POWER-BASED SIGNAL-TO-DIFFUSE RATIO ESTIMATION USING NOISY DIRECTIONAL MICROPHONES

Oliver Thiergart, Tobias Ascherl, and Emanuël A. P. Habets

International Audio Laboratories Erlangen*
Am Wolfsmantel 33, 91058 Erlangen, Germany
{oliver.thiergart, emanuel.habets}@audiolabs-erlangen.de

ABSTRACT

The signal-to-diffuse ratio (SDR), which describes the power ratio between the direct and diffuse component of a sound field, is an important parameter in many applications. This paper proposes a power-based SDR estimator which considers the auto power spectral densities obtained by noisy directional microphones. Compared to recently proposed estimators that exploit the spatial coherence between two microphones, the power-based estimator is more robust at lower frequencies given that the microphone directivities are known with sufficiently high accuracy. The proposed estimator can incorporate more than two microphones and can therefore provide accurate SDR estimates independently of the direction-of-arrival of the direct sound. We further propose a method to determine the optimal microphone orientations for a given set of directional microphones. Simulations show the practical applicability.

Index Terms— signal-to-diffuse ratio, array signal processing, directional microphones

1. INTRODUCTION

Sound fields in reverberant environments are often modeled as a sum of a direct sound (e. g., generated by a point sound source) and reverberant sound. The power ratio between both components, usually referred to as signal-to-reverberation ratio (SRR), represents an important measure in many applications, such as speech enhancement and dereverberation [1, 2], parametric spatial audio coding [3], or evaluation of beamforming performance [4]. In these applications, it is paramount that the SRR is accurately estimated with a high temporal and spectral resolution.

Usually, reverberation is modeled as a diffuse field and the SRR is equivalent to the signal-to-diffuse ratio (SDR) describing the power ratio of the direct and diffuse sounds. Recently, various methods have been proposed to estimate the SDR in the time-frequency domain. Most methods are based on the spatial coherence between two microphones. The authors of [5] consider the real part of the complex spatial coherence between two omnidirectional microphones and assume that the direct sound arrives at the broadside of the array. In [6], the real and imaginary part of the spatial coherence between two omnidirectional microphones is considered such that no specific assumption on the direction-of-arrival (DOA) of the direct sound is required. The estimator in [7] is also based on the coherence model but can use more than two omnidirectional microphones to estimate the power of the direct and diffuse sound. In general, these coherence-based approaches suffer from a high estimation variance at low frequencies since the omnidirectional microphone signals are strongly correlated even if the sound field is diffuse. This problem can be mitigated by employing directional microphones as in [8], where the SDR is estimated based on the spatial coherence between arbitrary setups of two first-order directional microphones. In order to provide unbiased results, the estimator considers the self-noise of the microphones in the signal model but assumes that the noise power spectral densities (PSDs) are known. Unfortunately, the SDR estimation performance of [8] depends strongly on the DOA of the direct sound. The authors of [9] have proposed an approach for estimating the SDR which is not based on the coherence model but considers the auto PSDs of two beamformer signals. This estimator has outperformed [6] but requires a microphone array which can provide two beamformers with equal directivity pattern towards different directions (e. g., a circular array of identical microphones). Moreover, the estimator does not consider the self-noise of the microphones, which becomes relevant at low frequencies where the white noise gain (WNG) of the beamformers is small.

In this paper, a power-based SDR estimator similar to [9] is proposed. In contrast to [9], the auto PSDs of two or more arbitrary directional microphones are used. Compared to the coherence-based approaches, the estimator is robust also at low frequencies given that the assumed microphone directivities are accurate at the frequency of interest. The estimator considers the self-noise of the microphones in the signal model to provide unbiased results. The required noise PSDs can be measured in advance or estimated together with the SDR. Throughout the paper, we derive the optimal microphone orientations for a given set of directional microphones which allows us to obtain accurate SDR estimates independent of the DOA of the direct sound. Simulation results show that the power-based SDR estimator can outperform the coherence-based estimators especially at lower frequencies and for specific DOAs of the direct sound.

