On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation

Oliver Thiergart,[a] Giovanni Del Galdo, and Emanuël A. P. Habets
International Audio Laboratories Erlangen, Am Wolfsmantel 33, 91058 Erlangen, Germany (A joint institution of the University Erlangen-Nuremberg and Fraunhofer IIS, Germany)

(Received 9 February 2012; revised 28 July 2012; accepted 20 August 2012)

Many applications in spatial sound recording and processing model the sound scene as a sum of directional and diffuse sound components. The power ratio between both components, i.e., the signal-to-diffuse ratio (SDR), represents an important measure for algorithms which aim at performing robustly in reverberant environments. This contribution discusses the SDR estimation from the spatial coherence between two arbitrary first-order directional microphones. First, the spatial coherence is expressed as function of the SDR. For most microphone setups, the spatial coherence is a complex function where both the absolute value and phase contain relevant information on the SDR. Secondly, the SDR estimator is derived from the spatial coherence function. The estimator is discussed for different practical microphone setups including coincident setups of arbitrary first-order directional microphones and spaced setups of identical first-order directional microphones. An unbiased SDR estimation requires noiseless coherence estimates as well as information on the direction-of-arrival of the directional sound, which usually has to be estimated. Nevertheless, measurement results verify that the proposed estimator is applicable in practice and provides accurate results. © *2012 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4750493]

PACS number(s): 43.55.Br, 43.55.Cs, 43.55.Nd [NX] Pages: 2337–2346

I. INTRODUCTION

Sound fields in reverberant environments are often modeled as a superposition of directional sound components (e.g., a plane wave) and reverberant sound components. The power ratio between both components, often referred to as the direct-to-reverberant ratio (DRR), represents an important parameter in signal processing and psychoacoustics. Assuming that the reverberant sound forms a diffuse field, it is more reasonable to refer to this parameter as the signal-to-diffuse ratio (SDR).[1] Typical applications of the SDR include dereverberation,[2,3] parametric spatial audio coding,[4] and evaluation of beamforming performance.[5] It is also widely assumed that the parameter influences the perceived distance of a listener from the source.[6,7] Many experiments on the sensitivity of human hearing to changes in the DRR have been carried out.[8,9]

Due to the importance of the SDR, various SDR estimators have been developed in the last four decades. Nowadays, many estimators are based on the spatial coherence between two microphones. Compared to more traditional estimators,[10] considering the spatial coherence is especially practical since it does not require one to determine room impulse responses. In order to derive an SDR estimator based on the spatial coherence, the theoretical spatial coherence function is required for the observed sound field. The theoretical spatial coherence function for purely diffuse sound fields was derived for various microphone setups[11–15] and was also verified via measurements in reverberant rooms.[16] A derivation of the power ratio between a single distant noise source and a collection of independent surrounding noise sources based on the spatial coherence between two omnidirectional microphones was presented by Piersol.[17] Jeub *et al.*[18] considered the real part of the complex spatial coherence between two omnidirectional microphones for estimating the SDR and assumed that directional sound arrives from the broadside of the array. Recently, an SDR estimator that is based on the complex spatial coherence was proposed for omnidirectional microphones.[19] The approach proposed by Hioka *et al.*[20] is closely related to the coherence based estimation techniques. The authors expressed the spectral densities between several omnidirectional microphones as a function of the directional sound power and diffuse sound power and proposed to estimate both quantities using a least-squares approach. Recently, Kuster[21] has derived an SDR estimator based on the spatial coherence between the coincident pressure and particle velocity, which represents a special case of the following contribution. The spatial coherence of directional microphones, and hence spatial coherence based SDR estimators for directional microphones, have received little attention.

In this contribution, we extend the aforementioned derivations to microphones with arbitrary first-order directivity where the two microphones can be placed and oriented nearly arbitrarily. First, the spatial coherence is derived and expressed as a function of the SDR. Secondly, a general solution to the SDR estimation problem is presented and discussed for various practical first-order microphone setups. The derived SDR estimators consider the spatial coherence as a complex function to exploit all relevant information. It follows from the derivation that the direction-of-arrival (DOA) of the directional sound needs to be considered to obtain an unbiased estimator. We study the performance of

a)Author to whom correspondence should be addressed. Electronic mail: oliver.thiergart@audiolabs-erlangen.de

the SDR estimator when microphone self-noise is present and when the DOA of the directional sound is estimated.

The paper is structured as follows: In Sec. II, we introduce the sound field model. Section III reviews the definition of the spatial coherence and presents its derivation for mixed sound fields. Solutions to the SDR estimation problem for practical microphone setups are presented in Sec. IV. The provided derivations and SDR estimators are discussed in Sec. V based on simulations and measurements. The conclusions are drawn in Sec. VI.

II. SOUND FIELD MODEL

We model the complex sound pressure $S(k, t, \mathbf{d})$ at an arbitrary point $\mathbf{d}$ in a Cartesian coordinate system at a time instant t and wavenumber $k = 2\pi f/c$ (frequency f, speed of sound c) as a superposition of *directional sound* and *diffuse sound*, i.e,

$$S(k, t, \mathbf{d}) = S_{\mathrm{dir}}(k, t, \mathbf{d}) + S_{\mathrm{diff}}(k, t, \mathbf{d}). \tag{1}$$

Note that lower-case boldface letters represent column vectors in the following. The directional sound $S_{\mathrm{dir}}(k, t, \mathbf{d})$ equals a single plane wave (far-field assumption) with propagation direction $\mathbf{n}_{\mathrm{dir}}(k)$, i.e.,

$$S_{\mathrm{dir}}(k, t, \mathbf{d}) = \sqrt{P_{\mathrm{dir}}(k, t, \mathbf{d})}\, e^{j\mu(k)+j\phi_0(k,t)}, \tag{2}$$

where $P_{\mathrm{dir}}(k, t, \mathbf{d})$ is the power of the wave and $\phi_0(k, t)$ is the phase of the wave at the origin of the coordinate system. Moreover, $\mu(k) = k\mathbf{n}_{\mathrm{dir}}^T(k)\mathbf{d}$ with $||\mathbf{n}_{\mathrm{dir}}(k)|| = 1$ is the phase shift from the origin to $\mathbf{d}$. The dependency of $\mathbf{n}_{\mathrm{dir}}(k)$ and $\mu(k)$ on k is omitted in the following. When considering a three-dimensional sound field, $\mathbf{n}_{\mathrm{dir}}$ can be expressed as

$$\mathbf{n}_{\mathrm{dir}} = \begin{bmatrix} \cos(\varphi_{\mathrm{dir}})\cos(\vartheta_{\mathrm{dir}}) \\ \sin(\varphi_{\mathrm{dir}})\cos(\vartheta_{\mathrm{dir}}) \\ \sin(\vartheta_{\mathrm{dir}}) \end{bmatrix}, \tag{3}$$

where φ_{dir} and ϑ_{dir} are the azimuth and elevation, respectively.

The diffuse sound field $S_{\mathrm{diff}}(k, t, \mathbf{d})$ in Eq. (1) is assumed to be spatially isotropic, meaning that the sound arrives with equal strength from all directions, and spatially homogeneous, meaning that its mean power,

$$P_{\mathrm{diff}}(k, t) = E\{|S_{\mathrm{diff}}(k, t, \mathbf{d})|^2\}, \tag{4}$$

does not vary with $\mathbf{d}$. In Eq. (4), $\mathrm{E}\{\cdot\}$ denotes the mathematical expectation. In the following, $S_{\mathrm{diff}}(k, t, \mathbf{d})$ and $S_{\mathrm{dir}}(k, t, \mathbf{d})$ are assumed to be uncorrelated.

The ratio between the power of the directional sound $P_{\mathrm{dir}}(k, t)$ and the power of the diffuse sound, Eq. (4), represents the SDR $\Gamma(k, t, \mathbf{d})$ that we want to estimate, i.e.,

$$\Gamma(k, t, \mathbf{d}) = \frac{P_{\mathrm{dir}}(k, t, \mathbf{d})}{P_{\mathrm{diff}}(k, t)}. \tag{5}$$

Note that $\Gamma(k, t, \mathbf{d})$ is by definition real-positive.

