Audio Engineering Society

Convention Paper 10695

Presented at the AES 156th Convention
2024 June 15–17, Madrid, Spain

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Covariance in non-coincident tetrahedral arrays

Alexis Favrot[1], Franz Zotter[2], and Christof Faller[1]

[1]*Illusonic GmbH, Stationsstrasse 20, 8606 Greifensee, Switzerland*
[2]*Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria*

Correspondence should be addressed to Alexis Favrot (alexis.favrot@illusonic.com)

ABSTRACT

The four signals of the non-coincident directional capsules, composing an Ambisonics microphone are combined in order to render resulting decoded signals with desired directional and diffuse responses. While the directional responses are derived from first order microphone patterns, the responses to diffuse sound are described using the cross-correlation coefficient, as a function of frequency, between the various microphone signal pairs.

The cross-correlation coefficient between the two signals of a microphone pair, in a diffuse sound field, is an important parameter when using multiple microphones systems, trying to control, attenuate or enhance the diffuse sound. Its properties, for first order and Nth-order differential microphone signals, have been investigated in detail. However when considering multiple non-coincident directional microphones, the resulting cross-correlation coefficient of two decoded signals will not follow the same properties. The corresponding cross-correlation coefficient is quantitively derived and its characteristics are further discussed in terms of application in Ambisonics signal manipulation.

1 Introduction

Virtual reality (VR) audio has revived interest in Ambisonics microphones [1, 2, 3, 4]. The A-format signal of the four non-coincident directional microphone capsules, mounted on a tetrahedral array, are combined to either be converted into a B-format signal, or decoded into multichannel audio for loudspeaker reproduction [5, 6, 7, 8, 9, 10], or binaural audio for headphones reproduction [11, 12, 13], according to the look direction of the moving picture of a VR camera.

Whereas the directional responses of the converted or decoded signals are straightforwardly specified as combinations of first order microphone patterns, the responses to diffuse sound have been defined based on the cross-correlation coefficients between the various microphone pairs, separately, namely without taking all microphones at once in consideration [14, 15].

In a closed acoustic environment, contrary to the direct sound and early reflections, the later and numerous reflections that have themselves been reflected create a diffuse sound field, making late reverberation and diffuse sound to be identified as practically the same thing. Diffuse sound can then be mathematically understood as independent sounds of equal strength arriving at the listening point from all directions. Therefore, in order to identify the properties of the diffuse sound field, it has been proposed to model it by a spherically isotropic noise field. Originally, in [16], Cook showed that the cross-correlation, between two signals recorded by two omnidirectional microphones in such a spherically isotropic noise field, corresponds to a sine

cardinal function varying with respect to both the frequency of the signals and the spacing between the two microphones, thereby providing an estimation of how the diffuse sound field is captured and can be characterized in a room. Generalizations, for various pairs of first-order and Nth-order differential microphones, in different positions, have been later further developed in [17]: deriving a generic cross-correlation coefficient, referred to as the spatial coherence function.

Also, for proper balanced conversion, decoding, or rendering of Ambisonics, the knowledge about the amount of diffuse sound is a crucial aspect in the way the acoustic environment is perceived [18, 19]. The spatial coherence function between two microphones in the tetrahedral array follows the properties of the spatial coherence derived in [17], and the diffuse sound field characteristics can thus be depicted based on pairwise spatial coherence functions [14], or analogously using a diffuseness index [6, 15].

However, all the previous contributions did not investigate the characteristics of the spatial coherence function when derived on a pair of decoded signals, namely a pair of virtual decoded microphone signals generated as a combination of all involved first order, non-coincident microphone signals composing the original A-format . In this specific case, the spatial coherence function appears to differ from the one of a similar real microphone pair, making the rendering of diffuse sound inaccurate without further proper considerations.

In this contribution, after recalling the definition and derivation of the spatial coherence function in Section 2, the specific spatial coherence function of first order Ambisonics (FOA) decoded signals is derived quantitatively in Section 3, and Section 4 provides measurements and results with explanations on how the deviations can be used in order to control the amount of diffuse sound in the rendering of Ambisonics signals.

