DIFFUSENESS ESTIMATION WITH HIGH TEMPORAL RESOLUTION VIA SPATIAL COHERENCE BETWEEN VIRTUAL FIRST-ORDER MICROPHONES

Oliver Thiergart[1], *Giovanni Del Galdo*[2*], *and Emanuël A. P. Habets*[1]

[1]International Audio Laboratories Erlangen[†], Erlangen, Germany
[2]Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany
Email: ame-info@iis.fraunhofer.de

ABSTRACT

The diffuseness of sound can be estimated with practical microphone setups by considering the spatial coherence between two microphone signals. In applications where small arrays of omnidirectional microphones are preferred, the diffuseness estimation is impaired by a high signal coherence in diffuse fields at lower frequencies, which is particularly problematic when carrying out the estimation with high temporal resolution. Therefore, we propose to exploit the spatial coherence between two virtual first-order microphones derived from the omnidirectional array. This represents a flexible method to accurately estimate the diffuseness in high-SNR regions at lower frequencies with high temporal resolution.

Index Terms— Spatial audio, Microphone arrays, Parameter Estimation, Spatial coherence

1. INTRODUCTION

Information on whether recorded sound is diffuse or non-diffuse is of great importance for a variety of audio applications. For instance speech dereverberation algorithms can exploit the information to suppress reverberant sound energy [1]. In source localization one can discard unreliable Direction-Of-Arrival (DOA) estimates corresponding to diffuse sound components [2]. Knowledge of the diffuseness can also be used to derive a parametric description of spatial sound, as in Directional Audio Coding (DirAC) [3].

The diffuseness can be estimated with practical microphone setups based on the Magnitude Squared Coherence (MSC) [4] between two microphone signals. The coherence value obtained for a certain field and frequency depends on the microphone setup, i. e., directivity, orientation, and spacing of the capsules. The MSC based diffuseness estimation benefits from setups providing a low signal coherence in diffuse sound fields. Such setups are especially important when the diffuseness has to be estimated with high temporal resolution (high update rate), e. g., as in speech applications where the diffuseness of the sound varies rapidly. The desired low coherence in diffuse fields can be assured using first-order microphones as presented in [5]. In many applications however, it is preferred to use small arrays of omnidirectional microphones, e. g., due to form factor constraints or when dictated by other algorithms running in parallel. Unfortunately, such setups do not provide the required low signal coherence in diffuse fields, especially at lower frequencies.

To overcome this problem, we propose to estimate the MSC between two *virtual* first-order microphones. The microphone signals are created by combining three closely spaced omnidirectional capsules, yielding the desired low signal coherence in diffuse fields. On this basis we define a diffuseness estimator, which, assuming a sufficiently high Signal-to-Noise Ratio (SNR), provides accurate results with high temporal resolution even at low frequencies.

The paper is organized as follows: Section 2 discusses the diffuseness estimation based on the traditional MSC. Section 3 derives the diffuse field MSC between two virtual coincident first-order microphones. The performance of the proposed algorithm is studied in Section 4. Section 5 draws the conclusions.

2. COHERENCE BASED DIFFUSENESS ESTIMATION

Let us assume a sound field composed by a superposition of a plane wave with arbitrary DOA (*direct sound*) and diffuse sound. In general, we define the diffuseness $\Psi(\kappa)$ in the frequency domain as

$$\Psi(\kappa) = \big(1 + \Gamma(\kappa)\big)^{-1}, \tag{1}$$

where κ is the wavenumber and $\Gamma(\kappa)$ is the Direct-to-Diffuse-Ratio (DDR) expressing the energy ratio between the direct and diffuse sound. Thus, $\Psi(\kappa) \in [0, 1]$ where 0 is obtained for $\Gamma(\kappa) = \infty$ (only direct sound), 1 for $\Gamma(\kappa) = 0$ (only diffuse sound), and 0.5 for $\Gamma(\kappa) = 1$ (both fields with equal energy). The sound field is measured using two microphones with spacing r. We assume that the microphone signals $S_1(\kappa)$ and $S_2(\kappa)$ contain independent white Gaussian microphone noise resulting in a specific SNR. The diffuseness $\Psi(\kappa)$ is estimated via the MSC $C_{12}(\kappa r)$ between $S_1(\kappa)$ and $S_2(\kappa)$. The MSC, defined e. g. in [4], can be computed with

