performed in the frequency domain and in the time-frequency domain. Here, we work in the short-time Fourier transform (STFT) domain. The frame index is denoted by ℓ and the discrete angular frequencies are denoted by ω_k, where $(k \in \{0, \ldots, K/2-1\})$ and K is the frame length of the STFT. It should be noted that the spatial coherence can either be (slowly) time varying or time invariant. While the developed algorithm can be used to simulate a (slowly) time varying spatial coherence, we assume that the predefined spatial coherence is time invariant.

Let us define a matrix $\tilde{\mathbf{\Gamma}}(\omega_k)$ for each ω_k that consists of the predefined spatial coherence values

$$\tilde{\mathbf{\Gamma}}(\omega_k) = \begin{bmatrix} \tilde{\gamma}_{11}(\omega_k) & \tilde{\gamma}_{12}(\omega_k) & \cdots & \tilde{\gamma}_{1M}(\omega_k) \\ \tilde{\gamma}_{21}(\omega_k) & \tilde{\gamma}_{22}(\omega_k) & \cdots & \tilde{\gamma}_{2M}(\omega_k) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\gamma}_{M1}(\omega_k) & \tilde{\gamma}_{M2}(\omega_k) & \cdots & \tilde{\gamma}_{MM}(\omega_k) \end{bmatrix}. \tag{7}$$

Examples of $\tilde{\gamma}_{pq}(\omega k)$ are given by Eqs. (5) and (6).

The prerequisites for the sensor signals can be summarized as follows.

(1) The spatial coherence $\gamma_{pq}(\omega k)$ between the pth and the qth sensor should be equal to a predefined spatial coherence $\tilde{\gamma}_{pq}(\omega k)$.
(2) The PSDs of the sensor signals should be equal.

Let us define a vector that consists of the STFT coefficients of the sensor signals $\mathbf{X}(\ell, \omega_k) = [X_1(\ell, \omega_k), \ldots X_M(\ell, \omega_k)]^T$ and the noise signals $\mathbf{N}(\ell, \omega_k) = [N_1(\ell, \omega_k), \ldots, N_M(\ell, \omega_k)]^T$. The STFT coefficients of the sensor signals are obtained by instantaneous mixing of the STFT coefficients of the noise signals, i.e.,

$$\mathbf{X}(\ell, \omega_k) = \mathbf{C}^H(\omega_k)\mathbf{N}(\ell, \omega_k), \tag{8}$$

where $\mathbf{C}(\omega_k)$ denotes the mixing matrix and $(\cdot)^H$ is the Hermitian operation.

In Sec. III, we generate M mutually independent noise signals $N_p(\ell, \omega_k)$ $p \in \{1, 2, \ldots, M\}$. In Sec. IV, we show how to compute the mixing matrix $\mathbf{C}(\omega_k)$ such that the spatial coherence of the sensor signals equals Eq. (7).

III. GENERATE NOISE SIGNALS

In this section we provide two alternatives to generate the M mutually independent noise signals. In the following we assume that the column vectors of the mixing matrix have equal norm. If the short-term PSDs of the noise signals are equal then the short-term PSDs of the sensor signals are equal. Hence, the PSDs of the sensor signals are consistent with the PSDs observed in a homogeneous noise field.

A. Perfectly homogeneous

We generate M mutually independent noise signals with short-term PSD $\Phi(\ell,\omega_k)$. The short-term PSD of the noise signal can be time varying or time invariant, and spectrally white or colored.

We generate the STFT coefficients of the pth noise signal as

$$N_p(\ell,\omega_k) = \sqrt{\Phi(\ell,\omega_k)}\exp(i\pi D_p(\ell,\omega_k)), \tag{9}$$

where $D_p(\ell,\omega_k)$ for $\ell \in \{1,2,\ldots\}$ denotes a random signal with uniform distribution $(D_p(\ell,\omega_k) \sim U[-1,1])$. Now M mutually independent noise signals can be generated using mutually independent signals $D_p(\ell,\omega_k)$ for $p \in \{1,2,\ldots,M\}$, and the same $\Phi(\ell,\omega_k)$.

