OPEN ACCESS Freely available online

F. Salmon, G. Berthomieu, J. Palacino, and M. Paquier,
"The Influence of Diffuse-Field Equalization on the Perception of Spatial Aliasing Introduced by Spherical Microphone Arrays,"
J. Audio Eng. Soc., vol. 72, no. 10, pp. 650–663 (2024 Oct.).
https://doi.org/10.17743/jaes.2022.0157.

The Influence of Diffuse-Field Equalization on the Perception of Spatial Aliasing Introduced by Spherical Microphone Arrays

FRANÇOIS SALMON,[1,*] *AES Associate Member*, **GAUTHIER BERTHOMIEU,**[2]
(francois.salmon@noisemakers.fr) (gauthier.berthomieu@univ-brest.fr)
JULIAN PALACINO,[3] **AND MATHIEU PAQUIER**[2]
(julian.palacino@feichter-electronics.com) (mathieu.paquier@univ-brest.fr)

[1]*Noise Makers, 3 Rue Brossay Saint-Marc, 35700 Rennes, France*
[2]*University of Brest, CNRS, Lab-STICC UMR 6285, 6 Avenue Victor Le Gorgeu, CS 93837, 29238 Brest Cedex 3, France*
[3]*Feichter Audio, Espace de Broglie, 5 Rue Louis de Broglie, 22300 Lannion, France*

When a sound field is sampled by a finite number of microphones, the upper-frequency range of the recorded content is affected by spatial aliasing. The frequency beyond which aliasing-induced artifacts occur is related to the distance between a microphone and its nearest neighbors. In this study, ambisonics signals of order $N = 7$ were encoded from simulated recordings made by virtual spherical microphone arrays of varying characteristics (radius, sampling scheme, and sampling grid order). A diffuse-field equalization, the role of which is to lessen the aliasing-induced spectral coloration, was also considered in the tested conditions. These aliased stimuli were then compared with unaliased stimuli of order $N = 7$. Results showed that the diffuse-field equalization had a great impact on the perceived differences. A perceptual threshold was calculated in order to determine whether aliasing artifacts are likely to be audible based on the amount of aliasing-induced encoding error.

0 INTRODUCTION

Recent technological advances offer new opportunities for academic research and the entertainment industry to create audiovisual virtual environments that mimic real-life situations. When reproducing realistic scenarios, the accurate representation of the auditory and visual information is of primary interest. Consequently, numerous methods have been proposed to capture and reproduce a sound field in a way that is indistinguishable from natural hearing.

In particular, spherical microphone arrays (SMAs) are designed to capture a sound field in a physically accurate way. They are commonly used to analyze the spatial properties of a sound field or to create virtual and augmented reality contents. SMA recordings are usually employed to provide a description of sound fields in the spherical harmonic (SH) domain by the use of the ambisonic formalism. This is a common approach to reproduce an auditory scene at the position of a listener's head that allows efficient playback through headphones with head tracking (i.e., a dynamic binaural rendering).

*To whom correspondence should be addressed, email: francois.salmon@noisemakers.fr. Last updated: June 26, 2024.

However, the accuracy of the SH representation of the sound field depends on the physical characteristics of the SMA, which include the size of the array and the number, directivity, and positions of the sensors, as well as the presence of a baffle and its properties. These characteristics define the spatial resolution employed in the SH representation and the frequency range for which the SH modes are correctly estimated.

0.1 A Limited Spatial Resolution

When a sound field is transformed to the SH domain, the pressure that applies on the continuous surface elements (e.g., on the spherical surface of an SMA) is described through an infinite sum of discrete SH of different degrees and modes. In practice, this infinite sum is limited up to a maximal SH order, yielding a finite resolution of the recorded sound field that introduces truncation errors [1].

The physical accuracy of the sound field reconstruction thus depends on the maximal SH order—hereafter referred to as the ambisonic order—used for its decomposition in the SH domain. The ambisonic order is limited by the number and positions of the microphones used in the SMA to sample the sound field [2]. This truncation can significantly impact

the spectral content of a sound scene rendering, reducing the localization accuracy as well as the externalization [3].