The paper is structured as follows: Section 2 introduces the signal model and formulates the problem. In Sec. 3, we derive the power-based SDR estimator and optimal microphone orientations. In Sec. 4, we evaluate the performance of the proposed SDR estimator based on simulations. The conclusions are drawn in Sec. 5.

2. SIGNAL MODEL AND PROBLEM FORMULATION

We consider a sound field where the sound pressure $S(k, t, \mathbf{d})$ in an arbitrary point $\mathbf{d}$ in a Cartesian coordinate system at time instant t and wavenumber $k = 2\pi f/c$ (frequency f, speed of sound c) is formed by a superposition of a direct component and a diffuse

This research was supported by a grant from the GIF, the German-Israeli Foundation for Scientific Research and Development.

*A joint institution of the Friedrich-Alexander-University Erlangen-Nürnberg (FAU) and Fraunhofer IIS, Germany.

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

component, i. e.,

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

The direct component $S_{\text{dir}}(k,t,\mathbf{d})$ is modeled as a single plane wave (far-field assumption) with DOA expressed by the unit-norm vector $\mathbf{n}_{\text{dir}}(k)$. We assume that the power of the direct component

$$\phi_{\text{dir}}(k,t) = \mathrm{E}\left\{|S_{\text{dir}}(k,t,\mathbf{d})|^2\right\} \quad (2)$$

is independent of $\mathbf{d}$, which is a reasonable assumption if the considered positions $\mathbf{d}$ are sufficiently close. If multiple sources (e. g., talkers) are active at the same time, the source signals must be sufficiently sparse (i. e., the signal overlap must be sufficiently small) such that the single-wave model in (1) holds. This assumption normally holds for speech signals in the time-frequency domain [10,11].

The diffuse component $S_{\text{diff}}(k,t,\mathbf{d})$ corresponds to a sound field that is assumed spatially isotropic, meaning that the sound arrives with equal strength from all directions, and spatially homogeneous, meaning that its mean power

$$\phi_{\text{diff}}(k,t) = \mathrm{E}\left\{|S_{\text{diff}}(k,t,\mathbf{d})|^2\right\} \quad (3)$$

does not vary with $\mathbf{d}$. In the following, $S_{\text{dir}}(k,t,\mathbf{d})$ and $S_{\text{diff}}(k,t,\mathbf{d})$ are assumed uncorrelated.

The power ratio between the direct component and diffuse component represents the SDR $\Gamma(k,t)$, defined as

$$\Gamma(k,t) = \frac{\phi_{\text{dir}}(k,t)}{\phi_{\text{diff}}(k,t)}. \quad (4)$$

Throughout this paper, we aim at estimating the SDR $\Gamma(k,t)$ with $M \geq 2$ directional microphones located in $\mathbf{d}_{1\ldots M}$. According to the sound field model given in (1), the i-th microphone signal with $i \in \{1,2,\ldots,M\}$ can be written as

$$X_i(k,t) = X_{\text{dir},i}(k,t) + X_{\text{diff},i}(k,t) + X_{\text{n},i}(k,t), \quad (5)$$

where $X_{\text{dir},i}(k,t)$ is the i-th microphone signal proportional to the sound pressure of the direct component, $X_{\text{diff},i}(k,t)$ is the measured diffuse component, and $X_{\text{n},i}(k,t)$ models the microphone self-noise as independent and identically distributed (i. i. d.) zero-mean complex Gaussian noise. The noise power

$$\phi_{\text{n},i}(k,t) = \mathrm{E}\left\{|X_{\text{n},i}(k,t)|^2\right\} \quad (6)$$

may differ for each microphone i, which is often the case in practice when using microphones with different directivities. The specific assumptions on $\phi_{\text{n},i}(k,t)$ are further discussed in Sec. 3.