The short-term PSD $\Phi(\ell,\omega_k)$ can represent the long-term PSD of speech or factory noise, such that $\Phi(\ell,\omega_k) = \Phi(\omega_k)$. Alternatively, the short-term PSD $\Phi(\ell,\omega_k)$ can be equal to the short-term PSD of a given babble speech or factory noise signal. When babble speech or factory noise is used it should be noted that the original phase spectrum is destructed. Therefore, the resulting babble speech or factory noise signals do not sound like the original signals anymore. In the case when we use babble speech, the resulting noise signals are also known as *babble noise*.

B. Approximately homogeneous

We generate M mutually independent signals that consist of babble speech or factory noise. In the following we assume that these noise signals are continuous, i.e., there are no periods of silence. As mentioned before it is important that the noise signals have the same power. Therefore, we normalize the power of the noise signals $2 \leqslant p \leqslant M$ such that they have the same power as the first noise signal. The noise signals are first transformed into the STFT domain to construct the STFT coefficients $N_p(\ell,\omega_k)$ for $p \in \{1,2,\ldots,M\}$. When these signals are used to generate the sensor signals, the simulated noise field is not completely homogeneous since the short-term PSDs of $N_p(\ell,\omega_k)$ and $N_q(\ell,\omega_k)$ are not equal for $p \neq q$. However, in some cases, the long-term PSDs are approximately equal. For example, due to the nature of speech this is the case for babble speech signals. Therefore, only small fluctuations in the long-term PSDs are expected. Hence, the simulated noise field is approximately homogeneous.

It is important to note that when these noise signals are mixed to generate the desired spatial coherence the resulting sensor signals sound like a mixture of M babble speech or factory noise signals; this is not the case when we use the noise signals suggested in Sec. III A.

IV. GENERATING COHERENCE

In this section we determine the mixing matrix $\mathbf{C}(\omega_k)$ that generates the desired spatial coherence between the sensor signals.

In terms of the mixing matrix $\mathbf{C}(\omega_k)$ we can formulate the prerequisites as follows.

(1) The inner product between the pth and the qth column vector should be equal to $\tilde{\gamma}_{pq}(\omega_k)$.
(2) The norm of the column vectors should be equal to 1.

A. Two sensors

In this section we develop an efficient technique to generate two sensor signals with a predefined spatial coherence. Here we assume that the predefined spatial coherence between the first and the second sensor $\tilde{\gamma}_{12}(\omega_k)$ is real valued.

Let us define the following mixing matrix:

$$\mathbf{C}(\omega_k) = \begin{bmatrix} 1 & \sin(\alpha_{12}(\omega_k)) \\ 0 & \cos(\alpha_{12}(\omega_k)) \end{bmatrix}. \tag{10}$$

It can easily be verified that the norm of the column vectors equals to 1, and that the inner product equals to $\sin(\alpha_{12}(\omega_k))$. Hence, the spatial coherence $\tilde{\gamma}_{12}(\omega_k) = \sin(\alpha_{12}(\omega_k))$.

Although the mixing matrix (10) satisfies our prerequisites the mixed signals are not always adequate from a perceptual point of view. Specifically, they are inadequate in case noise signals exhibit different short-term PSDs (as discussed in Sec. III B). Let us assume that two mutually independent babble speech signals are mixed using Eq. (10). The sensor signal $X_1(\ell,\omega_k)$ only consists of $N_1(\ell,\omega_k)$, while the sensor signal $X_2(\ell,\omega_k)$ consists of a mixture of $N_1(\ell,\omega_k)$ and $N_2(\ell,\omega_k)$. Since $N_2(\ell,\omega_k)$ is not present in $X_1(\ell,\omega_k)$ the sensors signals sound unnatural. To solve this problem we require that the contribution of the noise signals in each of the sensor signals is equal. In terms of the mixing matrix we require that $(C_{11})^2 = (C_{12})^2$ and $(C_{21})^2 = (C_{22})^2$, where C_{pq} denotes the element in the pth row and qth column of the matrix $\mathbf{C}$. Since the norm of the column vectors and the inner product between the column vectors are not affected by the rotation operation, a valid solution can be obtained by properly rotating the mixing matrix.