2 Spatial Coherence Function

The diffuse sound field is modelled by a spherically isotropic noise field, namely numerous independent plane waves, each having a complex amplitude coefficient $a(\mathbf{n})$, coming from a specific direction defined by its unit vector $\mathbf{n}$, as illustrated in Fig. 1:

$$\mathbf{n}^\top = [\cos\theta\cos\phi, \sin\theta\cos\phi, \sin\phi]\,. \tag{1}$$

Given the properties of the spherically isotropic diffuse field, all amplitudes are defined to have constant variance per direction and the amplitudes of two plane waves from different directions, $\mathbf{n}$ and $\mathbf{l}$, are mutually uncorrelated

$$E\{a(\mathbf{n})a^\star(\mathbf{l})\} = \tfrac{1}{4\pi}\delta(\mathbf{n}-\mathbf{l}). \tag{2}$$

where δ stands for the Dirac delta function, * denotes complex conjugate, and E$\{.\}$ is the statistical expectation. The factor $1/4\pi$ is a normalization constant in order to provide a unitary variance when integrating the expectation (2) for all directions over the unit sphere, i.e. $\mathbf{n},\mathbf{l} \in S^2$,

$$\iint_{S^2} E\{a(\mathbf{n})a^\star(\mathbf{l})\}\,\mathrm{d}\mathbf{n}\,\mathrm{d}\mathbf{l} = \int_{S^2} \tfrac{1}{4\pi}\,\mathrm{d}\mathbf{n} = 1. \tag{3}$$

Neither a shift of the origin, which corresponds to a direction-dependent phase shift, nor a rotation modifies the distribution of the plane-wave amplitudes in expected strength and correlation, hence second-order statistics in a spherically isotropic diffuse field are shift and rotation invariant.

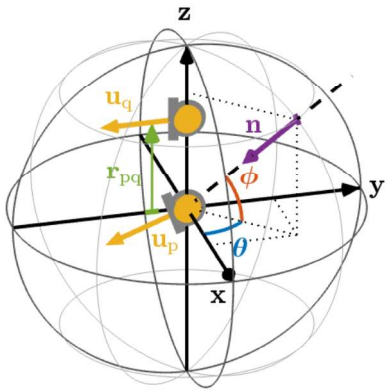


Fig. 1: Example of a plane wave arriving from a direction $\mathbf{n}$ and captured by two first-order microphones, p and q, distributed on the z-axis with a spacing r_{pq}.

2.1 Spatial coherence function of a First-Order Microphone Pair

By definition, a first-order microphone p has a directional response given as

$$D_{v_p}(\psi_p) = v_p + (1-v_p)\cos\psi_p\,, \tag{4}$$

where v_p is the weighting given to its omnidirectional component, while $1-v_p$ the weighting given to its dipole component. ψ_p stands for the angle between the incident wave vector $\mathbf{n}$ and the axis of the microphone, defined by the unit vector $\mathbf{u}_p$ as the look direction of the microphone:

$$\mathbf{u}_p^\top = [\cos\theta_p \cos\phi_p, \sin\theta_p \cos\phi_p, \sin\phi_p]\,. \quad (5)$$

Thus (4) can be re-written as

$$D_{v_p}(\mathbf{n}) = v_p + (1-v_p)\mathbf{u}_p^\top \mathbf{n}\,, \quad (6)$$

or differently scaled:

$$D_{w_p}(\mathbf{n}) = 1 + w_p \mathbf{u}_p^\top \mathbf{n}\,, \quad (7)$$

with $w_p = (1-v_p)/v_p$. Table 1 shows commonly used values for v_p, respectively w_p, and corresponding microphone type names.

Table 1: Parameter v_p, respectively w_p, for commonly used first order microphones.

v_p	w_p	Name
0	∞	Dipole (figure of eight)
1/4	3	Hyper-cardioid
1/3	2	Super-cardioid
0.5	1	Cardioid
> 0.5	< 1	Sub-cardioid
1	0	Omnidirectional

Considering now a pair of microphones, p and q, such that, for simplification and without loss of generality: one microphone is placed at the origin and aims towards $\mathbf{u}_p$, and the other one is shifted by r_{pq} along the Z-axis and aims towards $\mathbf{u}_q$, as shown in Fig. 1. The captured microphone signals are composed of diffuse-field components from all directions and can be derived as

$$s_p = \int_{S^2} a(\mathbf{n}) D_{w_p}(\mathbf{n})\,\mathrm{d}\mathbf{n} = \int_{S^2} a(\mathbf{n})(1 + w_p \mathbf{u}_p^\top \mathbf{n})\,\mathrm{d}\mathbf{n}, \quad (8)$$

and

$$\begin{aligned} s_q &= \int_{S^2} a(\mathbf{n}) D_{w_q}(\mathbf{n}) e^{ikr_{pq}\sin\phi}\,\mathrm{d}\mathbf{n} \\ &= \int_{S^2} a(\mathbf{n})(1 + w_q \mathbf{u}_q^\top \mathbf{n}) e^{ikr_{pq}\sin\phi}\,\mathrm{d}\mathbf{n}, \end{aligned} \quad (9)$$

where $k = \frac{2\pi}{\lambda} = \frac{\omega}{c}$ is the wave number, defined from the wave length λ or from the angular frequency ω, and the celerity c, such that the exponential represents the phase difference from one microphone to the other.

