Coherence-Based Diffuseness Estimation in the Spherical Harmonic Domain

Daniel P. Jarrett†, Oliver Thiergart†, Emanuël A. P. Habets† and Patrick A. Naylor∗

*Dept. of Electrical & Electronic Engineering, Imperial College London, UK
Email: {daniel.jarrett05,p.naylor}@imperial.ac.uk
†International Audio Laboratories Erlangen, Germany
Email: {oliver.thiergart,emanuel.habets}@audiolabs-erlangen.de

Abstract—The diffuseness of sound fields has previously been estimated in the spatial domain using the spatial coherence between a pair of microphones (omnidirectional or first-order). In this paper, we propose a diffuseness estimator for spherical microphone arrays based on the coherence between eigenbeams, which result from a spherical harmonic decomposition of the sound field. The weighted averaging of the diffuseness estimates over all eigenbeam pairs is shown to significantly reduce the variance of our estimates.

II. PROBLEM FORMULATION

In the following, we work in the short-time Fourier transform (STFT) domain, where \( k \) denotes the discrete frequency index and \( t \) denotes the time frame index.

A. Spatial and Spherical Harmonic Domain Signal Models

In the spatial domain, the signal \( X(k, r, t) \) received at a microphone position \( r = (r, \Omega) = (r, \theta, \phi) \) (in spherical coordinates) is modeled as the sum of a directional signal \( X_{\text{dir}} \), a diffuse signal \( X_{\text{diff}} \) and a sensor noise signal \( V \), i.e.,

\[
X(k, r, t) = X_{\text{dir}}(k, r, t, \Omega_{\text{dir}}) + X_{\text{diff}}(k, r, t) + V(k, r, t).
\]

(1)

The directional signal \( X_{\text{dir}} \) corresponds to a plane wave incident from a DOA \( \Omega_{\text{dir}} \). The diffuse signal \( X_{\text{diff}} \) is composed of an infinite number of independent plane waves with equal amplitude, random phase and uniformly distributed DOA [9]. The powers of the directional and diffuse signals received at a (virtual) omnidirectional reference microphone \( M_{\text{ref}} \) placed at the center of the array are denoted as \( P_{\text{dir}}(k, t) \) and \( P_{\text{diff}}(k, t) \), respectively.

When dealing with spherical microphone arrays, it is convenient to work in the SHD, particularly for rigid arrays whose scattering behaviour can be described analytically in the SHD. The spherical Fourier transform of a spatial domain signal \( X(k, r, t) \) is given by [10, p. 192]

\[
X_{lm}(k, t) = \int_{\Omega \in S^2} X(k, r, t) Y_{lm}^* (\Omega) d\Omega,
\]

(2)

where the order \( l \) and degree \( m \) are the transform parameters, \( \int_{\Omega \in S^2} d\Omega \triangleq \int_0^{2\pi} \int_0^\pi \sin \theta d\theta d\phi \), the basis functions \( Y_{lm} \) are the spherical harmonics [10, p. 190], and \((\cdot)^*\) denotes the complex conjugate. The spherical Fourier transform coefficient \( X_{lm} \) is often called the eigenbeam of order \( l \) and degree \( m \), due to the fact that the spherical harmonics are the eigensolutions of the acoustic wave equation in spherical coordinates.

The spherical Fourier transform in (2) requires the sound pressure \( X \) to be known on the entire sphere. In practice, however, the sound field is sampled by a number of microphones positioned at discrete points on the sphere, i.e., the integral in (2) is approximated by a sum over \( Q \) positions. All spatial sampling schemes require at least \( Q = (L + 1)^2 \) microphones to sample a sound field up to order \( l = L \). In the
following, we assume perfect spatial sampling; the effects of spatial aliasing [11] are therefore neglected.

Using the spherical Fourier transform in (2), the spatial domain signal model (1) can now be expressed in the SHD:

\[ X_{lm}(k, t) = X^\text{dir}_{lm}(k, t, \Omega_{\text{dir}}) + X^\text{diff}_{lm}(k, t) + V_{lm}(k, t) \quad (3) \]

where \( X^\text{dir}_{lm}, X^\text{diff}_{lm} \) and \( V_{lm} \) are respectively the spherical Fourier transforms of \( X \), \( X^\text{dir} \), and \( V \).