$$C_{12}(\kappa r) = \frac{\big|\mathrm{E}\{S_1(\kappa)\,S_2^*(\kappa)\}\big|^2}{\mathrm{E}\{|S_1(\kappa)|^2\}\mathrm{E}\{|S_2(\kappa)|^2\}}, \tag{2}$$

where $(\cdot)^*$ denotes complex conjugation. The expectation $\mathrm{E}\{\cdot\}$ in the numerator yields the cross Power Spectral Density (PSD) between $S_1(\kappa)$ and $S_2(\kappa)$, while in the denominator it yields the (auto) PSDs. When operating in the short-time frequency domain, we approximate $\mathrm{E}\{\cdot\}$ by averaging over K time frames. The MSC $C_{12}(\kappa r)$ obtained for a certain κ and $\Gamma(\kappa)$ depends on the microphone configuration, i. e., directivity, orientation, and spacing r of the capsules. Figure 1 shows the theoretical MSC between two omnidirectional microphones when no microphone noise is present. The minimum is obtained for purely diffuse sound (close to one at low κr) where it follows assuming spherically isotropic fields [4]

$$C_{12}^{\mathrm{dif}}(\kappa r) = \left(\frac{\sin(\kappa r)}{\kappa r}\right)^2 = \mathrm{sinc}^2(\kappa r). \tag{3}$$

*The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. ICT-2007-214793.

†A joint institution of Fraunhofer IIS and the University of Erlangen-Nuremberg.

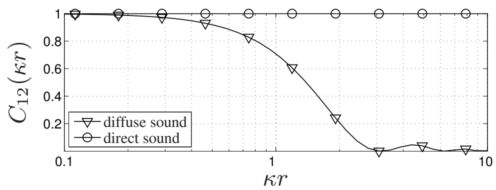


Figure 1: MSC between two omnidirectional microphones. For a typical microphone spacing of $r = 5.4\,\mathrm{cm}$ and $c = 340\,\mathrm{m/s}$, the x axis corresponds to frequency in kHz.

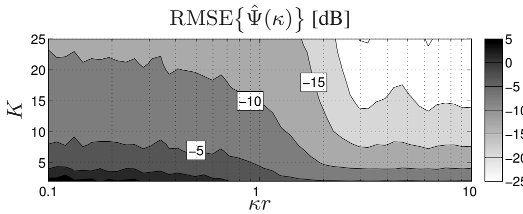


Figure 2: RMSE (dB) of $\hat{\Psi}(\kappa)$ in diffuse sound fields when using omnidirectional microphones. No microphone noise.

The MSC reaches its maximum $C_{12}^{\mathrm{dir}}(\kappa r) = 1$ for only direct sound. In practice, microphone noise reduces the MSC such that $C_{12}(\kappa r) = 0$ when only noise is present. For estimating the diffuseness, however, we assume high-SNR regions. Thanks to the knowledge of the maximum MSC $C_{12}^{\mathrm{dir}}(\kappa r)$ and minimum MSC $C_{12}^{\mathrm{dif}}(\kappa r)$, we can define a diffuseness estimator as

$$\hat{\Psi}(\kappa) = \frac{C_{12}^{\mathrm{dir}}(\kappa r) - \hat{C}_{12}(\kappa r)}{C_{12}^{\mathrm{dir}}(\kappa r) - C_{12}^{\mathrm{dif}}(\kappa r)}, \tag{4}$$