The expected cross-correlation between both microphone signals, as defined in (8) and (9),

$$\Gamma_{pq} = E\{s_p s_q^\star\}\,, \quad (10)$$

contains two independent integrals that collapse to a single one as in (3) by (2):

$$\Gamma_{pq} = \tfrac{1}{4\pi} \int_{S^2} (1 + w_p \mathbf{u}_p^\top \mathbf{n})(1 + w_q \mathbf{u}_q^\top \mathbf{n}) e^{ikr_{pq}\sin\phi}\,\mathrm{d}\mathbf{n}\,. \quad (11)$$

Replacing the integration element such that

$$\mathrm{d}\mathbf{n} = \mathrm{d}\theta\,\mathrm{d}z\,, \quad (12)$$

for integration on the surface S^2 with $0 \le \theta \le 2\pi$ and $-1 \le z \le 1$, and re-writing the vector $\mathbf{n}$ in (1):

$$\mathbf{n}^\top = [\cos\theta\sqrt{1-z^2}, \sin\theta\sqrt{1-z^2}, z]\,, \quad (13)$$

as well as the microphone specific direction vector $\mathbf{u}_p$ in (5):

$$\mathbf{u}_p^\top = [x_p, y_p, z_p]\,. \quad (14)$$

Thus, in (11), integration over θ yields a factor zero for the dependency on $\cos\theta$, $\sin\theta$, the factor π for the dependency on $\cos^2\theta$, $\sin^2\theta$, and 2π for θ-independent parts, leaving

$$\begin{aligned} \Gamma_{pq} = \tfrac{1}{4\pi} \int_{-1}^{1} \{ & 2\pi + 2\pi(w_p z_p + w_q z_q) z \\ & + w_p w_q [\pi(\mathbf{u}_p^\top \mathbf{u}_q - z_p z_q)(1 - z^2) \\ & + 2\pi z_p z_q z^2]\} e^{ikr_{pq}z}\,\mathrm{d}z. \end{aligned} \quad (15)$$

The integrand expanded into Legendre polynomials $P_0 = 1$, $P_1 = z$, and $P_2 = \frac{1}{2}(3z^2 - 1)$ becomes

$$\begin{aligned} \Gamma_{pq} = \tfrac{1}{4} \int_{-1}^{1} \Big[& \frac{6 + 2w_p w_q \mathbf{u}_p^\top \mathbf{u}_q}{3} P_0 + 2(w_p z_p + w_q z_q) P_1 \\ & + \frac{2w_p w_q (3 z_p z_q - \mathbf{u}_p^\top \mathbf{u}_q)}{3} P_2 \Big] e^{ikr_{pq}z}\,\mathrm{d}z\,, \end{aligned} \quad (16)$$

and allows to use the integral representation of the spherical Bessel functions, defined as $i^n j_n(\rho) = \frac{1}{2}\int_{-1}^{1} e^{i\rho z} P_n(z)\mathrm{d}z$, for an order n [20], such that we get

$$\begin{aligned} \Gamma_{pq} = & \frac{3 + w_p w_q \mathbf{u}_p^\top \mathbf{u}_q}{3} j_0(kr_{pq}) \\ & + i(w_p z_p + w_q z_q) j_1(kr_{pq}) \\ & - \frac{w_p w_q (3 z_p z_q - \mathbf{u}_p^\top \mathbf{u}_q)}{3} j_2(kr_{pq})\,. \end{aligned} \quad (17)$$

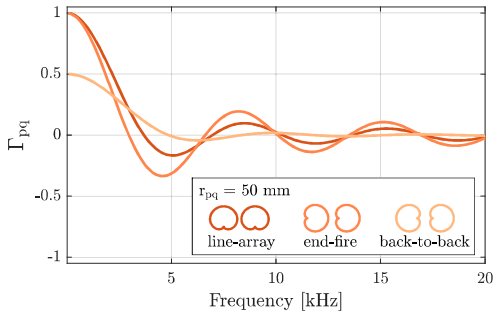


Fig. 2: Spatial coherence of two cardioids ($w_p = w_q = 0.5$) separated with a spacing $r_{pq} = 50$ mm in various orientations, respectively in line-array, end-fire and back-to-back configurations.

