DIRECTION-OF-ARRIVAL AND DIFFUSENESS ESTIMATION ABOVE SPATIAL ALIASING FOR SYMMETRICAL DIRECTIONAL MICROPHONE ARRAYS

Archontis Politis, Symeon Delikaris-Manias, Ville Pulkki

Department of Signal Processing and Acoustics
School of Electrical Engineering
Aalto University
02150 Espoo, Finland

ABSTRACT

A method for direction-of-arrival (DOA) and diffuseness estimation is presented, which proves to be effective above the spatial aliasing frequency of the microphone array in use. The method assumes symmetrical circular or spherical arrays of directional microphones or microphone mounted on a rigid baffle, and it exploits the inherent directionality of the array at high frequencies. The DOA and diffuseness estimators are shown to exhibit low estimation error above the spatial aliasing limit, compared to commonly used intensity-based estimators. A low-error broadband scheme that combines the intensity-based method and the new one is proposed for the ranges below and above aliasing respectively.

***Index Terms*—** DOA, diffuseness, direct-to-diffuse ratio, microphone arrays, spatial aliasing

1. INTRODUCTION

In parametric time-frequency processing of array recordings, many practical scenarios can be adequately modeled as the contribution of a single source, the direct sound, and reverberation. A further practical assumption considers the reverberant sound as perfectly diffuse and isotropic [1]. In this case, the model parameters that define the sound field properties are the direction-of-arrival (DOA) of the direct sound and the power of the direct and diffuse components. Equivalently, a power relation can be estimated such as the direct-to-diffuse ratio (DDR), or the diffuseness [1, 2] expressing the power ratio of the diffuse part over the total.

Estimation of DOA and diffuseness, or DDR, permits the application of beam-forming and optimal filtering techniques to the array signals with applications such as signal enhancement [3, 4], parametric spatial sound recording and reproduction [2, 5] and dereverberation [6]. Commonly used estimators for both DOA [7, 8, 9] and diffuseness [1, 2] are based on statistics of the sound intensity in a direct-plus diffuse sound field model. We call these estimators intensity-based (IB) ones. Spatial aliasing affects the performance of IB since the required pressure gradient signals are aliased above the limiting frequency of the array. This effect can be detrimental, if it occurs in the frequency range of interest. In applications where perceived spatial quality and coloration are critical, such as spatial sound reproduction, erroneous estimation results in loss of perceived quality [5].

Recently, a DDR estimator based on a least-squares solution on short-time power estimates of directional microphones was proposed in [10], which would be unaffected by spatial aliasing. That estimator requires separately a DOA estimate, which at high frequencies may be unreliable due to aliasing. DOA estimation under aliasing conditions have been studied to some extent, e.g. in [11] and references therein. Most of these approaches involve a search over the parameter space and consider broadband localization. Hence, computational efficiency and frequency resolution which are required for applications such as the aforementioned spatial audio recording and reproduction, are not necessarily met.

We propose a simple novel estimator of DOA and diffuseness that operates above the spatial aliasing limit. Similar DOA estimation was applied previously by the authors for a 3D tetrahedral array [5], without analysis of its properties. Certain conditions regarding the array properties should be met for unbiased estimation, which are detailed below. Furthermore, a new diffuseness estimator based on directional statistics is proposed. Their performance is evaluated with simulations and measurements of a spherical 3D array.

2. SIGNAL MODEL

The estimation is performed in the time-frequency domain, where k denotes discrete frequency index and t time-frame index. In this work a short-time Fourier transform (STFT) is utilized. Spherical coordinates are defined as $\mathbf{r} = (r, \Omega)$, with azimuth-elevation $\Omega = (\theta, \phi)$ or azimuth $\Omega = \theta$ for the 3D and 2D case respectively. Integration $\int_\Omega \mathrm{d}\Omega$ refers to integration over the unit sphere $\int_{-\pi}^{\pi}\int_0^{\pi} \sin\phi \,\mathrm{d}\phi\,\mathrm{d}\theta$ or over the circle $\int_{-\pi}^{\pi} \mathrm{d}\theta$ for the 3D and 2D case. A unit vector oriented at Ω is written as $\mathbf{n}(\Omega)$. The sound-field is modeled as a superposition of a source signal s_{dir}, and a diffuse signal s_{diff}

corresponding to diffuse sound such as late reverberation.