To estimate $\Gamma(k, t, \mathbf{d})$, we measure the sound pressure at two positions $\mathbf{d}_1$ and $\mathbf{d}_2$ using two microphones with arbitrary first-order directivity and orientation. The setup is depicted in Fig. 1. According to the sound field model in Eq. (1), the ith microphone signal with $i \in \{1, 2\}$ can be written as

$$X_i(k, t) = X_{\mathrm{dir},i}(k, t) + X_{\mathrm{diff},i}(k, t) + N_i(k, t), \tag{6}$$

where $X_{\mathrm{dir},i}(k, t)$ is the noiseless sound pressure resulting from the single plane wave and $X_{\mathrm{diff},i}(k, t)$ is the corresponding sound pressure resulting from the diffuse sound field. The microphone self-noise $N_i(k, t)$ is modeled as uncorrelated zero-mean complex Gaussian noise with mean power

$$P_{N,i}(k, t) = \mathrm{E}\{|N_i(k, t)|^2\}. \tag{7}$$

The pressure signal $X_{\mathrm{dir},i}(k, t)$ resulting from the directional sound can be expressed as

$$X_{\mathrm{dir},i}(k, t) = S_{\mathrm{dir}}(k, t, \mathbf{d}_i)g_i(\mathbf{n}_{\mathrm{dir}}), \tag{8}$$

where

$$g_i(\mathbf{n}_{\mathrm{dir}}) = \alpha_i - (1-\alpha_i)\mathbf{n}_{\mathrm{dir}}^T\mathbf{l}_i \tag{9}$$

is the first-order directional response of the microphone, $\mathbf{l}_i$ is a unit vector describing the orientation of the microphone, and $\alpha_i \in [0, 1]$ is the shape parameter. For instance, for $\alpha_i = 0.5$, the microphone possesses a cardioid directivity while it is omnidirectional for $\alpha_i = 1$. When measuring a diffuse field with directional microphones ($\alpha_i < 1$), we capture the sound field power only partially, i.e.,

$$\mathrm{E}\{|X_{\mathrm{diff},i}(k, t)|^2\} = Q_i^{-1}P_{\mathrm{diff}}(k, t), \tag{10}$$

where $Q_i \geq 1$ is the microphone directivity factor.[22] For first-order directional microphones in a spherically isotropic diffuse sound field ($\mathbf{n}_{\mathrm{diff}} \in \mathbb{R}^{3\times1}$), we have[13,22]

$$Q_i^{-1} = \alpha_i^2 + \frac{1}{3}(1-\alpha_i)^2. \tag{11}$$

Note that α_i, Q_i, and $g_i(\mathbf{n}_{\mathrm{dir}})$ can be frequency dependent. In the following sections, the dependency of all quantities on t is omitted for brevity.

III. COMPLEX SPATIAL COHERENCE

A. Definitions

The coherence $\gamma_{12}(kr)$ between two microphone signals $X_1(k)$ and $X_2(k)$ is defined as[23]

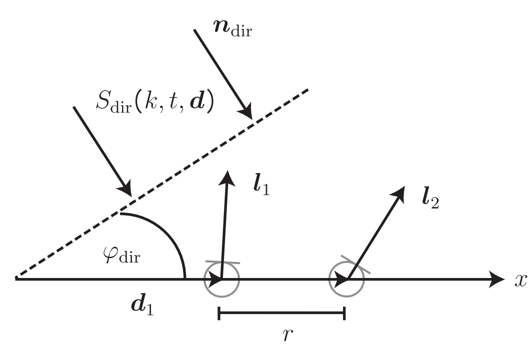

FIG. 1. Illustration of the setup used throughout this paper. Without loss of generality, both microphones are located on the x axis. Lower-case boldface letters represent column vectors.

$$\gamma_{12}(kr) = \frac{\Phi_{12}(kr)}{\sqrt{\Phi_{11}(k)}\sqrt{\Phi_{22}(k)}}, \tag{12}$$

where $r = ||\mathbf{d}_2 - \mathbf{d}_1||$ is the microphone spacing, $\Phi_{12}(kr)$ is the cross power spectral density (PSD) between the two microphone signals, and $\Phi_{11}(k)$ and $\Phi_{22}(k)$ are the corresponding auto PSDs. The PSDs can be determined with

$$\Phi_{ij}(\cdot) = \mathrm{E}\{X_i(k)X_j^*(k)\}, \tag{13}$$

where $(\cdot)^*$ denotes complex conjugate. In practice, the computation of $\mathrm{E}\{\cdot\}$ is usually replaced by time averaging. For the directional sound, Eq. (13) becomes with Eqs. (2) and (8)

$$\Phi_{\mathrm{dir},\,ii}(k) = g_i^2(\mathbf{n}_{\mathrm{dir}})P_{\mathrm{dir}}(k)\,, \tag{14a}$$

$$\Phi_{\mathrm{dir},\,12}(kr) = g_1(\mathbf{n}_{\mathrm{dir}})g_2(\mathbf{n}_{\mathrm{dir}})P_{\mathrm{dir}}(k)e^{-j\mu_{\mathrm{dir}}}\,, \tag{14b}$$

where $\mu_{\mathrm{dir}} = k\mathbf{n}_{\mathrm{dir}}^T\mathbf{r}$ with $\mathbf{r} = \mathbf{d}_2 - \mathbf{d}_1$ describes the phase shift of the plane wave from the first microphone to the second. Equation (14a) and (14b) assumes that the directional sound arrives with equal power at both microphone positions, i.e., $P_{\mathrm{dir}}(k, \mathbf{d}_1) = P_{\mathrm{dir}}(k, \mathbf{d}_2) = P_{\mathrm{dir}}(k)$. This assumption holds reasonably well in practice when the microphone spacing $r = ||\mathbf{r}||$ is sufficiently small compared to the distance to the sound source. The coherence $\gamma_{\mathrm{dir}}(kr)$ for purely directional sound follows by inserting Eq. (14) into Eq. (12), i.e.,

$$\gamma_{\mathrm{dir}}(kr) = \mathrm{sgn}\Big(g_1(\mathbf{n}_{\mathrm{dir}})g_2(\mathbf{n}_{\mathrm{dir}})\Big)\,e^{-j\mu_{\mathrm{dir}}}, \tag{15}$$

which depends on the DOA of the directional sound.

Similarly, the coherence $\gamma_{\mathrm{diff}}(kr)$ for a purely diffuse field can be expressed by substituting Eqs. (10) and (13) into Eq. (12), i.e.,

$$\gamma_{\mathrm{diff}}(kr) = \frac{\Phi_{\mathrm{diff},\,12}(kr)}{P_{\mathrm{diff}}(k)\sqrt{Q_1^{-1}Q_2^{-1}}}. \tag{16}$$

The coherence $\gamma_{\mathrm{diff}}(kr)$ was, for example, derived for omnidirectional microphones,[11] first-order directional microphones,[13] and for first-order differential microphone arrays.[15] A few theoretical examples for a spherically isotropic diffuse field are depicted in Fig. 2. For the two omnidirectional microphones (○–○), we obtain the well-known result $\gamma_{\mathrm{diff}}(kr) = \sin(kr)/kr$. For large kr (high frequencies or large microphone spacings), all depicted coherence functions approach zero, meaning that the microphone signals become uncorrelated. At low kr, the omnidirectional setup (○–○) and the two cardioid setups (Δ–Δ, ◁–▷) yield a relatively high coherence, i.e., the microphone signals become correlated although the sound field is diffuse. In contrast, the two dipole setups lead to zero coherence, i.e., uncorrelated signals, for all kr. Note that $\gamma_{\mathrm{diff}}(kr)$ can be complex.