Define the rotation matrix as

$$\mathbf{R}(\omega_k) = \begin{bmatrix} \cos(\beta(\omega_k)) & -\sin(\beta(\omega_k)) \\ \sin(\beta(\omega_k)) & \cos(\beta(\omega_k)) \end{bmatrix}. \tag{11}$$

The rotated mixing matrix is then given by

$$\begin{aligned} \mathbf{C}'(\omega_k) &= \mathbf{R}(\omega_k)\mathbf{C}(\omega_k) \\ &= \begin{bmatrix} \cos(\beta(\omega_k)) & \sin(\alpha_{12}(\omega_k) - \beta(\omega_k)) \\ \sin(\beta(\omega_k)) & \cos(\alpha_{12}(\omega_k) - \beta(\omega_k)) \end{bmatrix}. \end{aligned} \tag{12}$$

The additional requirement is fulfilled when

$$(\cos(\beta(\omega_k)))^2 = (\sin(\alpha_{12}(\omega_k) - \beta(\omega_k)))^2. \quad (13)$$

Given $\alpha_{12}(\omega_k)$ we can easily solve $\beta(\omega_k)$, viz., $\beta(\omega_k) = \alpha_{12}(\omega_k)/2 - \pi/4 + \pi n$, where $n \in \mathbb{Z}$.

B. Multiple sensors

In this section we develop an efficient technique to generate multiple sensor signals with a predefined spatial coherence.

The first technique that could be used is based on simulating the actual physical properties of the noise field, as shown in Ref. 10. However, a large number of mutually independent noise sources is required to approximate the desired spatial coherence.[10] Furthermore, the obtained spatial coherence depends on the positions of the noise sources. Therefore it is difficult, if not impossible, to find the appropriate positions to simulate an arbitrary spatial coherence.

The second technique, related to Eq. (10), is based on the Cholesky decomposition.[13] This technique is used in econometrics,[14] and in communications[15] to generate random variables with specific statistical properties. Accordingly, we obtain the mixing matrix by computing the Cholesky decomposition of the matrix $\tilde{\mathbf{\Gamma}}(\omega_k)$, i.e.,

$$\tilde{\mathbf{\Gamma}}(\omega_k) = \mathbf{C}^H(\omega_k)\mathbf{C}(\omega_k), \quad (14)$$

where $\mathbf{C}(\omega_k)$ is an upper triangle matrix. Since the mixing matrix is an upper triangle matrix we encounter a similar problem as in the two sensor case. Here, the Mth sensor signal consists of all noise signals while the first sensor signal only consists of the first noise signal. Furthermore, it should be noted that the Cholesky decomposition can only be calculated in case the matrix $\tilde{\mathbf{\Gamma}}(\omega_k)$ is positive definite. This requirement is fulfilled when Eq. (5) or Eq. (6) is used. In addition, the condition number of $\tilde{\mathbf{\Gamma}}(\omega_k)$ should not be too close to zero. The latter occurs when the spatial coherence corresponds to a directional sound source. In that case all values are nonzero and the rank of $\tilde{\mathbf{\Gamma}}(\omega_k)$ equals to 1. Hence, the response to a directional source cannot be simulated using this technique.