The following assumptions are made for the sound field. The direct and diffuse sound signals are uncorrelated: $\mathbf{E}\left[s_{\mathrm{dir}}(k,t)s_{\mathrm{diff}}^{*}(k,t,\Omega)\right] = 0$. Furthermore, the diffuse sound is uncorrelated for different directions and has equal power, P_{diff}, for all directions The theoretical DDR is then $\Gamma(k) = P_{\mathrm{dir}}(k)/(AP_{\mathrm{diff}}(k))$, with $A = \{2\pi, 4\pi\}$ for 2D or 3D fields respectively. The diffuseness is related to the DDR by

$$\psi(k) = 1/(1+\Gamma(k)). \tag{1}$$

Diffuseness is bounded between $\psi \in [0,1]$, with zero value indicating a single plane wave and unity a purely diffuse field.

Considering an array of Q microphones at positions $\mathbf{r}_q,\ q = 1,2,...,Q$, the signal vector captured by the microphone array is given by

$$\begin{aligned}\mathbf{x}(k,t) &= \mathbf{h}(k,\Omega_{\mathrm{dir}})s_{\mathrm{dir}}(k,t) \\ &+ \int_{\Omega} \mathbf{h}(k,\Omega)s_{\mathrm{diff}}(k,t,\Omega)\,\mathrm{d}\Omega,\end{aligned} \tag{2}$$

where the vector $\mathbf{h}(k,\Omega) = [h_1(k,\Omega),...,h_Q(k,\Omega)]^T$ consists of the steering vectors of the array to direction Ω and Ωdir the direction of the direct component.

3. DOA ESTIMATION BASED ON MAGNITUDE SENSOR RESPONSE

The proposed estimators are based on the following assumptions concerning the microphone array:

- The array consists of microphones that provide some directionality.
- The array consists of microphones with similar axisymmetric patterns, so that $|h_q(\Omega)| = |h(\alpha_q)|$ where $\alpha_q = \arccos(\mathbf{n}^T(\Omega)\mathbf{n}(\Omega_q))$ is the angle between the DOA and the orientation of the microphone Ω_q.
- The microphones in the array are at the same radius $||\mathbf{r}_q|| = R$ and are symmetrically distributed. The symmetry can be expressed as $\mathbf{N}_Q^T\mathbf{N}_Q = (\kappa/Q)\mathbf{I}_\kappa$, where $\mathbf{N}_Q = [\mathbf{n}(\Omega_1),...,\mathbf{n}(\Omega_Q)]^T$ is the matrix with the unit vectors pointing to the microphone positions and $\kappa = \{2,3\}$ for the 2 and 3-dimensional case respectively.

These conditions hold in practice for many practical cases, such as uniform circular arrays of directional microphones and symmetric arrangements of omnidirectional microphones mounted on rigid cylindrical or spherical baffles.

3.1. Single plane-wave case

First we consider the case of a single plane wave incident from Ω_{dir}. Assuming that we have a continuous array of directional elements and radius R, the pressure captured on its circumference or surface can be expressed by

$$p(k,t,R,\Omega,\Omega_{\mathrm{dir}}) = s_{\mathrm{dir}}(k,t)h(k,R,\alpha), \tag{3}$$

where $\alpha = \arccos(\mathbf{n}^T(\Omega_{\mathrm{dir}})\mathbf{n}(\Omega))$ is the angle between the DOA of the plane wave and the measurement point. The complex directional response $h(k,R,\alpha)$ includes phase differences at each array point. Examples of h for some common cases are