Since all terms in (5) are assumed mutually uncorrelated, the auto PSD of the i-th microphone signal can be written as

$$\phi_{x,i}(k,t) = \mathrm{E}\left\{|X_i(k,t)|^2\right\} \quad (7a)$$

$$= g_i^2(\mathbf{n}_{\text{dir}})\,\phi_{\text{dir}}(k,t) + Q_i\,\phi_{\text{diff}}(k,t) + \phi_{\text{n},i}(k,t), \quad (7b)$$

where $g_i(\mathbf{n}_{\text{dir}})$ is the directivity function of the i-th microphone depending on $\mathbf{n}_{\text{dir}}$. For instance for first-order directional microphones,

$$g_i(\mathbf{n}_{\text{dir}}) = \alpha_i + (1-\alpha_i)\,\mathbf{n}_{\text{dir}}^{\mathrm{T}}(k)\,\mathbf{l}_i, \quad (8)$$

where $\mathbf{l}_i$ is the look direction of the i-th microphone and α_i is the shape parameter (e. g., $\alpha_i = 0.5$ for a cardioid directivity). Note that $g_i(\mathbf{n}_{\text{dir}})$ may be frequency-dependent. The factor $Q_i \leq 1$ in (7b) is inversely proportional to the directivity factor [12] that describes the sensitivity of microphone i to the diffuse sound. For first-order directional microphones in a cylindrically isotropic diffuse sound field, we have [12, 13]

$$Q_i = \alpha_i^2 + \frac{1}{2}(1-\alpha_i)^2. \quad (9)$$

The estimation of $\Gamma(k,t)$ is explained in the next section.

3. POWER-BASED SDR ESTIMATION

In the following, we propose an approach for estimating the PSDs $\phi_{\text{dir}}(k,t)$ and $\phi_{\text{diff}}(k,t)$ based on the input PSDs $\phi_{x,i}(k,t)$ in (7a). Based on the estimated PSDs, we can compute the SDR with (4).

3.1. Assuming *a priori* information on the noise PSDs

In this section, we assume that the noise PSDs $\phi_{\text{n},i}(k,t)$ in (7b) are known *a priori* or can be estimated from the microphone signals in advance, e. g., during speech pauses when no direct sound or diffuse sound is present. This typically requires that the noise PSDs are slowly time-variant or time-invariant, i. e., $\phi_{\text{n},i}(k,t) = \phi_{\text{n},i}(k)$.

Let us rewrite (7b) in vector form for M microphones as

$$\boldsymbol{\Phi}_x(k,t) = \mathbf{G}(\mathbf{n}_{\text{dir}}) \begin{bmatrix} \phi_{\text{dir}}(k,t) \\ \phi_{\text{diff}}(k,t) \end{bmatrix} + \boldsymbol{\Phi}_{\text{n}}(k), \quad (10)$$

where $\boldsymbol{\Phi}_x(k,t) = [\phi_{x,1}(k,t)\ldots\phi_{x,M}(k,t)]^{\mathrm{T}}$ contains the input PSDs, $\boldsymbol{\Phi}_{\text{n}}(k) = [\phi_{\text{n},1}(k)\ldots\phi_{\text{n},M}(k)]^{\mathrm{T}}$ contains the known noise PSDs, and

$$\mathbf{G}(\mathbf{n}_{\text{dir}}) = \begin{bmatrix} g_1^2(\mathbf{n}_{\text{dir}}) & Q_1 \\ \vdots & \vdots \\ g_M^2(\mathbf{n}_{\text{dir}}) & Q_M \end{bmatrix}. \quad (11)$$

Equation (10) can be solved for $\phi_{\text{dir}}(k,t)$ and $\phi_{\text{diff}}(k,t)$, e. g., via the least-squares (LS) approach. In this case, the estimates of the direct sound and diffuse sound PSDs are given by

$$\begin{bmatrix} \widehat{\phi}_{\text{dir}}(k,t) \\ \widehat{\phi}_{\text{diff}}(k,t) \end{bmatrix} = (\mathbf{G}^{\mathrm{T}}\mathbf{G})^{-1}\mathbf{G}^{\mathrm{T}}\boldsymbol{\Phi}_{\text{s}}(k,t), \quad (12)$$

where $\boldsymbol{\Phi}_{\text{s}}(k,t) = \boldsymbol{\Phi}_x(k,t) - \boldsymbol{\Phi}_{\text{n}}(k)$. Computing (12) requires the information of $M \geq 2$ microphones, otherwise the problem is underdetermined. Moreover, the DOA of the direct component is required to compute the elements $g_i^2(\mathbf{n}_{\text{dir}})$ in $\mathbf{G}(\mathbf{n}_{\text{dir}})$. The DOA can be estimated with well-known narrowband estimators such as ESPRIT [14] or root MUSIC [15].