where $\hat{C}_{12}(\kappa r)$ is the measured MSC obtained with (2). The diffuseness estimator in (4) represents a linear scaling of the measured MSC to the range $[0, 1]$ such that $\hat{\Psi}(\kappa) = 1$ in purely diffuse fields and $\hat{\Psi}(\kappa) = 0$ in non-diffuse fields. For semi-diffuse fields, i. e., medium $\Gamma(\kappa)$, the linear scaling in (4) might yield biased diffuseness estimates depending on the DOA of the direct sound. In low-SNR situations, the microphone noise leads to an overestimation of the diffuseness due to the reduced coherence. The estimates can even exceed the theoretical maximum $\Psi(\kappa) = 1$, namely when MSC $\hat{C}_{12}(\kappa r)$ becomes smaller than the assumed minimum $C_{12}^{\mathrm{dif}}(\kappa r)$. In general however, the diffuseness overestimation for low SNRs is a acceptable behavior for most applications.

Even more critical for the diffuseness estimation is the estimation variance in $\hat{C}_{12}(\kappa r)$ resulting from approximating the expectations in (2) by a short temporal averaging. This is required when the diffuseness has to be estimated with high temporal resolution, as for instance in speech applications where the true diffuseness is highly time-variant. A short temporal averaging is particularly critical as it yields a high variance in $\hat{C}_{12}(\kappa r)$ even in high-SNR regions. This variance is strongly amplified in (4) when $C_{12}^{\mathrm{dif}}(\kappa r) \to C_{12}^{\mathrm{dir}}(\kappa r)$, i. e., when the signal coherence for diffuse fields is high. The impact of the variance amplification is depicted in Fig. 2 showing the Root Mean Squared Error (RMSE) of $\hat{\Psi}(\kappa)$ for a diffuse field for omnidirectional microphones. The sound field is generated in the frequency domain by summing $N = 500$ plane waves with random phases and uniformly distributed DOAs. $\hat{\Psi}(\kappa)$ is computed with (4) and (2). The expectation operators are approximated by averaging over K time frames. Although no microphone noise is present, we obtain a high RMSE especially for $\kappa r < 1$ and $K < 10$. Notice that $\kappa r = 1$ corresponds to 1 kHz for a typical microphone spacing of $r = 5.4\,\mathrm{cm}$. Moreover, $K = 10$ is a typical averaging length ($\approx$ 100 ms) when using a short-time Fourier transform at 48 kHz with 50% overlap (frame duration $\approx$ 10 ms). Figure 2 shows that $\hat{\Psi}(\kappa)$ cannot be determined accurately with high temporal resolution at low frequencies or for small r, even for very high SNRs. As mentioned above, the estimation of $\hat{\Psi}(\kappa)$ suffers mainly from a high coherence between the omnidirectional microphone signals in diffuse fields at low κr. Although directional microphones or larger r provide a lower coherence [4], they do not represent a feasible solution in many applications, e. g., when other algorithms running in parallel require small omnidirectional arrays. Thus, we propose in the next section the use of virtual first-order microphones to reduce the signal coherence in diffuse fields.

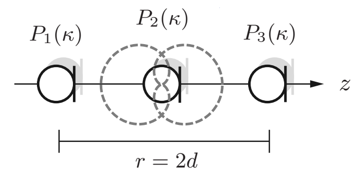


Figure 3: Array of $M = 3$ omnidirectional microphones

3. SPATIAL COHERENCE BETWEEN VIRTUAL FIRST-ORDER MICROPHONES

In this section we derive the MSC between two virtual coincident first-order microphones for spherically isotropic (diffuse) sound fields. This information is required in (4) when using such microphones for estimating the diffuseness of a sound field. Let us consider the array in Fig. 3 of $M = 3$ omnidirectional microphones with equal spacing d. The distance between the outer capsules is $r = 2d$. From the pressure signals $P_{1\ldots3}(\kappa)$ we can derive the outputs $S_1(\alpha, \kappa)$ and $S_2(\alpha, \kappa)$ of two coincident first-order microphones with opposite steering direction, i. e., [6],

$$S_i(\alpha, \kappa) = \alpha\, P_2(\kappa) + (-1)^{i+1}(1 - \alpha)\, \rho_0\, c\, V(\kappa), \tag{5}$$