The function Γ_{pq} enables to characterise generically the behaviour of a pair of two first-order microphones signals in a spherically isotropic diffuse sound field. For instance, Fig. 2 shows the resulting spatial coherence function considering two cardioids ($w_p = w_q = 0.5$) separated with a spacing $r_{pq} = 50$ mm for various orientations, namely line-array, end-fire and back-to-back cardioids, providing quantitative measures of what such microphone pairs can achieve in terms of diffuse sound capturing as a function of frequency.

In the following, general properties and characteristics of the spatial coherence function Γ_{pq} are described as well as illustrations of specific conspicuous cases.

2.2 Generalization and Noticeable Cases

The function Γ_{pq} is shift invariant in a diffuse sound field since the phase term, as expressed in (11), only depends on the relative distance between both microphones $r_{pq} = ||\mathbf{r}_{pq}||$. It is also rotation invariant, as rotation leaves the assumptions for a spherically isotropic diffuse field (2) and (3) unaffected. Considering the weights, w_p and w_q, and the aiming vectors, $\mathbf{u}_p$ and $\mathbf{u}_q$, of the first-order component for each microphone, (17) can be re-written for any angle and distance between the two microphones:

$$\Gamma_{pq} = \frac{3 + w_p w_q \mathbf{u}_p^\top \mathbf{u}_q}{3} j_0(kr_{pq}) + i(w_p \mathbf{u}_p^\top + w_q \mathbf{u}_q^\top)\mathbf{r}_{pq} j_1(kr_{pq}) - \frac{w_p w_q [3(\mathbf{u}_p^\top \mathbf{r}_{pq})(\mathbf{u}_q^\top \mathbf{r}_{pq}) - \mathbf{u}_p^\top \mathbf{u}_q]}{3} j_2(kr_{pq}). \quad (18)$$

As in Cook et al [16], sound pressure microphones, i.e. with omnidirectional characteristics ($w_p = w_q = 0$), simplify integration of (15) to be

$$\Gamma_{\mathrm{pq}_{|w=0}} = j_0(kr_{pq}) = \frac{\sin(kr_{pq})}{kr_{pq}}. \quad (19)$$

Fig. 3 shows the spatial coherence function of two omnidirectional microphone signals for different spacing r_{pq}.

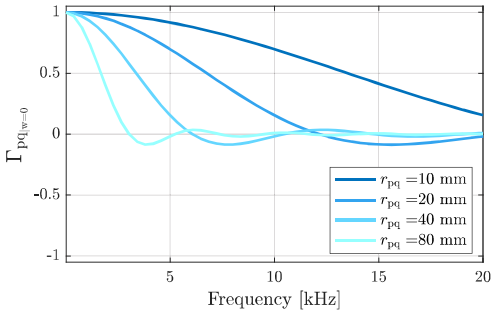


Fig. 3: Spatial coherence functions of two omnidirectional microphones ($w_p = w_q = 0$) for different spacing r in a spherically isotropic diffuse sound field.

As in Elko's work [17], coincident microphones, i.e. $r_{pq} = 0$, make the integrand in (15) a frequency-independent polynomial; or equivalently $j_n(0) = \delta[n]$ makes (18) frequency-independent so that it depends on the types of microphone, w_p and w_q, as well as on their aiming directions, $\mathbf{u}_p$ and $\mathbf{u}_q$, only:

$$\Gamma_{\mathrm{pq}_{|r=0}} = 1 + \tfrac{1}{3} w_p w_q \mathbf{u}_p^\top \mathbf{u}_q. \quad (20)$$

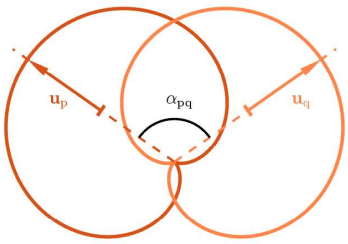


Fig. 4: Example of a pair of coincident ($\mathbf{r}_{pq} = 0$) cardioids ($w_p = w_q = 0.5$) microphones separated by an angle α_{pq}.

Fig. 5 reproduces the spatial coherence functions $\Gamma_{\mathrm{pq}_{|r=0}}$ of coincident microphone pairs of same types, i.e. $w_p = w_q$, as provided and described in [21], with respect to the stereo angle α_{pq} defined such that $\mathbf{u}_p^\top \mathbf{u}_q = \cos\alpha_{pq}$ as illustrated in Fig. 4 for a pair of cardioids ($w_p = w_q = 0.5$) .