For specific microphone setups and DOAs of the direct sound, the linear system in (10) can become ill-conditioned such that reliable estimates of the desired PSDs cannot be obtained in (12). For instance for $M = 2$, (12) simplifies to

$$\begin{bmatrix} \widehat{\phi}_{\text{dir}}(k,t) \\ \widehat{\phi}_{\text{diff}}(k,t) \end{bmatrix} = D(k) \begin{bmatrix} Q_2 & -Q_1 \\ -g_2^2(\mathbf{n}_{\text{dir}}) & g_1^2(\mathbf{n}_{\text{dir}}) \end{bmatrix} \boldsymbol{\Phi}_{\text{s}}(k,t), \quad (13)$$

where the determinant is given by

$$D(k) = \frac{1}{Q_2\,g_1^2(\mathbf{n}_{\text{dir}}) - Q_1\,g_2^2(\mathbf{n}_{\text{dir}})}. \quad (14)$$

Clearly, computing (13) requires that both microphones have different directivities (different α_i and Q_i) or different orientations $\mathbf{l}_i$, otherwise the denominator in (14) becomes zero for all $\mathbf{n}_{\text{dir}}(k)$. Even if different microphones or orientations are used, the denominator in (14) can approach zero, namely for specific $\mathbf{n}_{\text{dir}}(k)$, which depends on the microphone configuration. Figure 1(a) shows for which azimuth angles φ_0 the denominator in (14) becomes zero for an XY-stereophony setup. Here, we consider a two-dimensional sound field, i. e., $\mathbf{n}_{\text{dir}}(k) = [\cos\varphi_{\text{dir}}(k)\ \sin\varphi_{\text{dir}}(k)]^{\mathrm{T}}$, where $\varphi_{\text{dir}}(k)$ is the azimuth of the direct sound. The denominator becomes zero for $\varphi_{\text{dir}} = 0^\circ$ and $\varphi_{\text{dir}} = 180^\circ$. For these DOAs, both microphones

posses the same power for any SDR and the system in (10) is ill-conditioned. For DOAs $\varphi_{\text{dir}}(k)$ close to φ_0, the estimator has a poor robustness against noise since $D(k)$ in (14) becomes large. This problem can be avoided when using $M > 2$ microphones and selecting the microphone setup properly, as explained in Sec. 3.3.

3.2. Without *a priori* information on the noise PSDs

In many applications, the noise PSDs $\phi_{\text{n},i}(k,t)$ in (7b) are not known and difficult to estimate in advance. This can be the case for instance in very reverberant environments, where speech pauses without diffuse sound occur rarely. Moreover, microphones with different directivities do not necessarily contain the same self-noise power in practice, e. g., an omnidirectional microphone may contain less self-noise than a cardioid microphone. To account for different self-noise powers, we relate the noise PSDs of two microphones i and j by a specific factor $\beta_{ij}(k)$, i. e.,

$$\phi_{\text{n},i}(k,t) = \beta_{ij}(k)\phi_{\text{n},j}(k,t). \tag{15}$$

Typically, the factors $\beta_{ij}(k)$ can be assumed fixed for a given microphone configuration, whereas the noise PSDs $\phi_{\text{n},i}(k,t)$ and $\phi_{\text{n},j}(k,t)$ vary depending on the (unknown) input gain of the amplifier. Therefore, the factors $\beta_{ij}(k)$ can be determined in advance for the given microphone setup (e. g., by measuring the noise floor of the microphones) and thus, can be assumed known in the following.