where $i = \{1, 2\}$ and

$$V(\kappa) = \frac{P_1(\kappa) - P_3(\kappa)}{j\kappa r \rho_0\, c} \tag{6}$$

is a pressure gradient approximation, c is the speed of sound, ρ_0 is the mean density of air, $j = \sqrt{-1}$, and α is the shape parameter. For instance, $\alpha = 0.5$ yields cardioid directivities as shown in Fig. 3 (dashed lines). A single plane wave arriving with azimuth φ, elevation θ, and phase $\phi_0 = 0$ at the array in Fig. 3 yields

$$P_i(\kappa) = e^{j\mu_i} e^{j\phi_0} \tag{7}$$

where $i = \{1, 2, 3\}$ and $\mu_i = (1 - i)\, \kappa d \cos\theta$. The phase ϕ_0 has no effect on the derivation as it vanishes during the following steps. Moreover, (7) does not depend on the azimuth φ since the array is aligned with the z-axis. This orientation has no influence on the result since the sound field is isotropic. Inserting (7) into (5) with $\phi_0 = 0$ yields the directivity functions

$$T_i(\theta, \kappa r) = \alpha + (-1)^i (1 - \alpha) \frac{2}{\kappa r} \sin\left(\frac{\kappa r}{2} \cos\theta\right), \tag{8}$$

where $i = \{1, 2\}$. The diffuse field MSC is derived via [4]

$$C_{12}^{\mathrm{dif}}(\kappa r) = \left| \frac{N_{12}(\kappa r)}{D_{12}(\kappa r)} \right|^2, \tag{9}$$

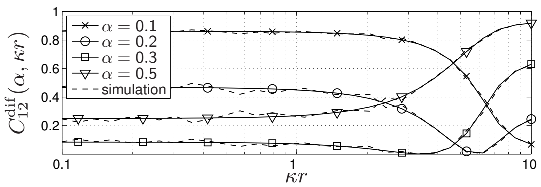


Figure 4: Solid lines: theoretical MSC between two virtual coincident first-order microphones in spherically isotropic fields. The spacing between the two outer microphones of the original array is given by r. Dashed lines: corresponding simulation results.

where the numerator $N_{12}(\kappa r)$ is defined as

$$N_{12}(\kappa r) = \frac{1}{4\pi} \int_{\mathcal{S}^2} T_1(\theta, \kappa r)\, T_2^*(\theta, \kappa r)\, e^{-j\boldsymbol{k}\cdot\boldsymbol{s}}\, \mathrm{d}\Omega \tag{10}$$

with $\int_{\mathcal{S}^2} \mathrm{d}\Omega \equiv \int_0^{2\pi}\int_0^{\pi} \sin\theta\, \mathrm{d}\theta\, \mathrm{d}\varphi$ and $\boldsymbol{k}$ being the wavevector. The vector $\boldsymbol{s}$ describes the displacement between the virtual microphones. Since the virtual microphones are coincident ($\boldsymbol{s} = \boldsymbol{0}$), we can omit the exponential in (10). The denominator $D_{12}(\kappa r)$ is

$$D_{12}(\kappa r) = \frac{1}{4\pi} \sqrt{\int_{\mathcal{S}^2} |T_1(\theta, \kappa r)|^2\, \mathrm{d}\Omega \int_{\mathcal{S}^2} |T_2(\theta, \kappa r)|^2\, \mathrm{d}\Omega}, \tag{11}$$

where both integrals are equal as the virtual microphones are identical besides their orientation. Inserting (8) into (10) yields

$$N_{12}(\kappa r) = \alpha^2 - \frac{(1-\alpha)^2}{\pi(\kappa r)^2} \int_{\mathcal{S}^2} \sin^2\left(\frac{\kappa r}{2}\cos\theta\right) \mathrm{d}\Omega. \tag{12}$$