A more general solution is obtained by calculating the eigenvalue decomposition (EVD) of the matrix $\tilde{\mathbf{\Gamma}}(\omega_k)$ as follows:

$$\tilde{\mathbf{\Gamma}}(\omega_k) = \mathbf{V}(\omega_k)\mathbf{D}(\omega_k)\mathbf{V}^H(\omega_k). \quad (15)$$

Now we can split the diagonal matrix $\mathbf{D}(\omega_k)$ to obtain

$$\tilde{\mathbf{\Gamma}}(\omega_k) = \mathbf{V}(\omega_k)\sqrt{\mathbf{D}(\omega_k)}\sqrt{\mathbf{D}(\omega_k)}\mathbf{V}^H(\omega_k). \quad (16)$$

The mixing matrix is then given by

$$\mathbf{C}(\omega_k) = \sqrt{\mathbf{D}(\omega_k)}\mathbf{V}^H(\omega_k). \quad (17)$$

For $M > 2$ the mixing matrix does not provide equal contribution of each of the noise signals in each of the sensor signals. However, the results of informal listening tests confirmed that the sensor signals generated using the mixing matrix (17) are perceptually satisfactory. It should be noted that the EVD can also be used to compute the response to a single directional noise source. In case the spatial coherence is related to a coherent sound field only one eigenvalue is larger than zero. Therefore, only the first row of the mixing matrix will contain elements larger than zero. Consequently, the sensor signals are all related to the first noise signal.

TABLE I. Summary of the developed algorithm that generates multisensor signals under a predefined spatial coherence constraint.

(1) Define a matrix $\tilde{\mathbf{\Gamma}}(\omega_k)$ for each ω_k that consists of the predefined spatial coherence values.
(2) Calculate the eigenvalue decomposition of the matrix $\tilde{\mathbf{\Gamma}}(\omega_k) = \mathbf{V}(\omega_k)\mathbf{D}(\omega_k)\mathbf{V}^H(\omega_k)$. The mixing matrix is then obtained by $\mathbf{C}(\omega_k) = \sqrt{\mathbf{D}(\omega_k)}\mathbf{V}^H(\omega_k)$.
(3) Generate M mutually independent complex random signals $N_p(\ell, \omega_k)$ (see Sec. III).
(4) For all ℓ and ω_k. Calculated $\mathbf{X}(\ell, \omega_k) = \mathbf{C}^H(\omega_k)\mathbf{N}(\ell, \omega_k)$, where $\mathbf{X}(\ell, \omega_k) = [X_1(\ell, \omega_k), \ldots, X_M(\ell, \omega_k)]^T$ and $\mathbf{N}(\ell, \omega_k) = [N_1(\ell, \omega_k), \ldots, N_M(\ell, \omega_k)]^T$.
(5) Finally, the sensor signals can be obtained by calculating the inverse STFT of $X_p(\ell, \omega_k)$ for $p \in \{1, 2, \ldots, M\}$.

V. ALGORITHM SUMMARY AND COMPUTATIONAL COMPLEXITY

A summary of the developed algorithm for generating M stationary or nonstationary sensor signals is provided in Table I. The first three steps are part of the initialization. The fourth step generates the STFT coefficients of the sensor signals. Finally, in the fifth step, the inverse STFT is used to obtain M discrete time signals that exhibit the predefined spatial constraint.

We now determine the computational complexity of the algorithm. Let us assume that the frame length of the STFT equals K. Firstly, we determine the computational complexity required to obtain the mixing matrix. Since the mixing matrix is time invariant we only need to compute it once. The complexity of constructing a mixing matrix for a single frequency is $O(M^3)$ for the EVD and $O(M^2)$ to construct the mixing matrix. For K frequencies the complexity is $O(K(M^2 + M^3))$. Hence, the computational complexity of the initialization grows rapidly for increasing K and M. Secondly, we determine the computational complexity required to compute Eq. (8). Specifically, we need to generate M STFT coefficients, one for each sensor signal, which yields $O(M^2)$ per time frame and frequency index. For K frequencies we then obtain $O(KM^2)$.