The directional signal \( X^\text{dir}_{lm}(k, t) \) can be expressed in the SHD as [12]

\[ X^\text{dir}_{lm}(k, t, \Omega_{\text{dir}}) = \sqrt{P_{\text{dir}}(k,t)} \varphi_{\text{dir}}(k, t) 4\pi b_l(k) Y^*_{lm}(\Omega_{\text{dir}}), \quad (4) \]

where \( \varphi_{\text{dir}}(k, t) \) is the wave phase (with \( |\varphi_{\text{dir}}(k, t)| = 1 \forall k, t \)).

The mode strength \( b_l(k) \) is a function of the array properties (configuration, microphone type, radius), mode strength expressions for various configurations (open, rigid, dual-sphere, etc.) can be found in [13].

The diffuse signal \( X^\text{diff}_{lm}(k, t) \) is expressed in the SHD as

\[ X^\text{diff}_{lm}(k, t) = \frac{P_{\text{diff}}(k,t)}{4\pi} \int_{\Omega \in S^2} \varphi_{\text{diff}}(k,t,\Omega) 4\pi b_l(k) Y^*_{lm}(\Omega) d\Omega, \quad (5) \]

where \( \varphi_{\text{diff}}(k, t, \Omega) \) is the phase of the wave with DoA \( \Omega \) (with \( |\varphi_{\text{diff}}(k, t, \Omega)| = 1 \forall k, t, \Omega \)). As the plane waves are independent, the wave phases satisfy the property \( E[\varphi_{\text{diff}}(k,t,\Omega)\varphi_{\text{diff}}^*(k,t',\Omega')] = \delta_{\Omega,\Omega'} \), where \( \delta \) is the Kronecker delta and \( E[\cdot] \) denotes mathematical expectation.

The signal received at the reference microphone \( M_{\text{ref}} \) is given by \( X_{00}(k,t)/\sqrt{4\pi b_0(k)} \) [14]. Using this relationship and the fact that \( |Y_{00}(\cdot)|^2 = (4\pi)^{-1} \), it can be verified that the powers of the directional and diffuse signals received at \( M_{\text{ref}} \) are given by \( P_{\text{dir}} \) and \( P_{\text{diff}} \), respectively.

**B. Signal to Diffuse Ratio and Diffuseness**

The signal to diffuse ratio (SDR) \( \Gamma \) at \( M_{\text{ref}} \) is given by

\[ \Gamma(k,t) = \frac{|X_{00}^\text{dir}(k,t,\Omega_{\text{dir}})|^2}{E[|X_{00}^\text{diff}(k,t)|^2]} = \frac{P_{\text{dir}}(k,t)}{P_{\text{diff}}(k,t)}. \quad (6) \]

The diffuseness \( \Psi \) of the sound field can be defined as [15]

\[ \Psi(k,t) = [1 + \Gamma(k,t)]^{-1}. \quad (7) \]

We have \( \Psi(k,t) \in [0, 1] \), where a diffuseness of 0 is obtained for \( \Gamma(k,t) \to \infty \) (purely directional field), 1 for \( \Gamma(k,t) = 0 \) (purely diffuse field), and 0.5 for \( \Gamma(k,t) = 1 \) (equal energy directional and diffuse fields).

In the following, we aim to estimate the diffuseness in (7) from the sound field observed using a spherical array.

**III. SDR ESTIMATION USING SPATIAL COHERSITY**

In this section, we propose a method to estimate the SDR using the spatial coherence between the SHD signals (i.e., the eigenbeams). The estimated SDRs are then mapped to obtain the estimated diffuseness values using (7).

\[ \text{It should be noted that in (4) and the expressions that follow, we have extracted the } 4\pi \text{ scaling factor from the mode strength given in [13].} \]