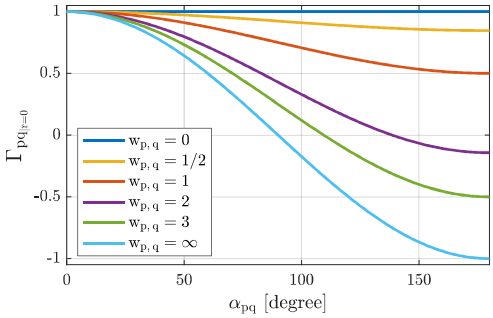


Fig. 5: Spatial coherence functions of coincident microphone pairs of same types ($w_p = w_q$) with respect to the angle α_{pq} between both microphones.

3 Spatial Coherence Function of Ambisonic Decoded Signals

Ambisonic signals are recorded by means of four non-coincident directional microphone capsules, mounted on a tetrahedral array, namely the A-format signal to be converted into a B-format signals, or directly as B-format signal. The B-format signal is eventually used a basis of decoding into multichannel audio for loudspeaker reproduction, or binaural audio for headphones reproduction.

3.1 Tetrahedral Arrays

If starting from an A format signal, tetrahedral or open-sphere arrays use radially aiming transducers of uniform shape $w_\mathrm{A} = w_p = w_q$. Any microphone pair, p and q, encloses an angle α_A such that $\mathbf{u}_p^\top \mathbf{u}_q = \cos\alpha_\mathrm{A}$, and aims symmetrically with regard to the connecting line $\mathbf{u}_p^\top \mathbf{r}_{pq} = -\mathbf{u}_q^\top \mathbf{r}_{pq} = \pm\sin\frac{\alpha_\mathrm{A}}{2}$, leading to

$$\Gamma^\mathrm{A}_{pq} = \frac{3 + {w_\mathrm{A}}^2\cos\alpha_\mathrm{A}}{3} j_0(kr_\mathrm{A}) - \frac{{w_\mathrm{A}}^2(3-\cos\alpha_\mathrm{A})}{6} j_2(kr_\mathrm{A}), \quad (21)$$

with $r_\mathrm{A} = ||\mathbf{r}_{pq}||$.

The covariance matrix $\mathbf{C}_A$ of a tetrahedral array is defined in the frequency domain and consists of the correlation coefficient of any pair of microphones $\mathbf{C}_A = [\Gamma^\mathrm{A}_{pq}]$.The A-format signals of a tetrahedral microphone array can then be encoded into B-format signals by a Hadamard matrix $\mathbf{H}$ separating the constant omni-directional component from the front-back, left-right, and up-down difference signals, followed by a diagonal weight b, directly related to the factor w_A. The covariance matrix of the encoded B-format microphone signals becomes

$$\mathbf{C}_\mathrm{B} = \mathrm{diag}\{\mathbf{b}\}\mathbf{H}\mathbf{C}_A\mathbf{H}\mathrm{diag}\{\mathbf{b}\}, \quad (22)$$

with

$$\mathbf{H} = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \end{bmatrix}, \quad \mathbf{b} = [1, b, b, b]. \quad (23)$$

3.2 Ambisonic Decoding

The B-format covariance matrix $\mathbf{C}_\mathrm{B}$ in (22) enables to derive the correlation coefficient of any pair of decoded signals with the weights $\mathbf{d}_p$, $\mathbf{d}_q$ as

$$\Gamma^\mathrm{D}_{pq} = \mathbf{d}_p^\top \mathbf{C}_\mathrm{B}\mathbf{d}_q, \quad (24)$$

Such a decoded, namely first order Ambisonics (FOA), signal can be interpreted as a virtual microphone signal because it is directly generated as a linear combination of the four channels of a B-format signal: W, X, Y, and Z, corresponding to an omni-directional signal and three figure-of-eight bidirectional signals pointing towards the three cartesian axes, respectively. Resulting in a first order directivity pattern v_p (see Table 1) and pointing toward the look direction (θ_p, ϕ_p):

$$\mathbf{d}_p = [v_p, (1-v_p)\mathbf{u}_p^\top]. \quad (25)$$

As an example, the spatial coherence function Γ^D_{pq} (24) of two decoded cardioids ($w_p = w_q = 0.5$), namely virtual cardioids, is compared to the spatial coherence functions of the corresponding true cardioid signals, $\Gamma_{pq_{|r=0}}$ and Γ_{pq}, respectively, assuming successively coincident cardioids, i.e. $r_{pq} = 0$ as in (20), and then, a spacing $r_{pq} = 20$ mm as in (18). This specific spacing has been chosen to match the dimension of a conventional A-format microphone, i.e. a tetrahedral array