Let us assume that we need to generate M sensor signals of length L. In case $R = K/4$ denotes the number of samples between two successive STFT frames, we need to compute $L' = \lceil L/R \rceil$ time frames. The computational complexity of computing the STFT coefficients for all time frames, frequencies, and sensor signals yields $O(L'KM^2)$. Finally, we compute the inverse STFT, which can be efficiently implemented using a weighted overlap-add technique. The discrete Fourier transform of length K can be computed using the fast Fourier transform and has a complexity of order $O(K \log_2(K))$. The total computational complexity to compute $L \times M$ samples is given by $O(L'K^2M^2 \log_2(K)) \approx O(LKM^2 \log_2(K))$. We can see that the computational complexity per sample grows for increasing K and M. For a

given M the overall computational complexity can be reduced by using a small value of K. The minimum value of the frame length K is determined by the minimum filter order. The influence of K on the accuracy of the spatial coherence is examined in Sec. VI.

The computational complexity can be reduced by using more efficient techniques to compute the EVD, by performing more efficient matrix multiplications, and by exploiting the fact that the spectrum is conjugate symmetric.

VI. PERFORMANCE EVALUATION

In this section we analyze the generated sensor signals. The spatial coherences between two sensors in spherical and cylindrical noise fields are calculated and depicted. Firstly, the sensor signals received in a homogenous noise field are generated as described in Sec. III A and analyzed. Secondly, babble speech is generated as described in Sec. III B and subsequently analyzed. For the following analysis we have generated two sensor signals since the spatial coherence can only be measured between two sensors. The authors conducted the analysis with more than two sensors and confirmed that the results are equivalent.

In this work we used a recorded single babble speech signal available in Ref. 16. Alternatively, one can generate a babble speech signal by mixing N mutually independent speech signals. In order to obtain a realistic babble speech signal one can convolve each speech signal with a different room impulse response (the source positions are uniformly spread in a room). The N room impulse responses (RIRs) can be generated using the image method proposed by Allen and Berkley[17] with a reverberation time between 0.3 and 0.6 s. This method can also be used to generate M babble speech signals that exhibit the characteristics of a 2D or a 3D diffuse noise field. In the latter case we require MN RIRs. In case multiple sensor signals are generated, it is recommended to use only the reverberant parts of the RIRs since the direct paths are coherent. When multiple babble speech signals are generated the spatial coherence function is determined by the geometry of the enclosure and acoustic properties of the walls. Specifically, a spherical (3D diffuse) noise field is obtained when the reflection coefficients of the six walls are larger than zero. A cylindrical (2D) noise field is obtained when the reflection coefficients of the floor and ceiling are zero, and the other reflection coefficients are larger than zero.

Here we have used the later method for comparison, i.e., we generated multiple sensor signals by mixing 50 reverberant noise signals; this method is indicated as the "image method". The reverberant noise signals are generated by convolving each noise source signal with the reverberant part of a RIR. The noise sources are positioned such that (i) the distance between each source and all walls is larger than 1 m, and (ii) the distance between each source and the center of the sensor array is larger than 1 m. In order to simulate a spherical noise field the reflection coefficients of the six walls are set to 0.8. To simulate a cylindrical noise field the reflection coefficients of the sidewalls are set to 0.8 while those related to the floor and ceiling are set to 0. It should be noted that the computational complexity of the developed algorithm is much smaller than that of the image method. The developed algorithm only requires the generation of M noise sources, while the image method requires the generation of N noise sources, where $N \gg M$ to approximate the desired noise field. For N noise sources and M sensors we need to compute MN RIRs. Subsequently, each noise source needs to be convolved with M filters. Hence, the image method requires MN filter operations while the developed algorithm requires only M^2 filter operations.