With Eqs. (16), (10), and (13), the diffuse field PSDs $\Phi_{\mathrm{diff},\,ij}(k)$ can be expressed as

$$\Phi_{\mathrm{diff},\,ii}(k) = \frac{P_{\mathrm{diff}}(k)}{Q_i}\,, \tag{17a}$$

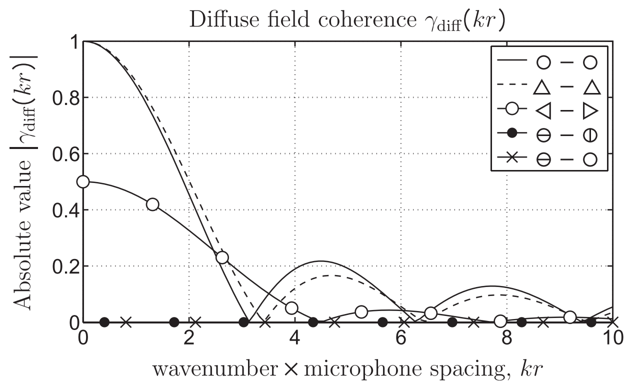

FIG. 2. Diffuse field coherence (absolute value) for a spherically isotropic field for different microphone setups. ○–○: Two displaced omnidirectional microphones. ▷: Cardioid microphone with orientation $(\varphi, \vartheta) = (0, 0)$. Δ: Cardioid $(90^\circ, 0)$. ⦶: Dipole $(0, 0)$. ⊖: Dipole $(90^\circ, 0)$. All arrays are aligned with the x axis. The coherence for the two latter setups (⊖–⦶, ⊖–○) is zero for all kr.

$$\Phi_{\mathrm{diff},\,12}(kr) = \gamma_{\mathrm{diff}}(kr)\frac{P_{\mathrm{diff}}(k)}{\sqrt{Q_1Q_2}}. \tag{17b}$$

Similarly, with Eqs. (7) and (13) the PSDs of the microphone self-noise can be expressed as

$$\Phi_{N,\,ii}(k) = P_{N,\,i}(k), \tag{18a}$$

$$\Phi_{N,\,12}(kr) = 0, \tag{18b}$$

assuming that the two noise signals are uncorrelated.

B. Solutions for mixed sound fields

For the signal model in Sec. II, where a mixture of the directional sound, diffuse sound, and microphone noise is measured, the PSDs $\Phi_{ij}(\cdot)$ in Eq. (13) equal the sum of the individual PSDs, i.e.,

$$\Phi_{ij}(\cdot) = \Phi_{\mathrm{dir},\,ij}(\cdot) + \Phi_{\mathrm{diff},\,ij}(\cdot) + \Phi_{N,\,ij}(\cdot), \tag{19}$$

which assumes that the three components are mutually uncorrelated. Let us define the *noiseless* spatial coherence as

$$\gamma_s(kr) = \frac{\Phi_{s,\,12}(kr)}{\sqrt{\Phi_{s,\,11}(k)}\sqrt{\Phi_{s,\,22}(k)}}, \tag{20}$$

where $\Phi_{s,\,ij}(\cdot) = \Phi_{\mathrm{dir},\,ij}(\cdot) + \Phi_{\mathrm{diff},\,ij}(\cdot)$ are the (noiseless) signal PSDs. Depending on the statistical properties of $X_{\mathrm{dir},\,i}(k)$, $X_{\mathrm{diff},\,i}(k)$, and $N_i(k)$, there exist different optimal estimators for $\Phi_{s,\,ij}(\cdot)$. When an estimate of the noise PSDs $\Phi_{N,\,ii}(k)$ is available, we can determine $\gamma_s(kr)$ by considering Eqs. (19) and (18), i.e.,

$$\gamma_s(kr) = \frac{\Phi_{12}(kr)}{\sqrt{\Phi_{11}(k) - \Phi_{N,11}(k)}\sqrt{\Phi_{22}(k) - \Phi_{N,22}(k)}}. \tag{21}$$

Clearly, disregarding the microphone noise, i.e., assuming $\Phi_{N,\,ii}(k) = 0$, will introduce a bias in the estimate of $\gamma_s(kr)$. This leads to biased SDR estimates as shown in Sec. V. Substituting Eqs. (5) and (14)–(17) into Eq. (20) yields

$$\gamma_s(kr) = \frac{\Gamma(k)\gamma_{\text{dir}}(kr)q_1q_2 + \gamma_{\text{diff}}(kr)}{\sqrt{\Gamma^2(k)q_1^2q_2^2 + \Gamma(k)(q_1^2+q_2^2)+1}}, \tag{22}$$

where the dependency of $\Gamma(k)$ on $\mathbf{d}$ was omitted and

$$q_i = |g_i(\mathbf{n}_{\text{dir}})|\sqrt{Q_i} \in [0,\, 1]. \tag{23}$$

A detailed derivation of Eq. (22) is provided in the first part of the Appendix. When both microphones possess an equal directivity and orientation, i.e., when $q_1 = q_2 = q$, Eq. (22) can be simplified to

$$\gamma_s(kr)|_{q_1=q_2} = \frac{\Gamma(k)\gamma_{\text{dir}}(kr)q^2 + \gamma_{\text{diff}}(kr)}{\Gamma(k)q^2+1}. \tag{24}$$

Equation (22) expresses the spatial coherence $\gamma_s(kr)$ between two arbitrary first-order directional microphones as a function of the SDR $\Gamma(k)$ for the sound field model introduced in Sec. II. As expected, we have $\gamma_s(kr) = \gamma_{\text{diff}}(kr)$ for purely diffuse fields $[\Gamma(k) = 0]$ and $\gamma_s(kr) = \gamma_{\text{dir}}(kr)$ when only the directional sound is present $[\Gamma(k) \to \infty]$. Clearly, $\gamma_s(kr)$ depends on the DOA of the directional sound for any microphone setup (disregarding the meaningless setup of two coincident omnidirectional microphones). Moreover, for specific first-order directional microphone setups, $\gamma_s(kr)$ is not a bijective function, i.e., different $\Gamma(k)$ can lead to the same $\gamma_s(kr)$. As is discussed in Sec. IV, this characteristic together with the DOA dependency of $\gamma_s(kr)$ can make it difficult to estimate the SDR.

IV. SDR ESTIMATION

A. General solution

In order to estimate the SDR from the spatial coherence, let us first introduce two complex variables $z_s(kr)$ and $z_{\text{diff}}(kr)$ as

$$z_s(kr) = \gamma^*_{\text{dir}}(kr)\gamma_s(kr), \tag{25a}$$

$$z_{\text{diff}}(kr) = \gamma^*_{\text{dir}}(kr)\gamma_{\text{diff}}(kr). \tag{25b}$$

The SDR is determined by solving Eq. (22) for $\Gamma(k)$. This leads to two results (of which only one is correct), namely

$$\Gamma(k) = \frac{A \pm \sqrt{A^2 + 4\left(1 - z_s^2(kr)\right)\left(z_s^2(kr) - z_{\text{diff}}^2(kr)\right)}}{2q_1q_2\left(1 - z_s^2(kr)\right)}, \tag{26}$$

where

$$A = z_s^2(kr)\left(\frac{q_1}{q_2} + \frac{q_2}{q_1}\right) - 2z_{\text{diff}}(kr). \tag{27}$$

The derivation of Eq. (26) is presented in the second part of the Appendix. Since the SDR $\Gamma(k)$ in Eq. (22) is real-positive by definition, at least one result of Eq. (26) must be real-positive as well [assuming no measurement uncertainties in $z_s(kr)$ and $z_{\text{diff}}(kr)$]. However, for specific microphone setups, both results in Eq. (26) can become real-positive so that the true SDR cannot be identified. This problem can appear when the term A and the denominator in Eq. (26) become real-valued, or complex numbers with equal phase. The former typically arises for coincident setups as shown in Sec. IV B.

For most spaced microphone setups, the term A and the denominator in Eq. (26) become complex numbers with a different phase. Therefore, at most one result of Eq. (26) can be real-positive, which is the correct SDR value. Note that in practice, when computing the SDR from the estimated coherence, we might obtain two complex results $\Gamma(k)$. In this case, we consider the $\Gamma(k)$ with smaller imaginary part. The final SDR value is then obtained by taking the real part of this $\Gamma(k)$.

The SDR estimation with Eqs. (25)–(27) requires the spatial coherence $\gamma_s(kr)$, the diffuse field coherence $\gamma_{\text{diff}}(kr)$, and the direct sound coherence $\gamma_{\text{dir}}(kr)$. The coherence $\gamma_s(kr)$ can be computed from the microphone signals, e.g., via Eqs. (13) and (21). The coherence $\gamma_{\text{diff}}(kr)$, which would have been obtained if the sound was purely diffuse, is available as *a priori* information for the given microphone setup for specific diffuse field characteristics (see Fig. 2). Alternatively, $\gamma_{\text{diff}}(kr)$ can be estimated in advance as shown with the measurements in Sec. V. The coherence $\gamma_{\text{dir}}(kr)$, which would have been obtained if only the directional sound was present, can be determined, e.g., via Eqs. (9) and (15). Clearly, this requires us to estimate the propagation direction $\mathbf{n}_{\text{dir}}$ of the directional sound, which might be difficult particularly when the sound field is more diffuse. Nevertheless, relatively accurate SDR estimates can be obtained in practice even with noisy $\mathbf{n}_{\text{dir}}$ estimates as shown in Sec. V. Note that $\mathbf{n}_{\text{dir}}$ is also required for computing the factors q_1 and q_2 in Eqs. (26) and (27) with Eq. (23).