Bernschütz et al. [4] and Bernschütz [5, SEC. 5.6.1] carried out quality-based listening experiments on binaural renderings of plane waves generated using multiple ambisonic orders. They observed no significant differences between full-resolution references and binaural renderings that used ambisonic orders greater than 11 for drums and guitar samples. The authors also concluded that the quality of the renderings started to be acceptable from an ambisonic order $N = 5$ and considered that an ambisonic order $N = 7$ provides the most reasonable trade-off between technical effort and perceptual quality.

Ahrens and Andersson [6] asked listeners to compare binaural renderings computed from SMA recordings of different resolutions with those of reference renderings obtained using mannequin-head measurements. Listeners rated to which extent the recordings and reference renderings differed in terms of spaciousness and timbre on two distinct scales. Stimuli of order $N \geq 8$ were rated as having no difference with the references when the sound sources were placed at a lateral position. Yet, ratings still exhibited small timbre differences up to order $N = 12$ when the sound sources were placed at a frontal position.

Lübeck et al. [7] evaluated points of subjective equality (PSEs) between binaural stimuli synthetized from SMA signals of varying ambisonic order and a binaural reference synthetized from an SMA of order $N = 29$ (for which the introduced artifacts can be neglected). PSEs were found to be between orders 17 and 20, thanks to an ABX three-interval/two-alternative forced-choice test design. These thresholds, in term of indistinguishability, are higher than the minimum orders of 8–12 found by Ahrens and Andersson's quality-based listening experiment. This can be accounted for by the use of a more conservative paradigm: small dissimilarities could be detected without leading to different ratings in terms of spectral quality and spaciousness.

The orders beyond which artifacts are no longer perceptible also depend on the method used for decoding the ambisonic signals to binaural signals. Ahrens and Andersson [6] performed binaural decoding in the SH domain, whereas Bernschütz et al. [4] and Lübeck et al. [7] convolved plane waves representing the sound field with the corresponding head-related impulse responses using the virtual loudspeaker method [8]. Several methods have been proposed to perceptually improve the binaural decoding of order-truncated ambisonic signals using optimized representation of head-related transfer functions (HRTFs) at the truncated ambisonic order. Most methods are based on a time alignment of binaural filters prior to interpolation [9, 10] or spectral equalization of HRTFs [11]. A common method uses magnitude-only optimization of the binaural filters at high frequencies [12].

A listening test conducted by Lübeck [13] showed that the preprocessing of binaural filters significantly reduces the perceived differences between order-truncated SMA renderings and a mannequin-head reference. Similarly, Arend et al. [14] reported that with the method described by Pörschmann et al. [11], a binaural rendering of a percussive frontal source required an order $N \geq 7$ to achieve indistinguishability from a full-resolution reference, whereas order $N \geq 13$ was required with unprocessed binaural filters.

No clear agreement on the minimum ambisonic order leading to a natural listening experience can be found in the literature, because it appears to depend on the sound content, source position, decoding method, and experimental protocol. In some cases, the reduction of the ambisonic order can have a great impact on the number of sensors involved in the SMA while impairing the perception of the rendering to a limited extent only. In practice, a reasonable compromise must then be found between the SMA technical complexity and the resulting perceptual quality.

0.2 A Limited Operational Bandwidth

The contribution of the different SH orders is not equal at all frequencies. At low frequency, when the SMA radius is small in comparison to the incident wavelength, the contribution of higher-order ambisonic components is very small [15]. The differences between microphone signals are too small to provide precise spatial information and need to be greatly amplified in order to properly estimate high-order ambisonic signals. However, this amplification has to be limited in practice to prevent the resulting ambisonic signals from being excessively polluted by the noise created from the addition of the uncorrelated self-noise of each sensor. Thus, at frequencies where this limit is effective, the spatial resolution of the SH representation is further reduced.

At high frequency, spatial ambiguities arise due to the spatial sampling of sound field with a limited number of sensors [16]. This phenomenon appears above the frequency for which the distance between a microphone and its nearest neighbors is equal to half a wavelength. Above this frequency, spatial aliasing occurs, and low-order ambisonic signals are being corrupted by spatial information that belongs to higher-order components.