Using the noise PSD of the j-th microphone signal as reference, (10) can be rewritten as

$$\boldsymbol{\Phi}_{x}(k,t) = \mathbf{F}(\mathbf{n}_{\text{dir}}) \begin{bmatrix} \phi_{\text{dir}}(k,t) \\ \phi_{\text{diff}}(k,t) \\ \phi_{\text{n},j}(k,t) \end{bmatrix}, \tag{16}$$

where $\mathbf{F}(\mathbf{n}_{\text{dir}}) = [\mathbf{G}(\mathbf{n}_{\text{dir}})\ \boldsymbol{\beta}_j(k)]$ with $\boldsymbol{\beta}_j(k) = [\beta_{1j}\ \beta_{2j} \dots \beta_{Mj}]^{\text{T}}$. An LS estimate of the PSDs of the direct sound, diffuse sound, and noise at the reference microphone is given by

$$\begin{bmatrix} \widehat{\phi}_{\text{dir}}(k,t) \\ \widehat{\phi}_{\text{diff}}(k,t) \\ \widehat{\phi}_{\text{n},j}(k,t) \end{bmatrix} = (\mathbf{F}^{\text{T}}\mathbf{F})^{-1}\mathbf{F}^{\text{T}}\boldsymbol{\Phi}_{x}(k,t). \tag{17}$$

The estimation requires $M \geq 3$ directional microphones and information on $\mathbf{n}_{\text{dir}}(k)$. In contrast to the previous subsection, the noise PSDs can be time-variant, which occurs for instance when the input gain of the system is adjusted during operation, e. g., by an automatic gain control (AGC). Similar to the previous section, the linear system in (16) can become ill-conditioned for specific DOAs and microphone setups. This problem appears for instance if all microphones have the same directivity (same α_i and Q_i) and noise sensitivity (β_{ij}). In this case, each microphone contains the same diffuse and self-noise powers. Thus, differentiating between both components is impossible. The next subsection provides further discussions.

3.3. Optimal microphone orientations

The estimators proposed in the previous subsections require to solve the linear system (10) and (16), respectively. In the following, we consider the condition number of the matrices $\mathbf{G}(\mathbf{n}_{\text{dir}})$ and $\mathbf{F}(\mathbf{n}_{\text{dir}})$ as a measure on how robust the estimators can perform for a specific microphone setup and $\mathbf{n}_{\text{dir}}(k)$.

In general, the condition number $C(\mathbf{A})$ of a matrix $\mathbf{A}$ is defined as the ratio between the largest and smallest singular value of $\mathbf{A}$. As an example, Fig. 1(b) shows the condition number $C(\mathbf{G})$ of $\mathbf{G}(\mathbf{n}_{\text{dir}})$

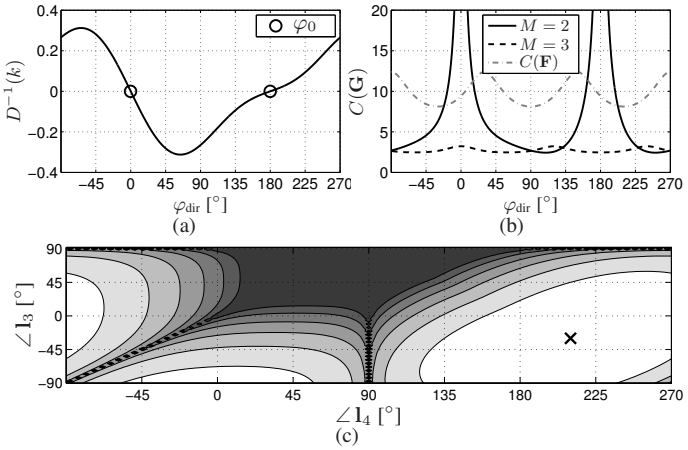


Fig. 1. Plot (a): inverse of the determinant (14), $M = 2$ cardioid microphones, look directions $\pm 45°$. Plot (b): condition number $C(\mathbf{G})$ (black lines) and $C(\mathbf{F})$ (gray dash-dot line) for different microphone setups. Plot (c): cost function $\mathcal{J}(\mathbf{F})$ (19) for $M = 4$ microphones.

as a function of the azimuth angle φ_{dir} for a setup of $M = 2$ cardioid microphones with look directions $\pm 45°$ and $M = 3$ cardioid microphones with look directions $\pm 60°$ and $180°$. When using the $M = 2$ microphones [black solid line Fig. 1(b)], $C(\mathbf{G})$ becomes large if the azimuth of the direct component φ_{dir} is close to $0°$ or $180°$. In this case, the linear system (10) becomes ill-conditioned. This problem is already visible in Fig. 1(a), where the inverse of the determinant approaches zero for the same azimuth angles. When using $M = 3$ microphones [black dashed line Fig. 1(b)], $C(\mathbf{G})$ is low for all DOAs φ_{dir}, i. e., the SDR can be estimated with almost equal sensitivity to noise for all DOAs of the direct sound.