In the following, we discuss the SDR estimator (26) in more detail for different practical microphone setups. The main outcomes are summarized in Table II.

B. Coincident microphone setups

For coincident setups ($r = 0$) of two first-order directional microphones with different orientation or directivity, we possibly obtain two valid (real-positive) results in Eq. (26). In fact, the diffuse field coherence $\gamma_{\text{diff}}(kr)$, the coherence $\gamma_{\text{dir}}(kr)$ in Eq. (15), and thus the spatial coherence $\gamma_s(kr)$ in Eq. (22) are all real-valued for $kr = 0$. Hence, $z_{\text{diff}}(kr = 0)$ and $z_s(kr = 0)$ in Eq. (25) are also real-valued numbers. When $|z_{\text{diff}}| > |z_s|$, we obtain two valid (real-positive) results in Eq. (26) (see the second part of the Appendix). This condition is fulfilled when the absolute value of the coherence $\gamma_s(kr)$ becomes smaller than the absolute value of the diffuse field coherence, i.e., when $|\gamma_s(kr)| < |\gamma_{\text{diff}}(kr)|$, which can occur for specific SDRs and propagation directions of the directional sound (see an example in Sec. V A). Besides this ambiguity problem, the SDR estimation with coincident setups also requires knowledge of the DOA of the directional sound, namely to compute the directivity factors $g_i(\mathbf{n}_{\text{dir}})$ in Eq. (23).

C. Spaced setups with equal directivity

For setups of two spaced microphones ($kr > 0$) with an equal directivity and orientation (e.g., two microphones of a typical linear array), we usually obtain a complex spatial coherence $\gamma_s(kr)$. Solving Eq. (24) for $\Gamma(k)$ leads to the unique solution

$$\Gamma(k)|_{q_1=q_2} = \frac{1}{q^2}\frac{z_s(kr) - z_{\mathrm{diff}}(kr)}{1 - z_s(kr)}, \tag{28}$$

where $z_{\mathrm{diff}}(kr)$ and $z_s(kr)$ can be complex. Note that for this estimator, computing $z_s(kr)$ and $z_{\mathrm{diff}}(kr)$ with Eqs. (15) and (25) requires information on the microphone directivities $g_i(\mathbf{n}_{\mathrm{dir}})$ as well as on the phase μ_{dir}.

D. Setups with zero diffuse field coherence

Zero diffuse field coherence, i.e., $\gamma_{\mathrm{diff}}(kr) = 0$, is obtained, for example, for a B-format microphone (which measures the coincident sound pressure and particle velocity components) or for most arbitrary microphone setups at sufficiently high kr (cf. Fig. 2). This leads to $z_{\mathrm{diff}}(kr) = 0$. Moreover, in Eq. (22) the spatial coherence $\gamma_s(kr)$ is in phase (or in anti-phase) with $\gamma_{\mathrm{dir}}(kr)$. Thus, $z_s^2(kr) = |\gamma_s(kr)|^2$. The estimator (26) can now be simplified to

$$\Gamma(k)|_{\gamma_{\mathrm{diff}}(kr)=0} = \frac{G \pm \sqrt{G^2 + 4\left(|\gamma_s(kr)|^{-2} - 1\right)}}{2q_1q_2\left(|\gamma_s(kr)|^{-2} - 1\right)}, \tag{29}$$

where $G = q_1/q_2 + q_2/q_1$. Since $|\gamma_s(kr)|^{-2} \geq 1$, one result of Eq. (29) is real-positive, while the other is zero or real-negative. This means that the SDR can be determined without ambiguities. Equation (29) shows that for this setup, the phase of the spatial coherence $\gamma_s(kr)$ contains no information on the SDR. Thus, the SDR estimation does not require the phase μ_{dir}, but the microphone directivities $g_i(\mathbf{n}_{\mathrm{dir}})$, namely for computing the factors q_i with Eq. (23).

A special solution is obtained for two displaced ($r > 0$) omnidirectional microphones, which have zero diffuse field coherence when kr is equal to multiples of π (assuming a spherically isotropic diffuse field) and, at least approximately, for sufficiently high kr. In this case, Eq. (29) becomes

$$\Gamma(k)\Bigg|_{\substack{\gamma_{\mathrm{diff}}(kr)=0\\ q_1=q_2=1}} = \frac{|\gamma_s(kr)|}{1 - |\gamma_s(kr)|}. \tag{30}$$

This represents the only case for which the SDR can be determined without *a priori* information on the DOA of the directional sound.

V. EXPERIMENTAL RESULTS

A. Discussion of the spatial coherence

Figure 3 illustrates some important characteristics of the theoretical spatial coherence $\gamma_s(kr)$ in Eq. (22). The plots show the absolute value and phase of $\gamma_s(kr)$ as a function of $\Gamma(kr)$ for different microphone setups (summarized in Table I), kr, and propagation directions $(\varphi_{\mathrm{dir}}, \vartheta_{\mathrm{dir}})$.

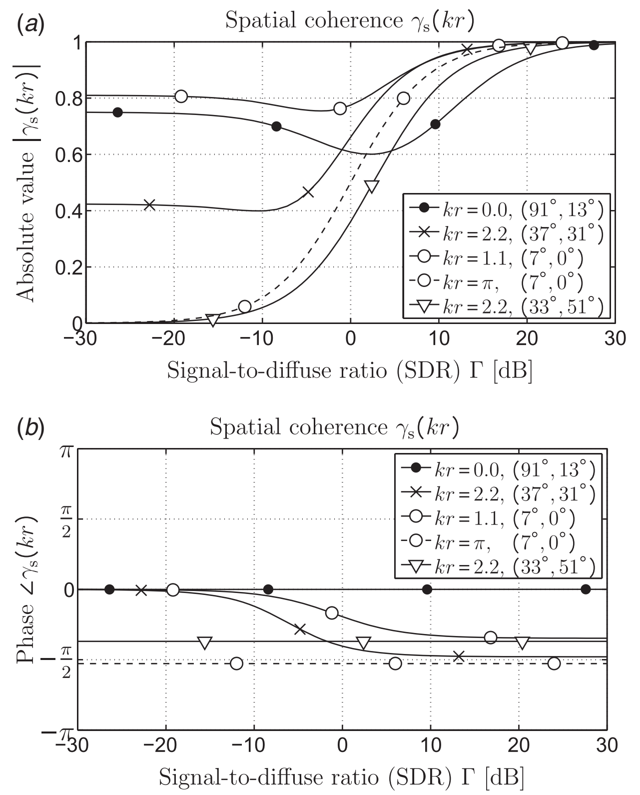

FIG. 3. Absolute value and phase of the complex spatial coherence $\gamma_s(kr)$ for a few specific microphone setups (three-dimensional sound field). For a typical microphone spacing of $r = 5.4$ cm, the kr values correspond to frequency in kilohertz.

The coincident setup (●) has a non-zero diffuse field coherence, i.e., the absolute value $|\gamma_s(kr)|$ in Fig. 3(a) is greater than zero for $\Gamma(k) \to -\infty$ dB. For increasing $\Gamma(k)$, the absolute value first decreases before it reaches its maximum of one. At the same time, the phase [Fig. 3(b)] is constant (zero) for all $\Gamma(k)$, i.e., it contains no information on the SDR. Obviously, the coherence function is not invertible for this microphone setup since different $\Gamma(k)$ can lead to the same $\gamma_s(kr)$. Thus, a unique SDR estimate cannot always be computed from $\gamma_s(kr)$. Note that such ambiguities can potentially appear for all coincident microphone setups with a non-zero diffuse field coherence.