$$h(k,R,\alpha) = \begin{cases} e^{i(\omega/c)R\cos\alpha}[\beta + (1-\beta)\cos\alpha], & \text{(a)} \\ B_0(\omega R/c) + 2\sum_{n=1}^{\infty} i^n B_n(\omega R/c)\cos n\alpha, & \text{(b)} \\ \sum_{n=0}^{\infty} i^n(2n+1)b_n(\omega R/c)P_n(\cos\alpha), & \text{(c)} \end{cases} \tag{4}$$

where (a) are open arrays of common first-order directional microphones with $\beta \in (0,1)$, both for circular and spherical ones, (b) omnidirectional microphones on a rigid cylinder, and (c) omnidirectional microphones on a rigid sphere. The quantities B_n, b_n are radial weights that depend on the radius of the array [12]. The pressure responses of (4) can be decomposed in a magnitude directional distribution and a phase distribution. The magnitude part is known to be symmetric around the DOA and with its maximum at it. For the case (a) of first-order microphones it is evident that it reduces to the magnitude of the first-order directional pattern $|h(k,R,\alpha)| = |\beta + (1-\beta)\cos\alpha|$ and is independent of frequency. Based on this property, a DOA estimator is proposed by taking the integral of the unit vector across all directions, weighted with the magnitude distribution as in

$$\begin{aligned}\mathbf{r}_{\mathrm{DOA}}(k,t,\Omega_{dir}) &= \int_{\Omega} |p(k,t,R,\Omega,\Omega_{\mathrm{dir}})|\mathbf{n}(\Omega)\,\mathrm{d}\Omega \\ &= |s_{\mathrm{dir}}(k,t)| \int_{\Omega} |h(k,R,\alpha)|\mathbf{n}(\Omega)\,\mathrm{d}\Omega.\end{aligned} \tag{5}$$

It is evident that in the case that the array does not exhibit any directionality, such as an open array of omnidirectional microphones with constant $|h(k,R,\alpha)|$, the integral vanishes and the estimator is zero. Otherwise, the unit vector in the integral is weighted with a symmetric pattern and the resulting vector points to the maximum of that pattern which is the DOA of the plane wave.

3.2. Discretization

In practice and for a real discrete array the DOA vector can be estimated by approximating the integration of (5) with the discrete summation of

$$\hat{\mathbf{r}}_{\mathrm{DOA}}(k,t) = \sum_{q=1}^{Q} |x_q(k,t)|\mathbf{n}(\Omega_q) \tag{6}$$

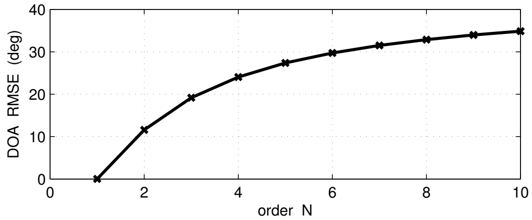

Fig. 1. Mean directional error for tetrahedral arrangement and cardioid sensors of spherical order ranging from $N = 1 : 10$.

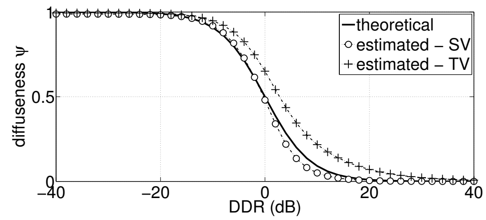

Fig. 2. Theoretical diffuseness curve and estimated one for a tetrahedral arrangement and ideal cardioid sensors.

for an array of Q uniformly arranged microphones where Ω_q the direction of each microphone. The product of the magnitude of the microphone signals x_q and the unit vectors $\mathbf{n}(\Omega_q)$ pointing at the microphones approximate the integral with zero error subject to some conditions. Essentially the integration error is determined by the spherical order of the magnitude directivity of the array $|h(k, R, \alpha)|$ and the number of microphones. The directivity functions form spherical or circular polynomials of certain order which can be integrated exactly by a sum on a finite uniformly arranged set of points. The estimator integrates the product of an N-th order pattern with the components of the unit vector, which themselves constitute first-order patterns (dipoles). Hence, the combined order of the product is $N + 1$. Examples of the minimum number of microphones and their arrangements that meet this requirement can be found tabulated in [13] and are termed t-designs, where $t = N+1$ refers to the order of the polynomial to be integrated.