For a given set of M microphones with specific directivity, one can determine the optimal orientation $\mathbf{l}_i$ of each microphone by minimizing the condition number over the expected angular region of the direct component, e. g.,

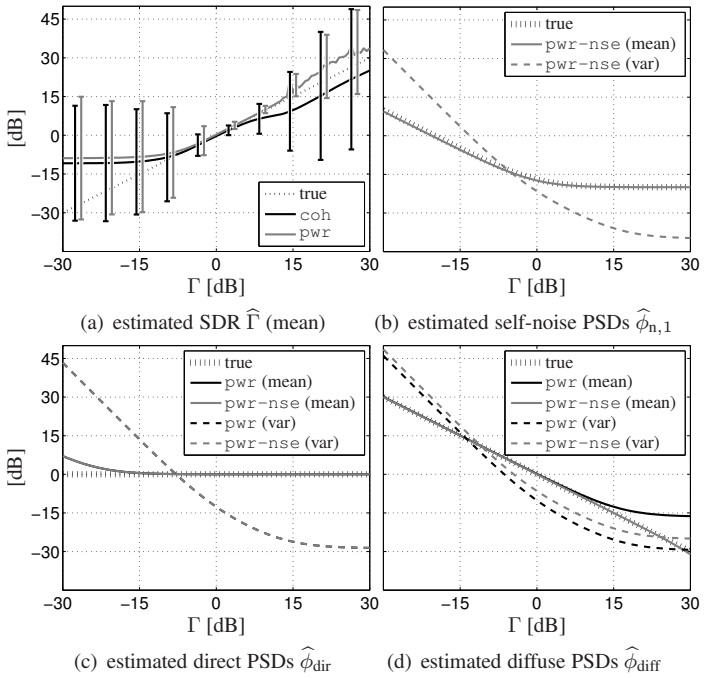
$$\{\mathbf{l}_1^{\text{opt}}, \dots, \mathbf{l}_M^{\text{opt}}\} = \underset{\mathbf{l}_1,\dots,\mathbf{l}_M}{\arg\min}\, \mathcal{J}(\mathbf{F}), \tag{18}$$

where the cost function to be minimized is given by

$$\mathcal{J}(\mathbf{F}) = \int_{-\pi}^{\pi} C(\mathbf{F})\, \mathrm{d}\varphi_{\text{dir}}. \tag{19}$$

Here, the direct component is expected to arrive from the azimuth directions $-\pi \leq \varphi_{\text{dir}} \leq \pi$. When using the estimator in Sec. 3.1, $\mathbf{F}(\mathbf{n}_{\text{dir}})$ is replaced by $\mathbf{G}(\mathbf{n}_{\text{dir}})$. The optimization problem (18) can be solved numerically when designing the microphone array.

An example cost function $\mathcal{J}(\mathbf{F})$ is illustrated in Fig. 1(c) where darker color indicates higher costs. Here, we have considered $M = 4$ microphones (one omnidirectional and three cardioid microphones). The first cardioid is oriented towards $90°$ and the look directions $\mathbf{l}_3$ and $\mathbf{l}_4$ of the other two cardioids are varied within the horizontal plane. We obtain a large $\mathcal{J}(\mathbf{F})$ if at least two of the cardioids possess similar look directions. The cross in Fig. 1(c) indicates the look directions of the two cardioids for which the cost function is minimal ($-30°$ and $210°$). The minimum of $\mathcal{J}(\mathbf{F})$ is found when orienting the three cardioids towards uniformly distributed directions. The condition number $C(\mathbf{F})$ corresponding to the optimal microphone orientations is depicted in Fig. 1(b) (gray dash-dot line). The condition number is similar for all DOAs of the direct component, i. e., the estimation of the direct, diffuse, and



(a) estimated SDR $\widehat{\Gamma}$ (mean) (b) estimated self-noise PSDs $\widehat{\phi}_{\mathrm{n},1}$