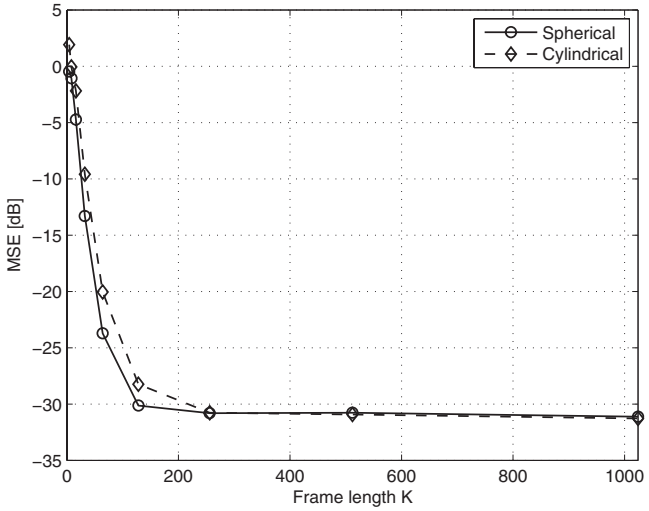

FIG. 1. Minimum square error between the theoretical spatial coherence and the spatial coherence of the generated signals versus the STFT frame length K.

Let us define the error between the spatial coherence of two generated signals and the theoretical spatial coherence by the normalized mean square error (MSE), i.e.,

$$\mathrm{MSE} = \frac{\sum_{k=0}^{K/2-1} (\hat{\gamma}_{pq}(\omega_k) - \tilde{\gamma}_{pq}(\omega_k))^2}{\sum_{k=0}^{K/2-1} (\tilde{\gamma}_{pq}(\omega_k))^2}, \tag{18}$$

where k denotes the discrete frequency index and $\hat{\gamma}_{\mathrm{pq}}(\omega_k)$ denotes the estimated spatial coherence of the desired spatial coherence $\tilde{\gamma}_{\mathrm{pq}}(\omega_k)$. The MSE is used to evaluate the image method and the developed algorithm. The coherence between the pth and the qth sensor was estimated using Welch's averaged periodogram method.[18] We used the fast Fourier transform of length 2048, a Hann window, and 75% overlap.

A. Minimum frame length

As discussed in Sec. V we prefer to use short STFT frames to achieve low computational complexity. In this section we determine the minimum length of the STFT frames, or in other words the minimum filter length. For the following experiment the mutually independent noise signals were generated as described in Sec. III A. Subsequently, we generated two sensor signals obtained in spherically and cylindrically isotropic noise fields as described in Sec. IV. The obtained MSE values for different values $K \in \{4, 8, 16, \ldots, 1024\}$ are shown in Fig. 1. For both noise fields we can see that the MSE is not significantly improved when $K \geq 256$. Hence, we chose $K = 256$, which provides a low MSE and results in a computationally efficient algo-

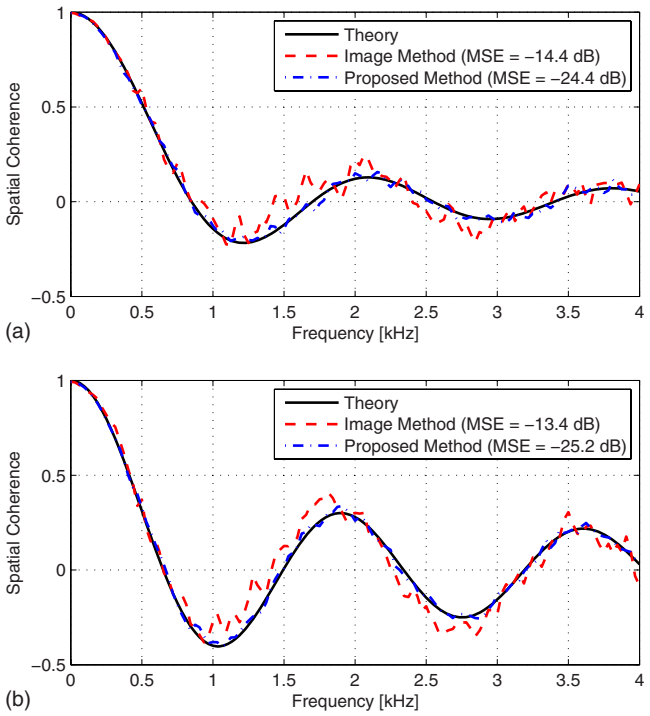

FIG. 2. (Color online) Spatial coherence between two nonstationary speechlike noise signals (d=20 cm): (a) spherically isotropic noise field and (b) cylindrically isotropic noise field.