For the spaced setups (×) and (○ solid line), the absolute value in Fig. 3(a) is also not an invertible function,

TABLE I. Microphone setups considered throughout the experiments. All arrays are aligned with the x axis.[a]

	Microphone setup	α_1	α_2	Geometry	SDR estimator
•	$r = 0$	0.5	0.5	Δ–▷	(26) + (25), (23), (15)
×	$q_1 = q_2$	0.5	0.5	Δ–Δ	(28) + (25), (23), (15)
○	$q_1 = q_2$	1.0	1.0	○–○	(28) + (25), (23), (15); (30) if $kr = \pi$
Δ	$\gamma_{\mathrm{diff}}(kr) = 0$	0.0	1.0	⊖–○	(29) + (23)

[a]Δ: Cardioid with orientation $\varphi = 90°$. ▷: Cardioid with 0°. ○: Omni. ⊖: Dipole with 90°.

TABLE II. Characteristics of the spatial coherence $\gamma_s(kr)$ with respect to its application to SDR estimation. The last column shows which DOA-specific *a priori* information is required for computing the SDR.

Microphone setup	SDR information contained in	Unique SDR results	DOA information required
Coincident ($r = 0$)	Real part of spatial coherence	Not necessarily	Yes [$g_i(\mathbf{n}_{\mathrm{dir}})$]
$r > 0$ and $q_1 = q_2$	Absolute value and phase	Yes	Yes [$g_i(\mathbf{n}_{\mathrm{dir}})$, μ_{dir}]
$\gamma_{\mathrm{diff}}(kr) = 0$	Absolute value	Yes	Yes[a] [$g_i(\mathbf{n}_{\mathrm{dir}})$]

[a]Not for two omnidirectional microphones.

similar to the coincident setup (●). Therefore, from only the absolute value of the spatial coherence, we cannot always determine a unique SDR estimate. However, the phase in Fig. 3(b) now also contains information on the SDR, i.e., the phase varies with increasing $\Gamma(k)$. Hence, unique SDR estimates can be computed from the spatial coherence by considering both the absolute value and phase (see Sec. IV).

For the microphone setups (▽) and (○ dashed line) with zero diffuse field coherence $\gamma_{\mathrm{diff}}(kr)$, the absolute value [Fig. 3(a)] is invertible. The phase [Fig. 3(b)] is constant and contains no information on the SDR. Therefore, a unique solution to $\Gamma(k)$ can be derived by considering only the absolute value of the spatial coherence. This is true for all microphone setups for which $\gamma_{\mathrm{diff}}(kr) = 0$ (see Sec. IV).

Table II summarizes the characteristics of the spatial coherence $\gamma_s(kr)$ with respect to its application to SDR estimation.

B. Discussion of the SDR estimation

1. Influence of noisy DOA estimates

In this simulation we compute the theoretical spatial coherence $\gamma_s(kr)$ with Eq. (22) for different SDRs and DOAs of the directional sound. As an example, we consider the spaced cardioid setup (×) in Table I at $kr = 2.2$. Microphone self-noise is not present. The SDR is then estimated from $\gamma_s(kr)$ as explained in Sec. IV (see also Table I). Information on the DOA of the directional sound, required to compute q_i and $\gamma_{\mathrm{dir}}(kr)$, is either available as *a priori* knowledge, or a specific DOA is assumed. Alternatively, $\gamma_{\mathrm{dir}}(kr)$ is estimated via

$$\hat{\gamma}_{\mathrm{dir}}(kr) = \frac{\gamma_s(kr)}{|\gamma_s(kr)|}, \tag{31}$$

which assumes that no diffuse sound is present. This estimator provides noisy estimates when diffuse sound is present.

Figure 4 shows the estimated SDR $\hat{\Gamma}(k)$ (coded in grayscale) as a function of the azimuth φ_{dir} and the true SDR when no *a priori* information on the DOA is available. The elevation of the directional sound is $\vartheta_{\mathrm{dir}} = 0$ in all simulations.

In Fig. 4(a), we assumed that the directional sound arrives from the broadside direction ($\varphi_{\mathrm{dir}} = 90°$). Therefore, we used $g_1(\mathbf{n}_{\mathrm{dir}}) = g_2(\mathbf{n}_{\mathrm{dir}}) = 1$ and $\gamma_{\mathrm{dir}}(kr) = 1$ for computing the SDR. The plot shows that the SDR estimation is correct only for the assumed direction $\varphi_{\mathrm{dir}} = 90°$. For all other directions, the SDR is strongly underestimated, even at high $\Gamma(k)$.

In Fig. 4(b), the coherence $\gamma_{\mathrm{dir}}(kr)$ was estimated with Eq. (31), while $g_1(\mathbf{n}_{\mathrm{dir}}) = g_2(\mathbf{n}_{\mathrm{dir}}) = 1$ was used in Eq. (23) as before. We obtain significantly more accurate results compared to Fig. 4(a), particularly when the sound is less diffuse (higher SDRs). This is especially interesting because the estimator for $\gamma_{\mathrm{dir}}(kr)$ in Eq. (31) is rather simple and inaccurate DOA information results in large estimation errors as shown in Fig. 4(a). However, the SDR estimation in Fig. 4 fails when the directional sound arrives close to the nulls of the microphones ($\varphi_{\mathrm{dir}} \to -90°$), as here the assumed $g_i(\mathbf{n}_{\mathrm{dir}})$ differ greatly from their true values.

Figure 4(c) shows the results for $kr \approx 3.43$ where $\gamma_{\mathrm{diff}}(kr)$ approaches zero. In this case, the SDR is estimated

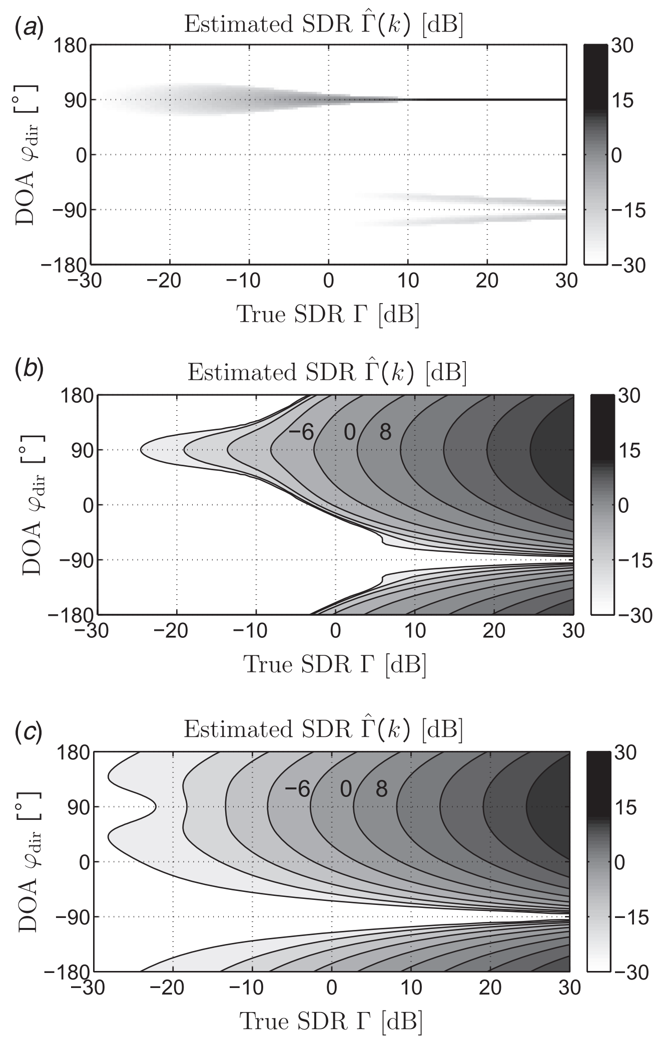

FIG. 4. Estimated SDR $\hat{\Gamma}(k)$ [dB] as a function of the true SDR $\Gamma(k)$ and azimuth φ_{dir}. No *a priori* information on the DOA of the directional sound was available.

with Eqs. (23) and (29), which requires information on $g_i(\mathbf{n}_{\mathrm{dir}})$, but not on the coherence $\gamma_{\mathrm{dir}}(kr)$. We obtain similar results compared to Fig. 4, especially for higher SDRs, indicating that the estimator for $\gamma_{\mathrm{dir}}(kr)$ in Eq. (31) provides relatively accurate results.