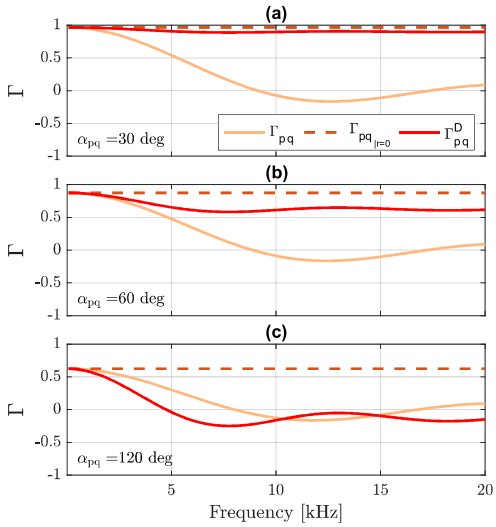


Fig. 6: Spatial coherence functions of two cardioids ($w_p = w_q = 0.5$), respectively: $\Gamma_{pq|r=0}$ for two coincident cardioids ($r_{pq} = 0$), Γ_{pq} considering two cardioids with $r_{pq} = 20$ mm spacing, and Γ^{D}_{pq} for two decoded cardioids from an A-format signal, these at various stereo angles: $\alpha_{pq} = 60, 90,$ and 120 degrees, successively in Panels (a), (b), and (c).

with a height $h = 12$ mm. Fig. 6 shows the resulting functions for three stereo decoding angles α_{pq}. These specific examples indicate that the diffuse sound, when rendering decoded signals from multi-microphone signals combination, can not be characterized or apprehended based on existing knowledge on the spatial coherence function of first-order microphone pairs: it needs deeper understanding on how the recording microphone signals are combined together.

Extending the results shown in Fig. 6, the differences between the spatial coherence functions Γ_{pq}, derived from (18), and Γ^{D}_{pq}, derived from (24), are illustrated in Fig. 7. Averaging the results over the full frequency range, the distance between the two functions Γ_{pq} and Γ^{D}_{pq} is derived for the specific case of two decoded cardioids at any stereo angle from 30 to 150 degrees as well as the 95% confidence interval.

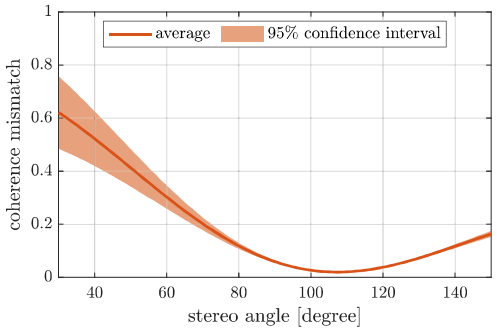


Fig. 7: Average distance and corresponding 95% confidence interval between the two spatial coherence functions Γ_{pq} and Γ^{D}_{pq} of two decoded cardioids over the full frequency range for stereo angles between 30 and 150 degrees.

4 Measurements and Applications

4.1 Diffuse Sound Field Measurements

A straightforward way to illustrate the above derivations is to measure a set of microphones within a diffuse sound environment: this has been practically carried out by using a loudspeaker emitting in a large room with a reverberation time (RT60) close to 1.2 s, and placing the microphones remotely from the loudspeaker. In order to get representative data, the measurement is repeated at 24 various positions, assuming that the diffuse sound field, independent from directions, should remain constant. The measured impulses responses are truncated in time to remove the direct and early parts, namely the direct sound and the early reflections, in order to only keep the late reverberation, or diffuse sound field.

In a first step two cardioid capsules mounted back-to-back, using the Schoeps CMXY stereo microphone, i.e with $\alpha_{pq} = 180$ degrees and a spacing of $r = 20$ mm, are measured to illustrate the theory of the spatial coherence function of two directional microphones as recalled in Section 2. Fig. 8 shows the theoretical spatial coherence function, Γ_{pq} (18), together with the measured ones Γ^{m}_{pq} from $M = 24$ various positions in Panel (a) and the corresponding average, $\overline{\Gamma}^{\mathrm{m}}_{pq}$, in Panel (b), respectively.