Solving the integral results in

$$N_{12}(\kappa r) = \alpha^2 - \frac{(1-\alpha)^2}{\pi(\kappa r)^2}\, 2\pi \left(1 - \mathrm{sinc}(\kappa r)\right). \tag{13}$$

Since $T_1(\theta, \kappa r)$ is real, we have $|T_1(\theta, \kappa r)|^2 = T_1^2(\theta, \kappa r)$. Hence, substituting (8) into (11) yields

$$\begin{aligned} D_{12}(\kappa r) &= \alpha^2 + \frac{(1-\alpha)^2}{\pi(\kappa r)^2} \int_{\mathcal{S}^2} \sin^2\left(\frac{\kappa r}{2}\cos\theta\right) \mathrm{d}\Omega \\ &= \alpha^2 + \frac{(1-\alpha)^2}{\pi(\kappa r)^2}\, 2\pi \left(1 - \mathrm{sinc}(\kappa r)\right). \end{aligned} \tag{14}$$

Finally, the diffuse sound MSC $C_{12}^{\mathrm{dif}}(\alpha, \kappa r)$ between the two virtual microphones can be computed using (9), (13), (14), i. e.,

$$C_{12}^{\mathrm{dif}}(\alpha, \kappa r) = \left| \frac{\alpha^2 - A(\alpha, \kappa r)}{\alpha^2 + A(\alpha, \kappa r)} \right|^2, \tag{15}$$

where

$$A(\alpha, \kappa r) = \frac{2(1-\alpha)^2}{(\kappa r)^2} \left(1 - \mathrm{sinc}(\kappa r)\right). \tag{16}$$

Figure 4 (solid lines) illustrates the theoretical $C_{12}^{\mathrm{dif}}(\alpha, \kappa r)$ computed with (15) and (16). The dashed lines show the corresponding simulation results using (5) in (2) where $K = 1000$. The theoretical functions match exactly the simulation results. The lowest signal coherence for low κr is obtained for $\alpha \approx 0.3$. All coherence functions approach one at high κr. Comparing Fig. 4 and Fig. 1 shows that the low-frequency coherence between the measured signals in a diffuse field can be significantly reduced when using the virtual first-order microphones.

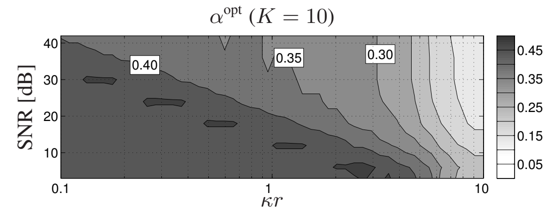


Figure 5: Numerical solution of the optimization problem in (17)

In non-diffuse fields, where only a single plane wave arrives at the microphone array, we obtain $C_{12}^{\mathrm{dir}}(\alpha, \kappa r) = 1$ for any DOA, unless a virtual microphone possesses an exact null for the DOA.

4. EVALUATION

This section presents simulation results to verify the proposed diffuseness estimation based on virtual first-order microphones. First, we study the influence of the shape parameter α in (5) on the diffuseness estimation. The optimal α to be used in practice can be determined by solving an optimization problem, e. g.,

$$\alpha^{\mathrm{opt}}(\kappa r) = \arg\min_{\alpha} \Big\{ \underbrace{\epsilon_{\mathrm{dir}}(\kappa r) + \epsilon_{\mathrm{dif}}(\kappa r)}_{\epsilon_{\mathrm{tot}}(\kappa r)} \Big\}, \tag{17}$$