(c) estimated direct PSDs $\widehat{\phi}_{\mathrm{dir}}$ (d) estimated diffuse PSDs $\widehat{\phi}_{\mathrm{diff}}$

Fig. 2. Simulation results for the coherence-based estimator [8] (coh), power-based estimator in Sec. 3.1 (pwr), and power-based estimator in Sec. 3.2 (pwr-nse)

noise PSD can be carried out with similar robustness for all DOAs. Moreover, we notice that the condition number $C(\mathbf{F})$ is higher than the condition number $C(\mathbf{G})$ (black dashed line), even though $M=4$ microphones are used for $\mathbf{F}(\mathbf{n}_{\mathrm{dir}})$ and only $M=3$ for $\mathbf{G}(\mathbf{n}_{\mathrm{dir}})$. This means that the estimator in Sec. 3.1, which only estimates the direct and diffuse sound PSD, is more robust then the estimator in Sec. 3.2, which also estimates the noise PSD.

4. EXPERIMENTAL RESULTS

We have carried out simulations to verify the proposed approaches. A direct sound component was modeled as a single plane wave with DOA $\varphi_{\mathrm{dir}} = 9^\circ$ and wavenumber $kr = 1.6$ (corresponding to $f = 875\,\mathrm{Hz}$ for $r = 10\,\mathrm{cm}$). A two-dimensional diffuse field was generated by summing 1000 plane waves with random phases, unit magnitudes, and uniformly distributed DOAs. Both sound components were summed yielding a sound field with a specific SDR. The sound was captured with $M=3$ cardioid microphones located on a circle at $\{\pm 60^\circ, 180^\circ\}$ facing outwards (microphone spacing r). Self-noise was added to the microphone signals resulting in a specific signal-to-noise ratio (SNR) (the signal power was defined as direct plus diffuse power). The input auto PSDs were computed with (7a) where the expectation was approximated by averaging over 30 realizations. This corresponds to a temporal averaging length of approximately 300 ms for a 1024-point STFT with 50% overlap.

Figure 2(a) shows the estimated SDR $\widehat{\Gamma}$ as function of the true SDR Γ. The SNR was 60 dB. The SDR was estimated with the coherence-based approach [8, Eq. (26)] (denoted as coh) and power-based approach in Sec. 3.1 (denoted as pwr). The self-noise was considered in both estimators and *a priori* information on the noise PSDs as well as on the DOA of the direct sound was provided. For the coh approach, which can incorporate only $M = 2$ microphones directly, we were estimating the SDR separately with each of the three microphone pairs and then combining the three SDR estimates via a weighted averaging as described in [16]. Note that this optimal averaging is difficult to carry out in practice as it requires the SDR estimation variances, which were available in the simulation. As shown in Fig. 2(a), both approaches resulted in a similar mean SDR $\widehat{\Gamma}$. While the SDR at higher Γ was slightly overestimated for pwr, it was slightly underestimated for coh. At lower Γ, both approaches were overestimating the SDR which mainly was resulting from the limited temporal averaging length. In terms of estimation variance (indicated by the error bars), the pwr approach was outperforming the coh approach, especially at higher Γ. Note that since the direct sound was arriving almost at the zero of one of the cardioid microphones, microphone pairs including this microphone were not able to contribute much to the SDR estimation using the coh approach (see the results in [8]). For other DOAs and higher kr, the coh approach may outperform the pwr approach.