The minimum order N, providing a representation of the sound field with an error of 4% in a sphere of radius r, is estimated by [17, 18]:

$$N = \lceil kr \rceil, \tag{1}$$

where k is the wave number $k = \dfrac{2\pi f}{c}$ and $\lceil . \rceil$ is the ceiling function. Hence, SH modes of orders $N > kr$ can be considered to have a negligible influence, and no significant spatial aliasing occurs if k and r satisfy $kr < N$. It follows that for a sound field sampled on a sphere of radius r_s according to a sampling grid that allows estimation of the ambisonic signals up to order N, the aliasing error is expected to be negligible below the frequency f_a, usually referred to as the aliasing frequency:

$$f_a = \frac{cN}{2\pi r_s}. \tag{2}$$

The lower N is, the lower f_a will be and the more spatial aliasing will induce audible artifacts. Above this frequency, components of orders greater than N will no longer be

negligible, and their aliasing in lower orders will lead to potential audible artifacts.

Ahrens and Andersson [6] compared SMA-based binaural signals to mannequin-head auralizations in terms of spaciousness and timbre. A drum audio signal was used, which could be either unfiltered or low-pass filtered with a cutoff frequency of 3 kHz. No significant difference was observed between the SMA and mannequin-head signals at any ambisonic order under investigation when the source was low-pass filtered. The authors thus attributed these spaciousness and timbre quality losses to high-frequency errors caused by spatial aliasing.

However, the aliasing properties, notably the pattern whereby high orders are aliased into low orders, of an SMA are also related to the sensor positions [16]. A given number of microphones can be distributed on a spherical array according to multiple sampling schemes [19]. For a same aliasing frequency, some sampling schemes require fewer microphones than other ones. As an example, a Gauss-Legendre [20] sampling scheme requires twice as many microphones than an extremal [21] sampling scheme for reaching the same aliasing frequency. Furthermore, Lecomte et al. [22] showed that a Lebedev sampling scheme produced less encoding error due to aliasing at capture compared with spherical t-design and Fliege sampling schemes while considering roughly the same number of sampling points.

Bernschütz [5, SEC. 5.6.3] led a perceptual study on the relation between the sampling scheme and spatial aliasing. He compared drums signals that were encoded following a simulation of two sampling schemes—Gauss-Legendre and Lebedev—at four ambisonics orders ($N = 4, 5, 6, 7$) while keeping a fixed aliasing frequency of $f_a = 2\,\text{kHz}$ thanks to an adaptation of the sphere radius. A multiple stimulus with hidden reference and anchor test reported that the stimuli encoded on a Gauss-Legendre grid led to significantly higher-quality ratings than the ones encoded on a Lebedev grid. This observation was, however, only made at the order $N = 6$ (no significant difference was highlighted between the two schemes at the three other orders).

Bernschütz also compared quality ratings of signals encoded following a simulation of spatial samplings of varying aliasing frequency either due to i) modification of the SMA radius or ii) modification of the number of microphones. It was reported that increasing the number of microphones appeared to be more effective than reducing the sphere radius to lessen the perception of artifacts. Consequently, the aliasing frequency alone can not be considered as a good estimator of the spatial aliasing perception, because it does not take into account the sampling scheme or the precise number of microphones over the sphere radius.

0.3 Correction of the Spectral Coloration

An equalization of the estimated components can be performed beyond the frequency where aliasing appears in order to lessen aliasing-induced spectral coloration [23–26]. Consider an ideal diffuse field represented in the SH domain for which the main properties are that i) the norm of each SH component is equal and ii) the correlation between two different SH components is zero [27]. If such a sound field is captured with an SMA, spatial aliasing occurs and higher-order components will corrupt lower-order components following a pattern that depends on the SMA properties [16]. Consequently, the estimated sound field represented in the SH domain will suffer from i) unequal norm across SH components and ii) nonzero correlation across SH components.

A diffuse-field equalization (DFE) can be applied to the encoding matrix in order to lessen the spectral coloration induced by such artifacts. The DFE consists in modifying the energy of each estimated component of such diffuse field to that of an ideal diffuse sound field by forcing equal norm for each SH component. Further processing can also be applied to suppress the interchannel correlation of the estimated components [28]. Deppisch et al. [29] proposed a preprocessing method of HRTF to create binaural decoding filters dedicated to the rendering of SMA recordings by considering the SMA properties. This method improves the high-frequency response by approximating the average magnitude response of HRTFs over the sphere, which is comparable with a DFE. Although it has been proven that such equalization improves the high-frequency response in terms of magnitude error, no listening test was performed, to the authors' knowledge, to demonstrate the relevance of a DFE on SMA-encoded signals.