**A. Spatial Coherence**

The complex spatial coherence between the eigenbeams \( X^\text{dir}_{lm}(k, t) \) and \( X^\text{dir}_{l'm'}(k, t) \) is defined for \( (l, m) \neq (l', m') \) as

\[ \gamma^\text{dir}_{lm,l'm'}(k, t) = \frac{\Phi^\text{dir}_{lm,l'm'}(k, t)}{\sqrt{\Phi^\text{dir}_{lm}(k, t) \Phi^\text{dir}_{l'm'}(k, t)}}, \quad (8) \]

where the power spectral densities (PSDs) \( \Phi \) are given by

\[ \Phi^\text{dir}_{lm,l'm'}(k, t) = E[X^\text{dir}_{lm}(k, t) X^\text{dir}_{l'm'}(k, t)]. \quad (9) \]

We now determine expressions for the spatial coherence in purely directional and purely diffuse fields, in order to express the coherence in a mixed field as a function of the SDR \( \Gamma \).

**For purely directional sound**, using (4) and (9) the PSD \( \Phi^\text{dir}_{lm,l'm'} \) is expressed as

\[ \Phi^\text{dir}_{lm,l'm'}(k, t) = P^\text{dir}(k,t)(4\pi)^2 b_l(k)b_{l'}(k)Y^*_{lm}(\Omega_{\text{dir}})Y_{l'm'}(\Omega_{\text{dir}}), \quad (10) \]

and the directional field coherence \( \gamma^\text{dir}_{lm,l'm'} \) is given by

\[ \gamma^\text{dir}_{lm,l'm'}(k, t) = b_l(k)b_{l'}(k)Y^*_{lm}(\Omega_{\text{dir}})Y_{l'm'}(\Omega_{\text{dir}}) \left[ \frac{b_l(k)b_{l'}(k)Y^*_{lm}(\Omega_{\text{dir}})Y_{l'm'}(\Omega_{\text{dir}})}{b_l(k)b_{l'}(k)Y^*_{lm}(\Omega_{\text{dir}})Y_{l'm'}(\Omega_{\text{dir}})} \right]^\frac{1}{2}. \quad (11) \]

For purely directional sound, the coherence \( \gamma^\text{dir}_{lm,l'm'} \) therefore has unit magnitude.

**For purely diffuse sound**, using (5), (9) and the orthonormality of the spherical harmonics [10], the PSD \( \Phi^\text{diff}_{lm,l'm'} \) is expressed as

\[ \Phi^\text{diff}_{lm,l'm'}(k, t) = P^\text{diff}(k,t)(4\pi)^2 b_l(k)b_{l'}(k)Y^*_{lm}(\Omega_{\text{dir}})Y_{l'm'}(\Omega_{\text{dir}}), \quad (12a) \]

\[ \times \int_{\Omega \in S^2} b_l(k)b_{l'}(k)Y^*_{lm}(\Omega)Y_{l'm'}(\Omega)d\Omega = P^\text{diff}(k,t)(4\pi)^2 b_l(k)b_{l'}(k)\delta_{l,l'}\delta_{m,m'}. \quad (12b) \]

The diffuse field coherence \( \gamma^\text{diff}_{lm,l'm'} \) is then given by

\[ \gamma^\text{diff}_{lm,l'm'}(k, t) = \frac{b_l(k)b_{l'}(k)}{|b_l(k)||b_{l'}(k)|} \delta_{l-l'}\delta_{m-m'} = 0 \quad (13) \]

providing \( (l, m) \neq (l', m') \).

The sensor noise \( V \) is assumed to be spatially incoherent noise of equal power \( P^N \) at each of the \( Q \) equidistant microphones. The SHD noise \( V_{lm} \) is therefore also incoherent across \( l \) and \( m \) and the PSD \( \Phi^N_{lm,l'm'} \) of the noise can be calculated as [10, eqn 7.31]

\[ \Phi^N_{lm,l'm'}(k, t) = E[V_{lm}(k, t)V^*_{l'm'}(k, t)] \quad (14a) \]

\[ = \frac{P^N(4\pi)^2}{Q} \delta_{l-l'}\delta_{m-m'}. \quad (14b) \]

The power of the noise at the reference microphone \( M_{\text{ref}} \) is then given by \( P^N/Q |b_0(k)|^2 \), i.e., it has been reduced by a factor of \( Q |b_0(k)|^2 \).