Figures 2(b)–(d) compare the power-based approaches proposed in Sec. 3.1 (pwr) and Sec. 3.2 (denoted as pwr-nse). We were using the same settings as before, however, a fourth microphone (omnidirectional) was added to the center of the microphone array and the SNR was 20 dB. The microphone setup was optimal for both approaches (see Sec. 3.3). Moreover, information on the noise PSDs was not available anymore and assumed unobservable in advance. Thus, for the pwr approach, we were estimating the SDR assuming there is no self-noise present, i. e., we were ignoring the presence of the noise. In contrast, the pwr-nse approach does not require information on the absolute noise power but provides an estimate of the noise PSD. We further were assuming that the relative noise sensitivities are known, which were given by $\beta_{11} = \beta_{21} = \beta_{31} = 1$ and $\beta_{41} = 0.5$, i. e., the self-noise level was 3 dB higher for the cardioid microphones than for the omnidirectional microphone. Figure 2(b) shows the noise PSD $\varphi_{\mathrm{n},1}$ (mean and variance) estimated with pwr-nse. The estimator provided unbiased results and the estimation variance was decreasing for lower levels of the self-noise power. Figure 2(c) depicts the direct sound PSDs estimated with pwr and pwr-nse. Both estimators provided accurate results and performed nearly identically in terms of mean and variance (the curves for the different estimators are lying upon each other). For both estimators, the direct sound power was overestimated at low Γ, where the direct sound was weak compared to the diffuse sound and self-noise. Figure 2(d) shows the estimated power of the diffuse field for pwr and pwr-nse. At high Γ, the diffuse PSDs were overestimated when using the pwr approach (black solid line), which is due to the self-noise that was ignored. In contrast, the pwr-nse approach provided unbiased results for all Γ (gray solid line). In terms of estimation variance (dashed lines), however, the pwr approach was outperforming the pwr-nse approach. This verifies that the pwr approach is in general more robust than the pwr-nse approach, as already discussed in Sec. 3.3.

5. CONCLUSIONS

A power-based signal-to-diffuse ratio (SDR) estimator was proposed that uses the auto power spectral density (PSD) of multiple directional microphones. The estimator considers the self-noise of the microphones in the signal model where the noise PSD can be estimated together with the SDR. Thus, the estimator can provide unbiased results in situations where the noise PSDs are unobservable in advance. The estimator can directly incorporate more than two directional microphones which allows us to carry out an accurate estimation for all directions-of-arrival (DOAs) of the direct sound. The estimator outperforms coherence-based approaches at lower frequencies and for specific DOAs of the direct sound.

6. REFERENCES

[1] P. J. Bloom, "Evaluation of a dereverberation technique with normal and impaired listeners.," *British Journal of Audiology*, vol. 16, no. 3, pp. 167–176, August 1982.

[2] M. Jeub, M. Schafer, T. Esch, and P. Vary, "Model-based dereverberation preserving binaural cues," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 18, no. 7, pp. 1732–1745, September 2010.

[3] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc*, vol. 55, no. 6, pp. 503–516, June 2007.

[4] 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 *Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011 Joint Workshop on*, June 2011, pp. 145–150.

[5] 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)*, August 2011, pp. 1347–1351.

[6] 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 *Acoustics Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, March 2012, pp. 309–312.

[7] Y. Hioka, K. Niwa, S. Sakauchi, K. Furuya, and Y. Haneda, "Estimating direct-to-reverberant energy ratio using D/R spatial correlation matrix model," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 8, pp. 2374–2384, November 2011.

[8] O. Thiergart, G. D. Galdo, and E. A. P. Habets, "On the spatial coherence in mixed sound fields and its application to signal-to-diffuse ratio estimation," *The Journal of the Acoustical Society of America*, vol. 132, no. 4, pp. 2337–2346, 2012.

[9] Y. Hioka, K. Furuya, K. Niwa, and Y. Haneda, "Estimation of direct-to-reverberation energy ratio based on isotropic and homogeneous propagation model," in *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on*, 2012.

[10] S. Rickard and Z. Yilmaz, "On the approximate W-disjoint orthogonality of speech," in *Acoustics, Speech and Signal Processing, 2002. ICASSP 2002. IEEE International Conference on*, April 2002, vol. 1.

[11] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *Signal Processing, IEEE Transactions on*, vol. 52, no. 7, pp. 1830–1847, July 2004.

[12] G. W. Elko, "Superdirectional microphone arrays," in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, Eds., chapter 10, pp. 181–237. Kluwer Academic Publishers, Dordrecht, 2000.

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

[14] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 37, no. 7, pp. 984–995, July 1989.

[15] B. Rao and K. Hari, "Performance analysis of root-music," in *Signals, Systems and Computers, 1988. Twenty-Second Asilomar Conference on*, 1988, vol. 2, pp. 578–582.

[16] Y. Chan, R. Hattin, and J. Plant, "The least squares estimation of time delay and its use in signal detection," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78.*, April 1978, vol. 3, pp. 665 – 669.