0.4 Aim of the Study

This study addresses the impact of the SMA properties on the perception of spatial aliasing using nonindividualized dynamic binaural playback of seventh-order ambisonic scenes. The aim of this study is to evaluate the role of the SMA properties and the DFE in the perception of spatial aliasing artifacts. A listening test was conducted to determine the number of microphones beyond which spatially aliased stimuli become indistinguishable from nonaliased stimuli using different radii and sampling schemes with and without DFE. The binaural decoding was performed using a well-known preprocessing method of the binaural filters, which is the magnitude least-square optimization of HRTFs (referred to as the MagLS method) [12].

The maximal SH order that an SMA can estimate is defined by its grid order, noted N_g, which depends on the number of sampling points that constitute this SMA. The experiment was performed using various N_g but without changing the ambisonic order of the rendering stage, noted $\tilde{N}$. The stimuli were decoded using a constant ambisonic order so that a modification in the number of microphones only had an impact on the ambisonic signal estimation without affecting the binaural decoding stage. The ambisonic order of the binaural reproduction was set to $\tilde{N} = 7$ because Bernschütz reported that it provides a reasonable trade-off between technical effort and perceptual quality [5].

In the following sections, the encoding process of SMA signals as well as the DFE will be presented. Then, the creation of the aliased and nonaliased stimuli will be described.

Finally, the results of the listening test will be presented and discussed. In order to make the results applicable to different SMA configurations, a metric will be introduced for quantifying the spatial aliasing error with regard to the SMA properties. The results of the listening test will then be discussed using this metric.

1 SOUND FIELD ENCODING IN THE AMBISONIC DOMAIN

In this study, the authors consider the case of an SMA that comprises Q omnidirectional microphones located on the surface of a perfectly rigid sphere with radius r_s.

1.1 Approximation of the Sound Pressure

The value of the sound pressure $p_q(k)$ at the frequency corresponding to a wave number k measured at the sensor position $\Omega_q = (\theta_q, \phi_q)$ in a spherical coordinate system can be approximated as

$$p_q(k) \approx \sum_{n=0}^{N} \sum_{m=-n}^{n} w_n\left(kr_s\right) Y_n^m\left(\Omega_q\right) b_{n,m}(k), \tag{3}$$

where Y_n^m is the value of the SH function of order n and degree m for direction Ω_q, $b_{n,m}$ is the SH component of order n and degree m of the incident sound field, and w_n is a radial transfer function that includes the effect of the incident sound field and the scattered sound field around the rigid sphere [19].

The vector $\mathbf{p}$ of the approximated sound pressure received at the Q microphones can be expressed as

$$\mathbf{p} = \mathbf{T}_N \mathbf{b}_N, \tag{4}$$

where

$$\mathbf{b}_N = \left[b_{0,0}, b_{1,-1}, ..., b_{n,m}, ..., b_{N,N}\right]^T,$$
$$\mathbf{T}_N = \mathbf{Y}_N \text{diag}(\mathbf{w}),$$

and

$$\mathbf{w} = \left[w_0\left(kr_s\right), w_1\left(kr_s\right), ..., w_n\left(kr_s\right), ..., w_N\left(kr_s\right)\right]^T,$$
$$\mathbf{Y}_N = \left[\mathbf{y}_N(\Omega_0), \mathbf{y}_N(\Omega_1), ..., \mathbf{y}_N(\Omega_Q)\right]^T,$$
$$\mathbf{y}_N(\Omega_q) = \left[Y_0^0\left(\Omega_q\right), ..., Y_n^m\left(\Omega_q\right), ..., Y_N^N\left(\Omega_q\right)\right]^T.$$

$\mathbf{T}_N$ represents the transfer matrix between the SH components of orders up to N and the sound pressure at the microphone positions. To satisfy the criteria of Eq. (1), the maximum order N of the SH representation must satisfy $kr_s \leq N$, where k can be obtained using an upper frequency limit $f_a = 20\,\text{kHz}$ to cover the whole audible spectrum. In this way, given an incident sound field representation $\mathbf{b}_N$, an approximation of the sound pressure at the position of each sensor constituting the SMA can be obtained.