In a second step, the corresponding back-to-back cardioid signals are generated virtually, after decoding the

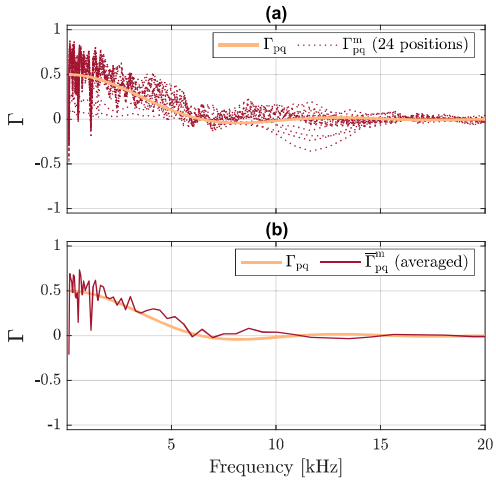


Fig. 8: Spatial coherence functions of two back-to-back cardioids ($\alpha_{pq} = 180$ degrees) measured within a diffuse sound field from 24 different positions in Panel (a) and the corresponding average curve in Panel (b), respectively, compared to the theoretical curve Γ_{pq}.

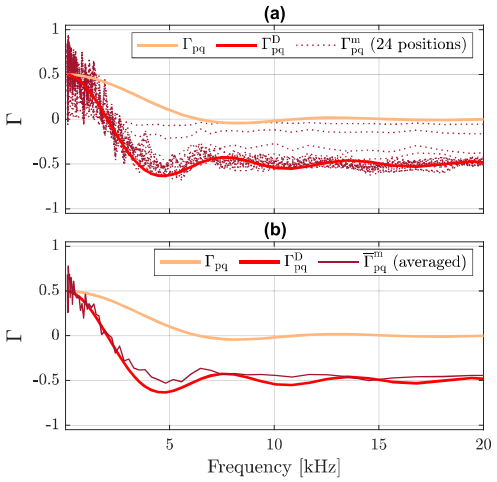


Fig. 9: Spatial coherence functions of two back-to-back cardioids ($\alpha_{pq} = 180$ degrees) decoded from an A-format microphone measured within a diffuse sound field from 24 different positions in Panel (a) and the corresponding average curve in Panel (b), respectively, compared to the theoretical curve Γ_{pq} and Γ^{D}_{pq}.

signals of a Soundfield SPS200 A-format microphone measured in the same way as the previous two cardioid capsules. The decoding, similarly as described in Section 3, involves the conversion from A- to B-format and the weights defined in (25). Fig. 9 shows the theoretical spatial coherence function, Γ^{D}_{pq} (24), and recalls the one, Γ_{pq} (18), for two cardioid capsules mounted back-to-back as shown in Fig. 8. Again the measured curves Γ^{m}_{pq} from the $M = 24$ various positions are shown in Panel (a) and the corresponding average, $\overline{\Gamma}^{m}_{pq}$, in Panel (b), respectively.

Both measurement examples verify the above derived theories and emphasize the differences between the original spatial coherence function of two first-order microphones and the spatial coherence function of the same virtual signals decoded from an Ambisonics microphone.

4.2 Decoding Applications

In light of these differences, the decoding of Ambisonics signals should be adapted for proper diffuse sound rendering.

As an example, as often, the decoding of Ambisonics signals follows a pair wise strategy i.e. generating stereo signals to be played back by loudspeaker symmetrically placed on both sides of the listener sweet spot. Whereas the spatial coherence functions of two signals with wide stereo angles in the range from 100 to 130 degrees are fairly close to the known behaviour described in [16] or [17]. Narrow stereo angles, i.e. from 30 to 60,will give a completely different behaviour for diffuse sound rendering as illustrated in Fig. 7. Thus the narrow front pairs of a multi-channel decoding will exhibit larger correlation compared to the wide side pairs which have negative correlation of diffuse sound.

Other examples involve parametric decoding techniques where the diffuse sound is purposely manipulated separately, by means of a diffuseness index [6], or post filtering using back-to-back cardioids [10] as illustrated in Fig. 9, require exact knowledge about the amount of diffuse sound such that it can be faithfully rendered.

It is also worth to notice that the same principles can be applied to any other recordings including multiple non-coincident directional microphones, such as Double or Triple MS, as well as any custom directional microphone arrays for VR recordings.

5 Conclusions

The spatial coherence function of a pair of microphone signals depends on the directivity patterns, on the angle, as well as on the spacing between these two microphones. The spatial coherence function of a pair of virtual microphone signals matches the one of real microphones when the source microphone signals are captured with coincident microphones, but will have modified diffuse sound properties if the source microphones (e.g. A-Format) are not coincident.