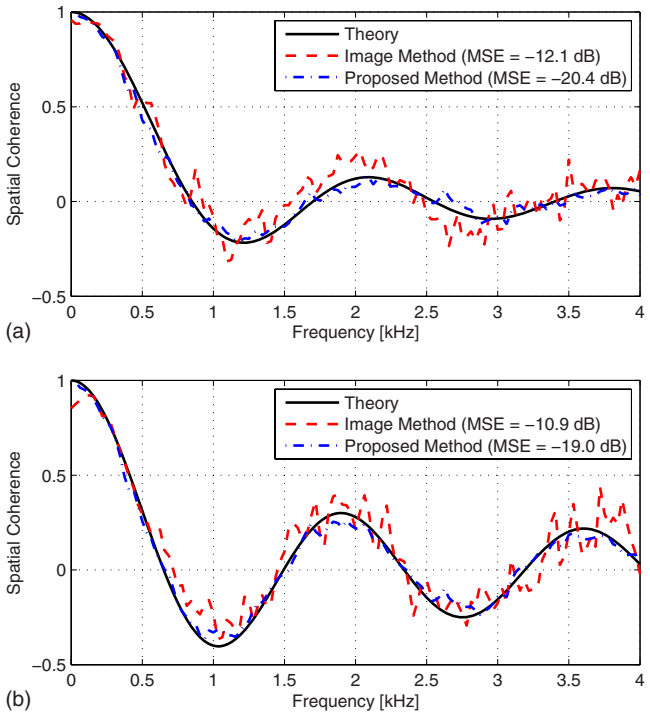

FIG. 3. (Color online) Spatial coherence between two babble speech sensor signals (d=20 cm): (a) spherical noise field and (b) cylindrical noise field.

rithm. Similar results were obtained when the noise signals are generated using the method described in Sec. III B.

B. Isotopic noise

The developed algorithm, as summarized in Table I, was used to generate the sensor signals obtained in spherically and cylindrically isotropic noise fields. Here we used nonstationary speechlike noise signals that were generated using the method described in Sec. III A. For comparison we also generated a set of sensor signals using the image method. We generated two sensor signal of 20 s and intersensor distance d=20 cm.

The simulation and theoretical results for the spherical and cylindrical noise fields are shown in Fig. 2. From the results shown in this figure we can see that the spatial coherence of the generated sensor signals obtained using the proposed method closely matches the theoretical spatial coherence for all frequencies. The spatial coherence of the generated sensor signals obtained using the image method only gives a good match at low frequencies. The MSE obtained by the proposed method is significantly lower than the MSE obtained by the image method for both noise fields.

C. Babble speech

Now the developed algorithm was used to generate the nonstationary sensor signals in spherically and cylindrically isotropic noise fields. The noise signals consist of babble speech signals that were generated using the method described in Sec. III B. We generated two sensor signals of 20 s and intersensor distance d=20 cm.

The simulation and theoretical results for the spherical and cylindrical noise fields are shown in Fig. 3. From these results we can see that the spatial coherence of the generated signals closely matches the theoretical spatial coherence. Again we see that the MSE of the proposed method is significantly smaller than the MSE of the image method. In Fig. 4 the long-term PSDs of sensor signals 1 and 2 are shown. We can see that the long-term PSDs of the multisensor babble speech signals are approximately equal although the short-term PSD of the babble speech signal at each time frame is known to be different. This demonstrates that we can accurately generate diffuse babble speech.