In general, Fig. 4 shows that accurate information on $\gamma_{\mathrm{dir}}(kr)$, if required, is crucial for the SDR estimation. Not considering the directivity factors $g_i(\mathbf{n}_{\mathrm{dir}})$ leads to a less severe, but still significant, estimation bias for specific DOAs.

2. *Influence of microphone self-noise*

In this simulation we assume that additive uncorrelated zero-mean complex Gaussian microphone noise with power $P_{N,1}(k) = P_{N,2}(k) = P_N(k)$ is present in the two microphone signals in Eq. (13). We further assume that in a practical application, the expectation in Eq. (13) is approximated by a sufficiently long time average such that the microphone noise disappears in the cross-PSD $\Phi_{12}(kr)$, while the power $P_N(k)$ remains in the auto-PSDs $\Phi_{ii}(k)$. If this is true, the PSDs $\Phi_{ij}(\cdot)$ in Eq. (13) can be computed analytically for different SDRs, signal-to-noise ratios (SNRs), and microphone setups using Eq. (19) together with Eqs. (14)–(18). We consider the microphone setups in Table I. Note that the SNR is defined as

$$\mathrm{SNR} = \frac{\mathrm{E}\{|S(k, \mathbf{d}_1)|^2\}}{P_N(k)}. \tag{32}$$

After computing $\Phi_{ij}(\cdot)$, the noisy coherence $\gamma_{12}(kr)$ is obtained with Eq. (12). The SDR is then determined with the estimators in Table I (*a priori* information on the DOA of the directional sound is available) assuming $\gamma_s(kr) = \gamma_{12}(kr)$. In doing so, we ignore the microphone self-noise. Consequently, the obtained estimates are biased. The performance of the SDR estimation is evaluated via the relation

$$\epsilon_N(k) = 10\log_{10}\left(\frac{\hat{\Gamma}(k)}{\Gamma(k)}\right), \tag{33}$$

where $\hat{\Gamma}(k)$ and $\Gamma(k)$ are the estimated and true SDR, respectively. This measure expresses the estimation error in a logarithmic domain where $\epsilon_N(k) = 0\,\mathrm{dB}$ means that no error occurred. Moreover, $\epsilon_N(k)$ indicates whether the estimates are overestimated or underestimated.

In general, Eq. (21) shows that disregarding the microphone self-noise reduces the absolute value of the spatial coherence $\gamma_s(kr)$, while the phase is not influenced. For microphone setups such as (∇) and ($\bigcirc$ if $kr = \pi$), for which the phase of $\gamma_s(kr)$ contains no information on the SDR, the reduced absolute value simply leads to an underestimation of the SDR. This becomes clear by considering Eqs. (29) and (30) or Fig. 3.

For other microphone setups such as ($\times$) and ($\bigcirc$), for which the phase of $\gamma_s(kr)$ contains information on the SDR, the effect of the reduced absolute value is less obvious. For this purpose, let us consider Fig. 5 in which the error $\epsilon_N(k)$ is depicted for setup ($\times$) under different conditions. Figure 5 shows the error as a function of the SNR and true SDR $\Gamma(k)$. The SDR becomes strongly underestimated [negative dB values for $\epsilon_N(k)$] at higher $\Gamma(k)$, even for higher SNRs. At smaller $\Gamma(k)$, the SDR is overestimated [positive dB values for $\epsilon_N(k)$] when the SNR is low. However, the estimation bias at higher $\Gamma(k)$, when the sound is less diffuse, appears more severe. This means that the microphone self-noise influences the SDR estimation particularly when the sound field is strongly directional.

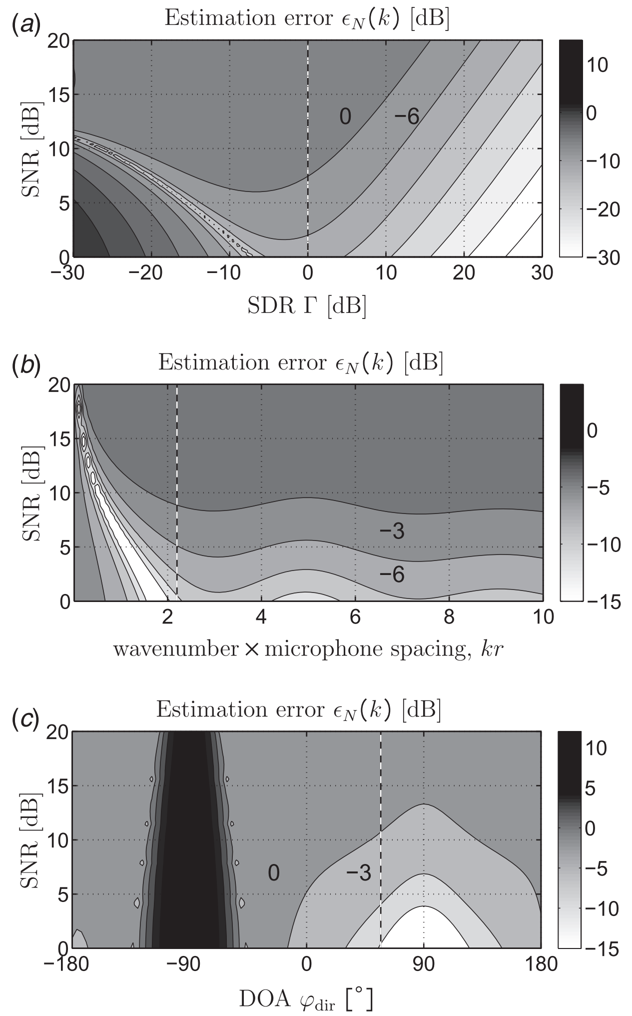

FIG. 5. Estimation error $\epsilon_N(k)$ [dB] when the microphone self-noise is disregarded. Microphone setup Δ–Δ. $\epsilon_N(k) = 0\,\mathrm{dB}$ means no error. The dashed lines indicate equal simulation parameters.

Figure 5(b) depicts the error for different kr and SNRs. The dashed lines in Figs. 5(a) and 5(b) indicate the same simulation settings, i.e., the error $\epsilon_N(k)$ is identical along both lines. The estimation error in Fig. 5(b) increases at kr values for which the diffuse field coherence $\gamma_{\mathrm{diff}}(kr)$ is higher (cf. Fig. 2), which is particularly the case toward small kr. Therefore, estimating the SDR becomes especially challenging at low frequencies or for small microphone spacings r. In fact, when $\gamma_{\mathrm{diff}}(kr)$ is high, the absolute value $|\gamma_s(kr)|$ of the true spatial coherence varies in a smaller range for the different SDRs [see for instance Fig. 3(a) plot $\bigcirc$]. Thus, a bias in the estimated spatial coherence becomes more relevant.

The results in Fig. 5(c) show the estimation error for different propagation directions φ_{dir} and SNRs. As expected, the error becomes high when the directional sound arrives close to the nulls of the microphones ($\varphi_{\mathrm{dir}} \rightarrow -90^{\circ}$).

C. Measurement results

1. Measurement setup

The presented theory was verified via measurements. For this purpose, a diffuse field $S_{\rm diff}(k, t, \mathbf{d})$ was generated by reproducing mutually uncorrelated white Gaussian noise signals from 25 loudspeakers on a hemisphere in a reverberant environment ($\mathrm{RT}_{60} \approx 360$ ms). Moreover, a directional sound field $S_{\rm dir}(k, t)$ with specific $\varphi_{\rm dir}$ was generated by reproducing a white Gaussian noise signal from a single loudspeaker in an anechoic chamber. In both measurements, the sound was captured at $f_s = 44.1$ kHz with four omnidirectional microphones arranged on the corners of a square with a diagonal of $r = 4.4$ cm. The recorded signals were transformed into the time-frequency domain with a 1024-point short-time Fourier transform with 50% overlap. The two sound fields were then added with specific SDR $\Gamma(k)$ to obtain the total sound pressure $X(k, t)$. The spatial coherence $\hat{\gamma}_s(kr)$ between two microphone signals was computed with Eqs. (12) and (21), where $\Phi_{N,11}(k) = \Phi_{N,22}(k) = 0$ was assumed due to a high SNR during the measurement. The expectation operators in Eq. (12) were approximated by averaging over K time frames. Finally, the SDR was estimated from $\hat{\gamma}_s(kr)$ via two different strategies:

(1) Via the complex spatial coherence $\hat{\gamma}_s(kr)$ using Eqs. (28) and (25) with Eq. (31) and $q_1 = q_2 = 1$ (no *a priori* information on the DOA of the directional sound).
(2) Via the real-valued spatial coherence $\mathrm{Re}\{\hat{\gamma}_s(kr)\}$ as proposed by Jeub *et al.*[18] assuming that the directional sound impinges broadside on the two microphones.