An example is given for a minimal 3D array using four cardioid microphones at the vertices of a tetrahedron, known in spatial sound recording literature as the A-format microphone [14]. A tetrahedral arrangement approximates exactly integration of patterns up to second order [13], hence the estimator should be unbiased only for cardioid microphones of $N = 1$. This effect is demonstrated in Fig.1, where the order of the cardioid is switched from $N = 1$ to $N = 10$ by following the relation $d_{\mathrm{card}}(\alpha) = (1/2)^N(1 + \cos\alpha)^N$. The root-mean squared error (RMSE) is computed between the true DOA and the estimated one, averaged over 162 DOAs uniformly distributed around the array. It is noted that the effect of the microphones to the estimation resembles the problem of spatial aliasing, but is fundamentally different. DOA estimators based on some kind of beamforming, including the IB ones, will become erratic above the spatial aliasing limit independently of the responses of the microphones. The proposed estimator does not depend on inter-microphone phase relations and is not affected by aliasing.

4. DIFFUSENESS AND DDR ESTIMATION

Estimation of the diffuseness and the DDR, in the case of IB estimators, is done through the normalized intensity-energy density ratio [1, 2], which suffer at high frequencies from spatial aliasing. Based on the DOA vector of (6), a potential alternative could be the estimator of [15] based on the temporal variation of intensity vectors, replaced by the DOA vectors of (6). However, that estimator considers the statistics of intensity vectors including their magnitude, and the DOA vectors of (6) are in general not proportional to intensity.

We propose a new diffuseness and DDR estimator, based on the spherical variance of the DOAs of (6) in the presence of diffuse sound, which does not consider the magnitude of the DOA vectors. More specifically, the mean vector of the observed DOAs is

$$\hat{\boldsymbol{\rho}}(k) = \mathbf{E}\left[\frac{\hat{\mathbf{r}}_{\mathrm{DOA}}(k,t)}{||\hat{\mathbf{r}}_{\mathrm{DOA}}(k,t)||}\right]. \tag{7}$$

An estimate of the plane-wave DOA in the presence of diffuse sound is then given by the direction of the mean vector

$$\mathbf{n}(k, \hat{\Omega}_{\mathrm{dir}}) = \frac{\hat{\boldsymbol{\rho}}(k)}{||\hat{\boldsymbol{\rho}}(k)||}. \tag{8}$$

Diffuseness can be computed from a normalized measure of the spherical spread of the DOA estimates, which should be perfectly concentrated for a plane wave, and uniformly dispersed for a purely diffuse field. This measure corresponds to the spherical variance [16], which is bounded between zero and one for these two cases and is given by

$$\hat{\psi}(k) = 1 - ||\hat{\boldsymbol{\rho}}(k)||. \tag{9}$$

From this definition of diffuseness, the DDR can be readily computed as

$$\hat{\Gamma}(k) = \frac{1}{\hat{\psi}} - 1 = \frac{||\hat{\rho}(k)||}{1 - ||\hat{\rho}(k)||}. \tag{10}$$

The expectation operator is approximated by a finite average over a number of time frames or a recursive scheme.

The fact that the estimator follows the theoretical diffuseness is shown in Fig. 2. A diffuse field is simulated as the sum of 162 complex exponentials with uniformly distributed random phases and with uniform DOAs, derived from subsequent divisions of the edges of an icosahedron. The direct

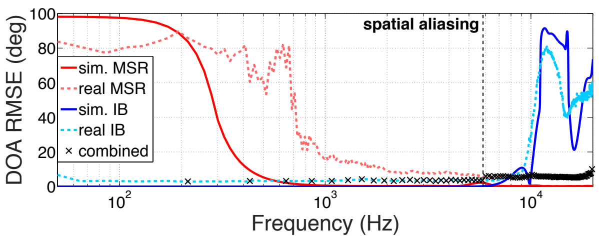

Fig. 3. Directional error for a measured (solid line) and simulated (dashed line) Eigenmike, with the proposed magnitude sensor response method (MSR) (red) and the IB method (blue). The combined estimator is marked on top (black).