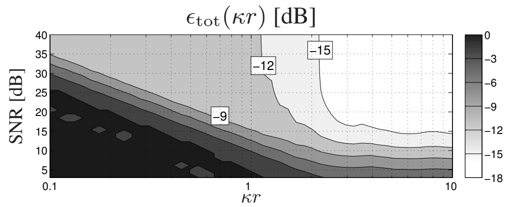
where $\epsilon_{\mathrm{dir}}(\kappa r)$ is the RMSE of the estimated diffuseness $\hat{\Psi}(\kappa)$ when only direct sound is present and $\epsilon_{\mathrm{dif}}(\kappa r)$ is the corresponding error for purely diffuse (spherical isotropic) fields. The optimization problem has to be formulated specifically for the application. Here, we define $\alpha^{\mathrm{opt}}(\kappa r)$ in terms of low estimation error of the diffuseness for both diffuse and non-diffuse fields. The solution to (17) is found numerically via simulations. The diffuse field is generated in the frequency domain by summing $N = 500$ plane waves with random phases and uniformly distributed DOAs. For the direct sound we simulate a single plane wave arriving orthogonal to the microphone array, which for our applications represents the most relevant direction. The virtual microphone signals $S_i(\alpha, \kappa)$ are computed for different α using (5), where white Gaussian microphone noise is added to the pressure signals $P_i(\kappa)$ to achieve a specific SNR. The diffuseness $\hat{\Psi}(\kappa)$ is estimated using (4) together with (15) and (2) with $K = 10$ (relatively high temporal resolution). Figure 5 depicts $\alpha^{\mathrm{opt}}(\kappa r)$ for different κr and SNRs. We obtain $\alpha^{\mathrm{opt}} \approx 0.4$ for lower κr and SNRs, while $\alpha^{\mathrm{opt}} \to 0$ for higher κr and SNRs. Thus, at lower frequencies and for smaller r and SNRs we prefer a virtual directivity which is more cardioid, whereas at higher frequencies we prefer bidirectional directivities.

In the following, the proposed diffuseness estimator is compared with two similar approaches in terms of estimation performance. All three methods employ the omnidirectional microphone array in Fig. 3. Specifically, the diffuseness is estimated by

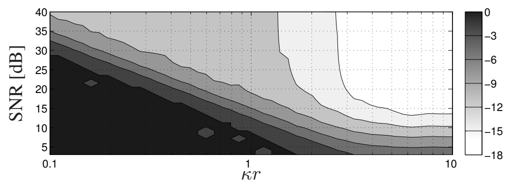
(A) computing $\hat{\Psi}_i(\kappa)$ for each omnidirectional microphone pair $i = \{1, 2, 3\}$ with the traditional method in Section 2. Combining the results with maximal ratio combining [7].

(B) computing $\hat{\Psi}(\kappa)$ using (4) where the Generalized Magnitude Squared Coherence (GMSC) algorithm [8] is used to estimate $\hat{C}_{12}(\kappa r)$ between all $M = 3$ microphones. $C_{12}^{\mathrm{dif}}(\kappa r)$ is found numerically.

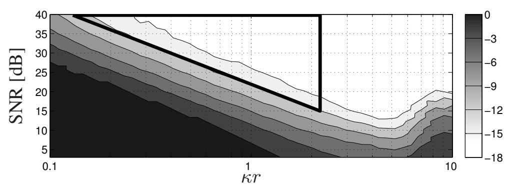
(C) computing $\hat{\Psi}(\kappa)$ via the proposed virtual first-order microphones (considering α^{opt}) as explained in the beginning of this section.



(a) traditional, approach (A)



(b) GMSC, approach (B)

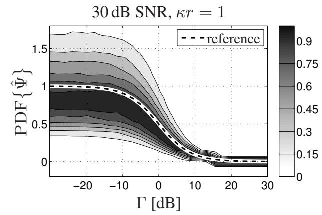


(c) proposed, approach (C)

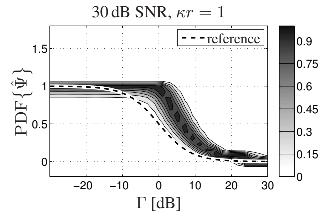
Figure 6: RMSE of the estimated diffuseness $\hat{\Psi}(\kappa)$ considering non-diffuse and diffuse fields. In all simulations $K = 10$ is used.