Note that the diffuse field coherence $\gamma_{\rm diff}(kr)$ for the specific measurement setup was determined in a pre-processing step via Eq. (12) from the diffuse field recording. In practice, of course, this diffuse field coherence might not always be available for a specific recording situation. We discuss this problem in the following.

2. Measurement results

Figure 6 shows the diffuse field coherence $\gamma_{\rm diff}(kr)$ measured for the two pairs of opposing microphones. A long temporal averaging over $K = 5500$ frames (≈ 60 s) was applied to approximate the expectation operators in Eq. (12). It can be observed that the two coherence functions are approximately equal showing that the diffuse field was spatially isotropic as required for the derivations in Sec. III. The average of both functions in Fig. 6 is used in the following as the diffuse field coherence $\gamma_{\rm diff}(kr)$.

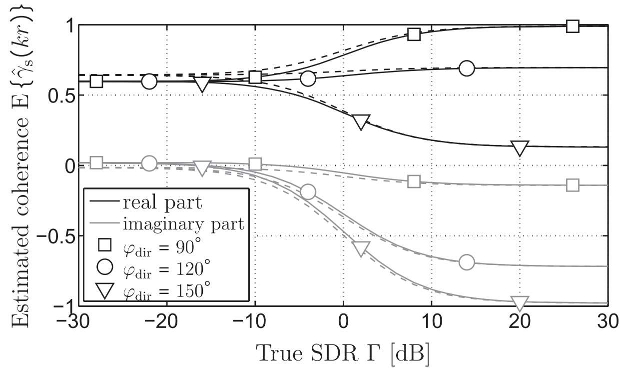

FIG. 7. Estimated spatial coherence (solid lines) for different propagation directions of the directional sound and $kr = 2.0$ (corresponding to $f = 2.48$ kHz). The temporal averaging length was $K = 10$ frames (≈ 110 ms). The dashed lines show the theoretical functions.

The mean estimated spatial coherence $\mathrm{E}\{\hat{\gamma}_s(kr)\}$ is depicted in Fig. 7 (solid lines) for different SDRs $\Gamma(k)$ and propagation directions $\varphi_{\rm dir}$ of the directional sound. A relatively short temporal averaging over $K = 10$ frames (≈ 110 ms) was applied, which is more typical for practical applications. For comparison, the dashed lines show the theoretical spatial coherence computed with Eq. (24) (using the measured diffuse field coherence). The estimated coherence follows accurately the theoretical function. The observed discrepancies toward lower $\Gamma(k)$ result from the small averaging length K used for approximating the expectation operators in Eq. (12).

The mean estimated SDR $\mathrm{E}\{\hat{\Gamma}(k)\}$ computed from the real-valued coherence $\mathrm{Re}\{\hat{\gamma}_s(kr)\}$ as proposed by Jeub *et al.*[18] is depicted in Fig. 8 as a function of the true SDR $\Gamma(k)$. For sound arriving from the broadside direction ($\varphi_{\rm dir} \approx 90^\circ$), we obtain accurate results for medium $\Gamma(k)$. For the other propagation directions, however, the SDR estimation fails. The SDR $\mathrm{E}\{\hat{\Gamma}(k)\}$ estimated from the complex $\hat{\gamma}_s(kr)$ with Eq. (28) is illustrated in Fig. 9. The SDR was accurately estimated for all DOAs particularly for medium

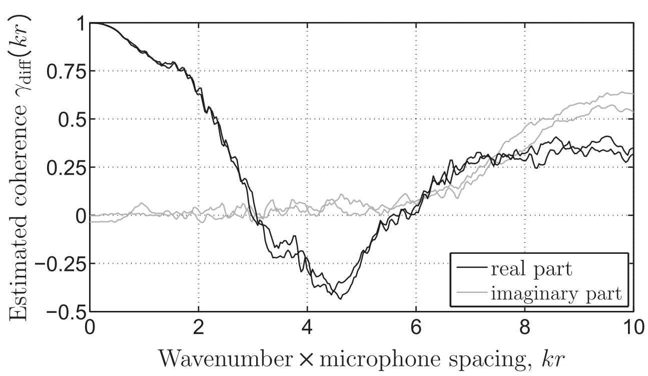

FIG. 6. Measured diffuse field coherence for two identical microphone pairs (one rotated by 90°). Diffuse field generated by 25 loudspeakers on a hemisphere. $K = 5500$.

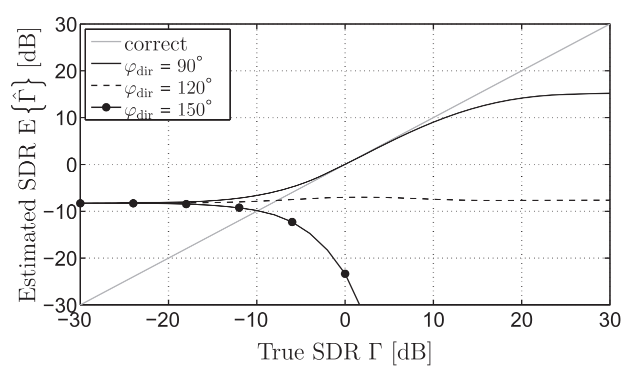

FIG. 8. SDR $\hat{\Gamma}(k)$ estimated from the real-valued spatial coherence $\mathrm{Re}\{\gamma_s(kr)\}$ as a function of the true SDR $\Gamma(k)$ at $kr = 2.0$ ($f = 2.48$ kHz) with $K = 10$ (≈ 110 ms).

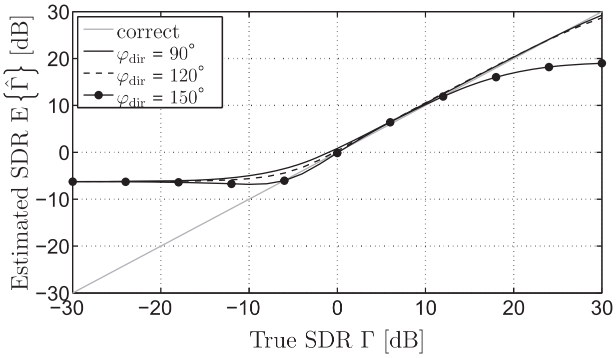

FIG. 9. SDR $\hat{\Gamma}(k)$ estimated from the complex spatial coherence $\hat{\gamma}_s(kr)$ as a function of the true SDR $\Gamma(k)$ at $kr = 2.0$ ($f = 2.48$ kHz) with $K = 10$ (≈ 110 ms).

$\Gamma(k)$. The overestimation at lower $\Gamma(k)$ follows from Fig. 7. The theoretical coherence function is relatively flat at low (also at high) $\Gamma(k)$ and thus, even a small estimation bias in $\hat{\gamma}_s(kr)$ yields a large bias in the estimated SDR $\hat{\Gamma}(k)$. The same problem occurs when the assumed diffuse field coherence $\gamma_{\text{diff}}(kr)$ differs from the true diffuse field coherence for the particular room. In this case, an SDR estimation bias appears for low SDRs. Similarly, the SDR estimation is impaired at high $\Gamma(k)$ for specific DOAs mainly due to the microphone self-noise and the noisy DOA estimates.

VI. CONCLUSIONS

The spatial coherence between two arbitrary first-order directional microphones was derived as a function of the SDR for mixed sound fields where directional sound (a single plane wave) is superimposed on a diffuse field. Based on the spatial coherence, different SDR estimators were presented for practical first-order microphone setups. It was shown that to accurately estimate the SDR in all scenarios, the coherence needs to be expressed by a complex function, in contrast to state-of-the-art methods that use either the real part or the magnitude only. Moreover, estimating the SDR requires knowledge of the DOA of the directional sound. Simulations have verified that accurate SDR estimates can be obtained even when employing noisy DOA estimates. The influence of microphone self-noise on the SDR estimation was studied showing that disregarding the noise power yields an underestimation of the SDR in strongly directional fields. Measurement results have shown that considering the complex spatial coherence together with a practical DOA estimator leads to accurate SDR estimates. With the theory presented in this contribution, one can develop an optimal SDR estimator for a given arbitrary first-order directional microphone setup that exploits the relevant information contained in the spatial coherence.