**In a mixed sound field**, both the directional and diffuse sound fields \( X^\text{dir} \) and \( X^\text{diff} \) are present, in addition to incoherent noise \( V \). We assume they are mutually uncorrelated, such that the PSD \( \Phi_{lm,l'm'} \) is equal to the sum of the individual PSDs, i.e.,

\[ \Phi_{lm,l'm'}(k, t) = \Phi^\text{dir}_{lm,l'm'}(k, t) + \Phi^\text{diff}_{lm,l'm'}(k, t) + \Phi^N_{lm,l'm'}(k, t). \quad (15) \]
We define the *noiseless coherence* as
\[
\gamma_{lm,l'm'}(k, t) = \frac{\Phi_{lm,l'm'}^b(k, t)}{\sqrt{\Phi_{lm,l}(k, t)\Phi_{l'm',l'm'}^b(k, t)}} \tag{16}
\]
where the noiseless PSD \(\Phi_{lm,l'm'}^b(k, t)\) is defined as \(\Phi_{lm,l'm'}^b(k, t) = \Phi_{lm,l'm'}^\text{dir}(k, t) + \Phi_{lm,l'm'}^\text{diff}(k, t)\). Using (10) and (12), the noiseless PSD can be expressed as
\[
\Phi_{lm,l'm'}^b(k, t) = 4\pi b_l(k) b_{l'}(k) \times \left[ 4\pi P_{\text{dir}}(k, t) Y_{lm}^\ast(\Omega_{\text{dir}}) Y_{l'm'}(\Omega_{\text{dir}}) + P_{\text{diff}}(k, t) \delta_{l-l'} \delta_{m-m'} \right] \tag{17}
\]
By substituting (17) in (16), and using (6), it can straightforwardly be shown that
\[
\gamma_{lm,l'm'}^\prime(k, t) = \frac{\Gamma(k, t) \gamma_{lm,l'm'}^\text{dir}(k, t) c_{lm} c_{l'm'}}{\sqrt{2\Gamma(k, t) c_{lm}^2 c_{l'm'}^2 + \Gamma(k, t) (c_{lm}^2 + c_{l'm'}^2) + 1}} \tag{18}
\]
where we have defined \(c_{lm} = \sqrt{4\pi} Y_{lm}(\Omega_{\text{dir}})\).

The noiseless PSDs in (16) cannot be directly observed, however as the noise \(V_{lm}\) is incoherent across \(l\) and \(m\), with sufficient time averaging the noise cross PSD \(\Phi_{lm,l'm'}^\text{dir}\) will average to zero in \(\Phi_{lm,l'm'}^\text{n}\). The noiseless auto PSD can be estimated providing an estimate of the noise power \(P_N\) is available. For simplicity, in this work we will assume a sufficiently high signal to noise ratio (SNR) and estimate the noiseless coherence directly from the noisy signals, i.e., we will not compensate for the noise. The effect of sensor noise on the estimation will be discussed in Sec. V.

**B. SDR Estimation**

The SDR is determined by first computing the coherence between pairs of eigenbeams \(X_{lm}(k, t)\) and \(X_{l'm'}(k, t)\). The SDR for each specific eigenbeam pair is then found by solving for \(\Gamma(k, t)\) in (18)\(^2\), as in [4]:
\[
\hat{\Gamma}_{lm,l'm'}(k, t) = \frac{G + \sqrt{G^2 + 4 \left( \left| \gamma_{lm,l'm'}^\prime \right|^2 - 1 \right)}}{2c_{lm} c_{l'm'} \left( \left| \gamma_{lm,l'm'}^\prime \right|^2 - 1 \right)}, \tag{19}
\]
where we have defined
\[
G = \frac{c_{lm}}{c_{l'm'}} + \frac{c_{l'm'}}{c_{lm}}. \tag{20}
\]
In order to compute \(c_{lm}\), the DOA \(\Omega_{\text{dir}}\) must be estimated; a robust DOA estimation method for spherical arrays is presented in [16].

The possible combinations of the pair \((l, m)\) form a set \(A\) with \((L + 1)^2\) elements, where \(L\) is the array order. The SDR can be estimated using (19) for all possible combinations of \((l, m)\) and \((l', m')\) (i.e., the set \(A^2\)) excluding identical pairs for which \((l, m) = (l', m')\); however we also exclude duplicate pairs \(((l', m'), (l, m))\) that provide the same information as \(((l, m), (l', m'))\) due to the symmetry of the coherence function. The reduced set thereby obtained is denoted as \(\widehat{L}\) and contains \((L + 1)^4 - (L + 1)^2\) \(2\) elements.