component is modeled as another complex random exponential incident from the north pole. The relative power of the two components is adjusted to result in a specific DDR, for which the theoretical diffuseness is computed from (1). For the same tetrahedral first-order array, the components are encoded to the array signals and the spherical variance (SV) estimator of (9) is applied for 100 realizations. The same is repeated for the temporal variation (TV) estimator of [15]. The proposed SV estimator approximates closely the theoretical curve, with minor deviation due to finite averaging effects, while the TV estimator exhibits a clear bias.

5. MEASUREMENTS AND EVALUATION

Real-world sensors will exhibit a frequency-dependent directionality, with a response approaching omnidirectional at low frequencies, and usually becoming more directional at high frequencies. This effect is inherent in array designs based on a scattering body, such as microphones mounted on a rigid sphere or cylinder [12]. To test the performance of the method under realistic conditions, DOA estimation is performed for a dense grid of measurements of a 3D spherical array of 32 microphones mounted on a rigid sphere of radius $R = 4.2\,\mathrm{cm}$ (Eigenmike[1]), along with its simulated ideal version. The array geometry is that of a truncated icosahedron and is near-symmetrical. Apart from the proposed estimator of (6), the IB estimator is also applied, as proposed in [2]. The aliasing frequency limit is determined with the approximate formula $f_{al} = c/(2R\gamma)$, where γ is the angle between two microphones. Fig. 3 depicts the root mean-square error (RMSE) of the DOA and it is evident that the two estimators are complementary; the proposed has very low error above aliasing and performs worse at lower frequencies due to noise and decreasing directionality, while the IB estimator performs well at this range and breaks above aliasing. Hence, broadband low-error estimation can be achieved through a combination of the two methods at different ranges, as is shown in Fig. 3 with the combined line. An RMSE below 8^o for the whole audible bandwidth is achieved with the combined approach.

[1] http://www.mhacoustics.com/products#eigenmike1

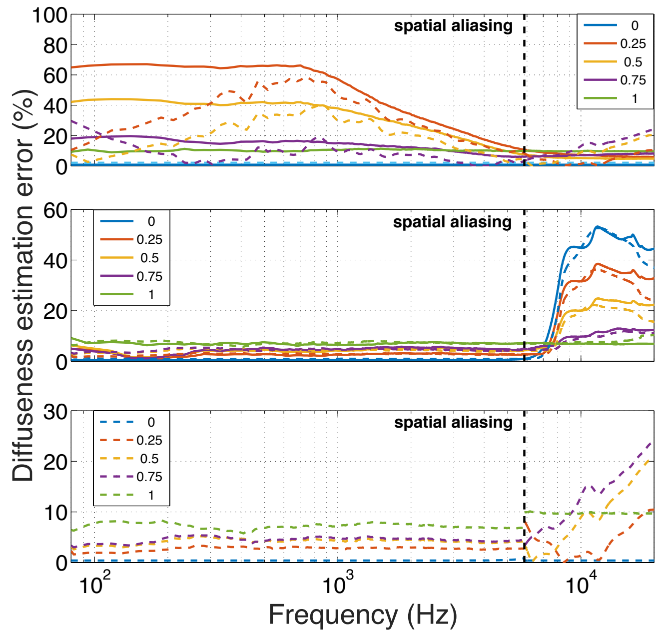

Fig. 4. Diffuseness error (%) for proposed method (top), the IB method (middle) and the combined estimator (bottom) for five true diffuseness values $\psi = 0 : 0.25 : 1$. Results are shown both for an ideal simulated Eigenmike (continuous lines) and based on a measured one (dashed lines). Note the different error scale in each plot.

The diffuseness estimation error, following a similar process as in Fig. 2, is show in Fig. 4 as percentage error (%) for different target diffuseness values. A similar trend as the DOA RMSE is evident. At low frequencies, low directionality makes the proposed estimator non-robust to noise and the slight asymmetry of the array, while above aliasing the error drops towards zero. The resulting error in the combined estimator is on average below 10%.