The derivation of the corresponding spatial coherence functions given two such first order virtual microphone signals has been proposed, analyzed, and compared to the spatial coherence functions of similar real microphone pairs.

Eventually the proposed theory has been verified through measurements obtained from room impulse response measurements in a large room showing how this theory matches the way the diffuse sound is recorded using an Ambisonics microphone. Both theoretical and measurement results, as well as the differences compared to real microphones, respectively, can be advantageously considered to draw conclusions on how the diffuse sound should be considered and comprehended when decoding Ambisonics signals.

References

[1] Gerzon, M. A., "The design of precisely coincident microphone arrays for stereo and surround sound," in *Preprint 50th Conv. Aud. Eng. Soc.*, 1975.

[2] Craven, P. G. and Gerzon, M. A., "Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs," *U.S. Patent Specification 4042779*, 1977.

[3] Farrar, K., "Soundfield Microphone: Design and Development of Microphone and Control Unit," *Wireless World*, pp. 48–50, 1979.

[4] Farrar, K., "Soundfield Microphone - 2: Detailed Functioning of Control Unit," *Wireless World*, pp. 99–103, 1979.

[5] Gerzon, M. A., "Optimum Reproduction Matrices for Multispeaker Stereo," *J. Audio Eng. Soc.*, 40(7/8), pp. 571–589, 1992.

[6] Pulkki, V., "Spatial Sound Reproduction with Directional Audio Coding," *J. Aud. Eng. Soc.*, 55(6), 2007.

[7] Berge, S. and Barrett, N., "High Angular Resolution Planewave Expansion (HARPEX)," in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, pp. 4–9, 2010.

[8] Zotter, F. and Frank, M., "All-Round Ambisonic Panning and Decoding," *J. Audio Eng. Soc.*, 60(10), pp. 807–820, 2012.

[9] Politis, A., Tervo, S., and Pulkki, V., "Coding and multidirectional parameterization of ambisonic sound scenes," in *Pro. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6802–6806, 2018.

[10] Favrot, A. and Faller, C., "B-Format Decoding Based on Adaptive Beamforming," in *Preprint 146th Conv. Aud. Eng. Soc.*, 2019.

[11] Mckeag, A. R. and McGrath, D. S., "Sound Field Format to Binaural Decoder with Head Tracking," in *Pro. 6th Australian Regional Audio Eng. Soc. Convention*, 1996.

[12] Zaunschirm, M., Schörkhuber, M., and Höldrich, C., "Binaural Rendering of Ambisonic Signals by Head-Related Impulse Response Time Alignment and a Diffuseness Constraint," *J. Acoust. Soc. Am.*, 6(143), pp. 3616–3627, 2018.

[13] Gorzel, M., Allen, A., Kelly, I., Kammerl, J., Gungormusler, A., Yeh, H., and Boland, F., "Efficient Encoding and Decoding of Binaural Sound with Resonance Audio," in *Pro. Audio Eng. Soc. International Conference on Immersive and Interactive Audio*, 2019.

[14] Kuster, M., "Spatial correlation and coherence in reverberant acoustic fields: Extension to microphones with arbitrary first-order directivity," *The Journal of the Acoustical Society of America*, 123(1), pp. 154–162, 2008.

[15] A. Perez-Lopez, N. S., "Analysis of spherical isotropic noise fields with an A-Format tetrahedral microphone," *The Journal of the Acoustical Society of America*, 146(4), pp. 329–334, 2019, doi:10.1121/1.5127736.

[16] Cook, R. K., Waterhouse, R. V., Berendt, R. D., Edelman, S., and Thompson, M. C., "Measurement of Correlation Coefficients in Reverberant Sound Fields," *Journal of the Acoustical Society of America*, 27(6), pp. 1072–1077, 1955.

[17] W. Elko, G., *Spatial Coherence Functions for Differential Microphones in Isotropic Noise Fields*, chapter 4, pp. 61–85, Springer, 2001.

[18] Mershon, D. H., , and King, L. E., "Intensity and reverberation as factors in the auditory perception of egocentric distance," *Perception & Psychophysics*, 18(6), pp. 409–415, 1975.

[19] Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization*, The MIT Press, Cambridge, Massachusetts, USA, revised edition, 1997.

[20] Abramowitz, M. and Stegun, I. A., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, 1964.

[21] Wittek, H., Haut, C., and Keinath, D., "Doppel-MS : eine Surround-Aufnahmetechnik unter der Lupe," in *24th Tonmeistertagung*, 2006.