We then form an estimate of the SDR \(\hat{\Gamma}\) by taking a weighted average of the SDR estimates \(\hat{\Gamma}_{lm,l'm'}\), i.e.,
\[
\hat{\Gamma}(k, t) = \sum_{(l,m,l',m')\in\widehat{L}} \alpha_{lm,l'm'}(k) \hat{\Gamma}_{lm,l'm'}(k, t), \tag{21}
\]
where \(\alpha_{lm,l'm'}\) is a normalized weighting function. Ideally, the optimal weights \(\alpha_{lm,l'm'}\) depend on the variances of the SDR estimates. Since the variances are usually unknown, we propose to compute the weights as the geometric mean of the SNRs of the eigenbeams involved, i.e.,
\[
\alpha_{lm,l'm'}(k) = \frac{\sqrt{\text{SNR}_{lm}(k)\text{SNR}_{l'm'}(k)}}{\sqrt{\text{SNR}_{lm}(k)}\sqrt{\text{SNR}_{l'm'}(k)}}, \tag{22}
\]
where \(\text{SNR}_{lm}\) denotes the SNR at order \(l\) and degree \(m\) and is defined as
\[
\text{SNR}_{lm}(k) = \frac{\left| X_{lm}(k, t, \Omega_{\text{dir}}) \right|^2}{E \left| V_{lm}(k, t) \right|^2} = \left( P_k \right)^{-1} 4\pi Q P_{\text{dir}}(k, t) b_l(k) Y_{lm}^\ast(\Omega_{\text{dir}})^2. \tag{23a}
\]
The weighting function can then be simplified to
\[
\alpha_{lm,l'm'}(k) = \frac{\left| b_l(k) b_{l'}(k) Y_{lm}^\ast(\Omega_{\text{dir}}) Y_{l'm'}(\Omega_{\text{dir}}) \right|^2}{\sum_{(l,m,l',m')\in\widehat{L}} \left| b_l(k) b_{l'}(k) Y_{lm}^\ast(\Omega_{\text{dir}}) Y_{l'm'}^\ast(\Omega_{\text{dir}}) \right|^2}. \tag{24}
\]
Due to the chosen SNR definition, (24) depends only on the DOA and not on the wave or noise powers.

The weighted averaging of the SDR estimates, which is not performed in spatial domain coherence-based approaches with two microphones, aims to reduce the estimate variance, at the expense of increased computational complexity.

**IV. DIFFUSENESS ESTIMATION USING THE PSEUDointENSITY VECTOR**

We compare the proposed (coherence based) method with the previously proposed *coefficient of variation* (CV) method [17]. The CV method exploits the temporal variation of the intensity vector \(I\), and estimates the diffuseness as
\[
\Psi_{\text{CV}}(k, t) = \sqrt{1 - \frac{\| \left| E[I(k, t)] \right| \|}{\| E[I(k, t)] \|}} \tag{25}
\]
where \(\| \cdot \|\) denotes the \(\ell_2\) norm.

As previously shown in [16], the intensity vector can be estimated using a linear combination of first-order eigenbeams obtained with a spherical microphone array. The resulting vector, which is proportional to the intensity vector, is called a *pseudointensity vector*. The reader is referred to [16] for details of the computation of the pseudointensity vector from \(X_{00}, X_{1(-1)}, X_{10}\text{ and } X_{11}\). We hereafter refer to the estimation of the diffuseness using the CV method based on pseudointensity vectors as the *modified CV method*.

It should be noted that while the modified CV method only makes use of first-order eigenbeams, all \(Q\) microphone signals
are used to compute the pseudointensity vector, unlike in previous approaches where the intensity vector was estimated using either an acoustic vector sensor or four pressure microphones.

V. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed SHD coherence-based method, and compare it to the performance of the modified CV method.