6. CONCLUSIONS

This work presents an estimation approach for DOA, diffuseness and DDR, that exploits the directionality of the magnitude response of a symmetrical array in a single-source plus diffuse-sound model. DOA estimation performs well above the spatial aliasing limit, subject to the properties of the microphones and some geometrical considerations for their arrangement. A novel diffuseness and DDR estimator is proposed as well, based on the spherical variance of the predicted DOA of the direct-sound, in the presence of diffuse sound. It is shown through simulations and measurements that both estimators complement effectively the commonly used intensity-based ones, and their combination offers broadband low-error estimation at all frequencies.

7. REFERENCES

[1] Giovanni Del Galdo, Maja Taseska, Oliver Thiergart, Jukka Ahonen, and Ville Pulkki, "The diffuse sound field in energetic analysis," *The Journal of the Acoustical Society of America*, vol. 131, no. 3, pp. 2141–2151, 2012.

[2] Ville Pulkki, "Spatial sound reproduction with directional audio coding," *Journal of the Audio Engineering Society*, vol. 55, no. 6, pp. 503–516, 2007.

[3] Iain McCowan and Hervé Bourlard, "Microphone array post-filter based on noise field coherence," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 709–716, 2003.

[4] Symeon Delikaris-Manias and Ville Pulkki, "Cross pattern coherence algorithm for spatial filtering applications utilizing microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 11, pp. 2356 – 2367, 2013.

[5] Archontis Politis and Ville Pulkki, "Broadband analysis and synthesis for directional audio coding using a-format input signals," in *131st International Convention of the Audio Engineering Society*, New York, NY, USA, 2011.

[6] Marco Jeub, Magnus Schäfer, Thomas Esch, and Peter Vary, "Model-based dereverberation preserving binaural cues," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1732–1745, 2010.

[7] Oliver Thiergart, Richard Schultz-Amling, Giovanni Del Galdo, Dirk Mahne, and Fabian Kuech, "Localization of sound sources in reverberant environments based on directional audio coding parameters," in *127th International Convention of the Audio Engineering Society*, New York, NY, USA, 2009.

[8] Sakari Tervo, "Direction estimation based on sound intensity vectors," in *17th European Signal Processing Conference (EUSIPCO 2009)*, Glasgow, UK, 2009, pp. 700–704.

[9] Dovid Levin, Emanuël AP Habets, and Sharon Gannot, "On the angular error of intensity vector based direction of arrival estimation in reverberant sound fields," *The Journal of the Acoustical Society of America*, vol. 128, no. 4, pp. 1800–1811, 2010.

[10] Oliver Thiergart, Tobias Ascherl, and Emanuël AP Habets, "Power-based signal-to-diffuse ratio estimation using noisy directional microphones," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 7440–7444.

[11] Vinod Veera Reddy, Andy WH Khong, and Boon Poh Ng, "Unambiguous speech doa estimation under spatial aliasing conditions," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, no. 12, pp. 2133–2145, 2014.

[12] Heinz Teutsch, *Modal array signal processing: principles and applications of acoustic wavefield decomposition*, Springer, 2007.

[13] Ronald H Hardin and Neil JA Sloane, "Mclaren's improved snub cube and other new spherical designs in three dimensions," *Discrete & Computational Geometry*, vol. 15, no. 4, pp. 429–441, 1996.

[14] Michael A Gerzon, "The design of precisely coincident microphone arrays for stereo and surround sound," in *50th International Convention of the Audio Engineering Society*, London, UK, 1975.

[15] Jukka Ahonen and Ville Pulkki, "Diffuseness estimation using temporal variation of intensity vectors," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2009 (WASPAA'09)*, New Paltz, NY, USA, 2009, pp. 285–288.

[16] Kanti V. Mardia and Peter E. Jupp, *Directional statistics*, John Wiley & Sons, 2009.