APPENDIX: DERIVATIONS

1. Spatial coherence between two arbitrary first-order microphones as function of the SDR

This section presents the derivation of the spatial coherence function (22).

Let us first rewrite the cross-PSD of the directional sound $\Phi_{\text{dir},12}(kr)$ in Eq. (14b) as

$$\Phi_{\text{dir},12} = \operatorname{sgn}(g_1 g_2)\, e^{-j\mu_{\text{dir}}} |g_1||g_2| \frac{\sqrt{Q_1 Q_2}}{\sqrt{Q_1 Q_2}} P_{\text{dir}}, \tag{A1}$$

where the dependencies on $\mathbf{n}_{\text{dir}}$, k, and kr have been omitted for brevity. Inserting Eqs. (15) and (23), we obtain

$$\Phi_{\text{dir},12} = \gamma_{\text{dir}} \frac{q_1 q_2}{\sqrt{Q_1 Q_2}} P_{\text{dir}}. \tag{A2}$$

Moreover, we can rewrite Eq. (20) as

$$\gamma_s = \frac{\Phi_{\text{dir},12} + \Phi_{\text{diff},12}}{\sqrt{(\Phi_{\text{dir},11} + \Phi_{\text{diff},11})(\Phi_{\text{dir},22} + \Phi_{\text{diff},22})}}. \tag{A3}$$

Inserting Eqs. (A2), (14a), (17a), and (17b) leads to

$$\gamma_s = \frac{\gamma_{\text{dir}} q_1 q_2 P_{\text{dir}} + \gamma_{\text{diff}} P_{\text{diff}}}{\sqrt{g_1^2 g_2^2 Q_1 Q_2 P_{\text{dir}}^2 + (g_1^2 Q_1 + g_2^2 Q_2) P_{\text{dir}} P_{\text{diff}} + P_{\text{diff}}^2}}. \tag{A4}$$

Dividing both the numerator and denominator of Eq. (A4) by P_{diff} and inserting Eqs. (5) and (23) results in Eq. (22).

2. SDR as function of the spatial coherence

Here, the derivation of the SDR estimator in Eq. (26) from the complex spatial coherence (22) is presented.

Multiplying Eq. (22) with the complex conjugate of the spatial coherence $\gamma_{\text{dir}}^*(kr)$, inserting Eq. (25), and squaring leads to

$$z_s^2(kr) = \frac{(\Gamma(k) q_1 q_2 + z_{\text{diff}}(kr))^2}{\Gamma^2(k) q_1^2 q_2^2 + \Gamma(k)(q_1^2 + q_2^2) + 1}. \tag{A5}$$

Note that $\gamma_{\text{dir}}(kr)\gamma_{\text{dir}}^*(kr) = 1$. Multiplying Eq. (A5) with the denominator at the right-hand side and equating to zero yields

$$0 = \Gamma^2(k) q_1 q_2 \left(z_s^2(kr) - 1\right) + \Gamma(k) A + \frac{z_s^2(kr) - z_{\text{diff}}^2(kr)}{q_1 q_2}, \tag{A6}$$

where A is given in Eq. (27). The solution of this quadratic equation has two results, namely

$$\Gamma(k) = \frac{A \pm \sqrt{A^2 - \left[4\left(z_s^2(kr) - 1\right)\left(z_s^2(kr) - z_{\text{diff}}^2(kr)\right)\right]}}{2 q_1 q_2 \left(1 - z_s^2(kr)\right)}. \tag{A7}$$

This solution is identical to Eq. (26). For real-valued $z_s(kr)$ and $z_{\text{diff}}(kr)$, and when $|z_s(kr)| < |z_{\text{diff}}(kr)|$, this estimator yields two valid (real-positive) results. In fact, for these conditions the term inside the square brackets is positive and therefore, the result of the square root is smaller than A. This results in two real-positive results for $\Gamma(k)$ [note that the argument of the square root cannot be negative for these conditions, since otherwise only complex $\Gamma(k)$ values are found].

[1]P. A. Naylor, E. A. P. Habets, J. Y.-C. Wen, and N. D. Gaubitch, "Models, measurement and evaluation," in *Speech Dereverberation*, edited by P. A. Naylor and N. D. Gaubitch (Springer, London, 2010), Chap. 2, pp. 21–56.

[2]P. J. Bloom, "Evaluation of a dereverberation technique with normal and impaired listeners," Br. J. Audiol. **16**, 167–176 (1982).

[3]M. Jeub, M. Schafer, T. Esch, and P. Vary, "Model-based dereverberation preserving binaural cues," IEEE Trans. Audio, Speech, Lang. Process. **18**, 1732–1745 (2010).

[4]V. Pulkki, "Spatial sound reproduction with directional audio coding," J. Audio Eng. Soc. **55**, 503–516 (2007).

[5]D. P. Jarrett, E. A. P. Habets, M. R. P. Thomas, N. D. Gaubitch, and P. A. Naylor, "Dereverberation performance of rigid and open spherical microphone arrays: Theory & simulation," in *Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)* (2011), pp. 145–150.

[6]S. H. Nielsen, "Auditory distance perception in different rooms," J. Acoust. Soc. Am. **41**, 755–770 (1993).

[7]P. Zahorik, "Assessing auditory distance perception using virtual acoustics," J. Acoust. Soc. Am. **111**, 1832–1846 (2002).

[8]P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," J. Acoust. Soc. Am. **112**, 2110–2117 (2002).

[9]E. Larsen, N. Iyer, C. R. Lansing, and A. S. Feng, "On the minimum audible difference in direct-to-reverberant energy ratio," J. Acoust. Soc. Am. **124**, 450–461 (2008).

[10]E. Larsen, C. D. Schmitz, C. R. Lansing, W. D. O'Brien, Jr., B. C. Wheeler, and A. S. Feng, "Acoustic scene analysis using estimated impulse responses," in *Conference Record of the 37th Asilomar Conference on Signals, Systems and Computers* (2003), Vol. 1, pp. 725–729.

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

[12]F. Jacobsen, "The diffuse sound field," Ph.D. thesis, The Acoustics Laboratory, Technical University of Denmark, 1979.

[13]G. W. Elko, "Spatial coherence functions for differential microphones in isotropic noise fields," in *Microphone Arrays: Signal Processing Techniques and Applications*, edited by M. Brandstein and D. Ward (Springer, Berlin, 2001), Chap. 4, pp. 61–85.

[14]M. Kuster, "Spatial correlation and coherence in reverberant acoustic fields: Extension to microphones with arbitrary first-order directivity," J. Acoust. Soc. Am. **123**, 154–162 (2008).

[15]O. Thiergart, G. Del Galdo, and E. A. P. Habets, "Diffuseness estimation with high temporal resolution via spatial coherence between virtual first-order microphones," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)* (2011), pp. 217–220.

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

[17]A. G. Piersol, "Use of coherence and phase data between two receivers in evaluation of noise environments," J. Sound Vib. **56**, 215–228 (1978).

[18]M. Jeub, C. M. Nelke, C. Beaugeant, and P. Vary, "Blind estimation of the coherent-to-diffuse energy ratio from noisy speech signals," in *19th European Signal Processing Conference (EUSIPCO 2011)* (2011), pp. 1347–1351.

[19]O. Thiergart, G. Del Galdo, and E. A. P. Habets, "Signal-to-reverberant ratio estimation based on the complex spatial coherence between omnidirectional microphones," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)* (2012), pp. 309–312.

[20]Y. Hioka, K. Niwa, S. Sakauchi, K. Furuya, and Y. Haneda, "Estimating direct-to-reverberant energy ratio based on spatial correlation model segregating direct sound and reverberation," in *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)* (2010), pp. 149–152.

[21]M. Kuster, "Estimating the direct-to-reverberant energy ratio from the coherence between coincident pressure and particle velocity," J. Acoust. Soc. Am. **130**, 3781–3787 (2011).

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

[23]J. S. Bendat and A. G. Piersol, *Engineering Applications of Correlation and Spectral Analysis*, 1st ed. (Wiley, New York, 1980).

Copyright of Journal of the Acoustical Society of America is the property of American Institute of Physics and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.