VII. CONCLUSION

We have developed a computationally efficient algorithm to generate sensor signals that exhibit a predefined spatial coherence. Different noise types such as babble

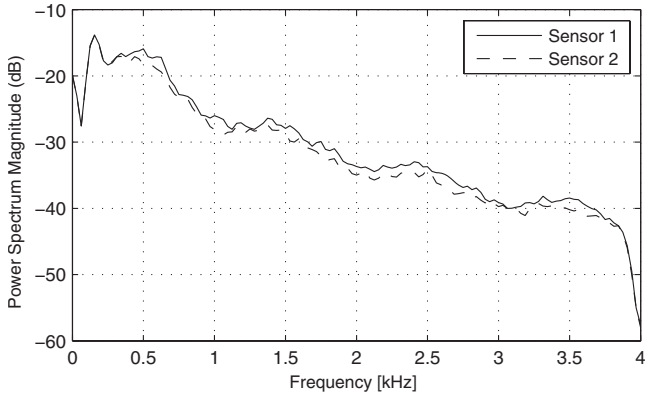

FIG. 4. Estimated PSD of sensor signals 1 and 2.

speech and factory noise can easily be generated. The algorithm consists of two steps: Firstly, we generate M mutually independent noise signals in the STFT domain, where M is equal to the number of sensors. Secondly, we calculate a frequency dependent mixing matrix, which will induce the predefined spatial coherence. We generate the spectral coefficients of the sensor signals for each time frame by multiplying the mixing matrix with the spectral coefficients of the noise signals. The sensor signals are then obtained by computing the inverse STFT. The major benefits of the developed algorithm are that (i) we can induce any spatial coherence and (ii) we only require M mutually independent noise signals. The performance evaluation showed that the spatial coherence of the generated sensor signals closely resembles the desired spatial coherence. Therefore the generated signals are useful for the evaluation and analysis of various signal processing algorithms. In addition, we showed that the MSE between the desired spatial coherence and the spatial coherence of generated signals is smaller than the MSE between the desired spatial coherence and the spatial coherence obtained using the image method that is frequently used in the acoustic signal processing community.

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

[2]S. Gannot and I. Cohen, "Speech enhancement based on the general transfer function GSC and postfiltering," IEEE Trans. Speech Audio Process. **12**, 561–571 (2004).

[3]J. Bitzer, K. Simmer, and K. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'99), 1999, Vol. **5**, pp. 2965–2968.

[4]J. Bitzer, K. Simmer, and K. Kammeyer, "Multimicrophone noise reduction by postfilter and superdirective beamformer," in Proceedings of the International Workshop on Acoustic Echo and Noise Control (IWAENC'99), 1999, pp. 100–103.

[5]S. Nordholm, I. Claesson, and B. Bengtsson, "Adaptive array noise suppression of hands-free speaker input in cars," IEEE Trans. Veh. Technol. **42**, 514–518 (1993).

[6]N. Dal-Degan and C. Prati, "Acoustic noise analysis and speech enhancement techniques for mobile radio applications," Signal Process. **18**, 43–56 (1988).

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

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

[9]M. Goulding and J. Bird, "Speech enhancement for mobile telephony," IEEE Trans. Veh. Technol. **39**, 316–326 (1990).

[10]E. Habets and S. Gannot, "Generating sensor signals in isotropic noise fields," J. Acoust. Soc. Am. **122**, 3464–3470 (2007).

[11]H. Cox, "Spatial correlation in arbitrary noise fields with application to ambient sea noise," J. Acoust. Soc. Am. **54**, 1289–1301 (1973).

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

[13]E. Kreyszig, *Advanced Engineering Mathematics*, 8th ed. (Wiley, New York, 1998).

[14]R. Tsay, *Analysis of Financial Time Series*, Probability and Statistics (Wiley, New York, 2002).

[15]B. Natarajan, C. Nassar, and V. Chandrasekhar, "Generation of correlated Rayleigh fading envelopes for spread spectrum applications," IEEE Commun. Lett. **4**, 9–11 (2000).

[16]TNO-Perception, "Babble speech recording," http://spib.rice.edu/spib/data/signals/noise/babble.html (last viewed June, 2008).

[17]J. Allen and D. Berkley, "Image method for efficiently simulating small room acoustics," J. Acoust. Soc. Am. **65**, 943–950 (1979).

[18]P. Stoica and R. Moses, *Spectral Analysis of Signals* (Prentice-Hall, Englewood Cliffs, NJ, 2005).