Figure 6 depicts for the three approaches the total error $\epsilon_{\mathrm{tot}}(\kappa r)$ defined in (17). The error is computed via simulations as explained in the beginning of this section. The traditional approach (A) and the GMSC based approach (B) yield nearly the same performance. The error $\epsilon_{\mathrm{tot}}(\kappa r)$ increases towards lower κr where it remains relatively high even for better SNRs. The proposed estimator (C) shown in (c) provides a clearly higher accuracy at lower κr when the SNR is sufficiently high (marked area). For instance for an SNR of 30 dB at $\kappa r = 1$, the error using (C) is 6 dB lower compared to (A). Notice that this SNR and κr represents a typical operation point in speech processing (especially during the important speech onsets). In general, the proposed approach becomes even more beneficial when further increasing the temporal resolution (smaller K). Nevertheless, (C) performs slightly worse compared to (A) and (B) for small SNRs due to the pressure gradient approximation in (6) which is highly sensitive to microphone noise at low frequencies. At higher frequencies ($kr > \pi$), spatial aliasing appears in the pressure gradient approximation in (6). However, as shown by the low error in Fig. 6(c), the estimation via (C) is not influenced by this effect.

The traditional approach (A) and the proposed approach (C) are compared in Fig. 7 for different DDRs Γ. The plots depict the Probability Density Function (PDF) of $\hat{\Psi}(\kappa)$ for 30 dB SNR and $\kappa r = 1$. The dashed line shows the ideal diffuseness defined in (1). Estimator (C) yields a clearly lower estimation variance for small Γ (diffuse fields). For large Γ (direct sound), both approaches show nearly the same high accuracy. The linear scaling in (4) leads to a diffuseness overestimation for medium Γ when using (C). The same effect would appear for estimator (A) when the direct sound did not arrive from broadside direction. The biased diffuseness estimation for medium Γ is of minor relevance for many practical applications, such as spatial sound reproduction with [3] or source localization with [2]. The effect is therefore left open to future research.



(a) traditional, approach (A)



(b) proposed, approach (C)

Figure 7: PDF of the estimated diffuseness $\hat{\Psi}$ when $K = 10$

5. CONCLUSIONS

This contribution discusses the estimation of the diffuseness of sound via the spatial coherence between two microphone signals. Using this method with small omnidirectional microphone arrays, which are preferred in many applications, suffers from a poor estimation performance at low frequencies resulting from a high coherence between the microphone signals in diffuse sound fields. To overcome this problem, we propose to compute the spatial coherence between two virtual first-order microphones generated from the omnidirectional array. It is shown analytically that the virtual microphones provide signals with lower coherence in diffuse sound fields at low frequencies. This is particularly beneficial when estimating the diffuseness with high temporal resolution, as required for instance in speech applications where the true diffuseness of the field varies strongly over time. The proposed method outperforms similar approaches mainly at frequencies and for signal-to-noise ratios which are especially relevant in speech processing.

6. REFERENCES

[1] E. A. P. Habets, S. Gannot, and I. Cohen, "Dual-microphone speech dereverberation in a noisy environment," in *Signal Processing and Information Technology, 2006 IEEE International Symposium on*, August 2006, pp. 651–655.

[2] O. Thiergart, R. Schultz-Amling, G. Del Galdo, D. Mahne, and F. Kuech, "Localization of sound sources in reverberant environments based on directional audio coding parameters," in *Audio Engineering Society Convention 127*, New York NY, USA, October 2009.

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

[4] 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. Springer, 2001, ch. 4, pp. 61–85.

[5] M. Kuster, "Spatial correlation and coherence in reverberant acoustic fields: Extension to microphones with arbitrary first-order directivity," *The Journal of the Acoustical Society of America*, vol. 123, no. 1, pp. 154–162, January 2008.

[6] F. J. Fahy, *Sound Intensity*. Elsevier Applied Science, 1989.

[7] G. L. Stüber, *Principles of Mobile Communication*, 1st ed. Kluwer Academic Publishers, 1996.

[8] D. Ramirez, J. Via, and I. Santamaria, "A generalization of the magnitude squared coherence spectrum for more than two signals: definition, properties and estimation," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, April 2008, pp. 3769–3772.