A. Experimental Setup

We simulated the SHD signals received by a rigid spherical array of radius 4.2 cm up to an order $L$ (either $L = 1$ or $L = 3$). The directional source signal consisted of complex white Gaussian noise, with a DOA of $(0^\circ, 0^\circ)$ (elevation, azimuth). This DOA was assumed to be known for the estimation of the SDR in (19) and the weights in (24). The diffuse signal was generated by summing 200 plane waves with random phase and uniformly distributed DOAs; the diffuse signal power was set according to the desired SDR.

The noise signal consisted of additive complex white Gaussian noise; the noise power was set such that the desired SNR was obtained at the reference microphone $M_{ref}$, i.e.,

$$\text{SNR} = \frac{\mathbb{E} \left[ |X_{00}(k,t,\Omega_{dir})|^2 \right]}{\mathbb{E} \left[ |V_{00}(k,t)|^2 \right]}.$$

(26)

The noise power was therefore fixed for all values of SDR. We chose to compute the SNR at $M_{ref}$ because the directional signal power is different at each sensor, particularly for a rigid array. As noted in Sec. III-A, the noise power at $M_{ref}$ is reduced by a factor of $Q |b_0(k)|^2$ with respect to the sensors; the noise power at $M_{ref}$ is therefore lowest at low frequencies, where $b_0(k)$ is highest. With $Q = 32$ microphones, at low frequencies an SNR of 25 dB at $M_{ref}$ corresponds to an SNR of around 10 dB based on the noise power at the sensors.

Processing was performed in the STFT domain with a sampling frequency of 8 kHz, a window length of 16 ms and 50\% overlap between consecutive frames, giving a hop length of $\tau_{hop} = 8$ ms. The expectations in (8) and (25) were estimated using moving averages over a given number of time frames $N_{frames}$, which is related to the time averaging length $\tau_{avg}$ via the expression $\tau_{avg} = (N_{frames} + 1) \tau_{hop}$. The performance results shown were averaged over 15 s of data.

B. Results

In Fig. 1 we plot the mean diffuseness estimated by the proposed and modified CV methods as a function of SDR, as well as the ideal diffuseness as given by (7). In this experiment, the time averaging length was 88 ms, and the proposed method exploited eigenbeams up to order $L = 3$. We find that for high SDRs, the proposed method estimates the diffuseness more accurately, particularly at low frequencies. For low SDRs, the proposed method has a slightly higher bias than the modified CV method, due to the limited time averaging, as in [4]. In addition, as the SNR decreases from 35 dB to 25 dB, for both methods the bias at low frequencies and high SDRs increases, however for the proposed method this bias is in part due to the lack of compensation for the noise power, as in [4].

We also plot the standard deviation of the diffuseness estimates as a function of the SDR in Fig. 2. It can be seen that at high SDRs, the estimates obtained using the proposed method have significantly lower variance than those obtained using the modified CV method, due to the averaging of the coherence estimates over all eigenbeam pairs. The proposed method’s estimates also have lower variance at high frequencies and low SDRs.
In order to illustrate the effect of increasing the time averaging, in Fig. 3 we plot the mean diffuseness estimated by the two methods for two different averaging lengths (88 ms and 328 ms). As expected we see that the increase in time averaging significantly reduces the bias for the proposed method. With increased time averaging, the bias for the two methods is essentially the same at low SDRs, and is lower for the proposed method at high SDRs.

Finally in Fig. 4 we plot the standard deviation of the estimates obtained for array orders of $L = 1$ and $L = 3$. We find that by averaging over a larger number of SDR estimates, the variance of the final estimate is greatly reduced at low SDRs (except at low frequencies). We also note that even for $L = 1$, the variance of the proposed method’s estimates is lower than those obtained using the modified CV method, which also uses only zero and first-order eigenbeams.

VI. CONCLUSIONS

In this contribution, we proposed a diffuseness estimator based on the coherence between eigenbeams. We showed that at high SDRs, the proposed method has a lower bias than a previously proposed spatial domain method (the modified CV method), and that the underestimation of the diffuseness at low SDRs can be reduced by increased time averaging. Finally we found that increasing the array order significantly reduces the variance of the diffuseness estimates, and that even using a first-order array yields estimates with lower variance than those obtained with the modified CV method.

REFERENCES