1.2 Estimation of the SH Components

Let $\mathbf{b}_{N_g}$ be the vector of SH components that represents the incident sound field up to order $N_g \leq N$. When $Q \geq (N_g + 1)^2$, the problem of finding the values of the SH components from the microphone signals can be seen as an overdetermined system where one wants to find the components that minimize the quantity $\left\|\mathbf{p} - \mathbf{T}_{N_g}\mathbf{b}_{N_g}\right\|^2 + \beta^2\left\|\mathbf{b}_{N_g}\right\|^2$, where β is a regularization parameter. In practice, regularization prevents excessive amplification of measurement noise at low frequency by adding a constraint on the norm of the SH components [30, 31].

The solution of this problem is the estimated SH components, noted $\hat{\mathbf{b}}_{N_g}$, that are obtained from the microphone signals using the encoding matrix $\mathbf{E}_{N_g}$:

$$\hat{\mathbf{b}}_{N_g} = \mathbf{E}_{N_g}\mathbf{p}. \tag{5}$$

If the number of microphones and their position over the sphere make $\mathbf{T}_{N_g}$ an invertible matrix, $\mathbf{E}_{N_g}$ is calculated by inverting the matrix $\mathbf{T}_{N_g}$. The encoding matrix $\mathbf{E}_{N_g}$ can be solved using a Tikhonov regularization as described by Moreau et al. [32]:

$$\mathbf{E}_{N_g} = (\mathbf{T}_{N_g}^H\mathbf{T}_{N_g} + \beta^2\mathbf{I}_{N_g})^{-1}\mathbf{T}_{N_g}^H, \tag{6}$$

where β is the regularization parameter and $\mathbf{I}_{N_g}$ is the $(N_g + 1)^2 \times (N_g + 1)^2$ identity matrix.

1.3 Spatial Aliasing Error

The estimated SH components $\hat{\mathbf{b}}_{N_g}$ of an incident sound field represented by the vector $\mathbf{b}_N$ can be expressed as follows:

$$\hat{\mathbf{b}}_{N_g} = \mathbf{\Gamma}\mathbf{b}_N, \tag{7}$$

where $\mathbf{\Gamma} = \mathbf{E}_{N_g}\mathbf{T}_N$. Now, decompose the error obtained between the estimated SH components and the SH components up to order N_g to analyze the contribution of the aliasing error in the same manner as in [22]. The spatial aliasing error can be expressed by splitting the $\mathbf{\Gamma}$ matrix and $\mathbf{b}_N$ vector in two parts:

$$\hat{\mathbf{b}}_{N_g} = \left[\mathbf{\Gamma}_{N_g}\ \mathbf{\Gamma}_\Delta\right]\begin{bmatrix}\mathbf{b}_{N_g}\\ \mathbf{b}_\Delta\end{bmatrix}, \tag{8}$$

where $\mathbf{\Gamma}_\Delta$ and $\mathbf{b}_\Delta$ represent the higher-than-order-N_g elements. The encoding error ϵ_{N_g}, which results from the sound field encoding in the SH domain up to order N_g, can be expressed as

$$\epsilon_{N_g} = \frac{\left\|\mathbf{b}_{N_g} - \hat{\mathbf{b}}_{N_g}\right\|}{\left\|\mathbf{b}_{N_g}\right\|} \tag{9}$$

$$= \frac{\left\|\mathbf{b}_{N_g} - \mathbf{\Gamma}_{N_g}\mathbf{b}_{N_g} - \mathbf{\Gamma}_\Delta\mathbf{b}_\Delta\right\|}{\left\|\mathbf{b}_{N_g}\right\|}. \tag{10}$$

In this study, the authors use virtual SMA signals that induce no measurement noise in the microphone signals. In this ideal case, the amplification of the SH components at low frequency is not limited, thus avoiding a low-frequency error in the estimation of the SH components. Therefore, $\beta = 0$, $\mathbf{\Gamma}_{N_g} = \mathbf{I}_{N_g}$, and the decoding error corresponds only to the aliasing error:

$$\epsilon_{a,N_g} = \epsilon_{N_g|\beta=0} = \frac{\|\mathbf{\Gamma}_\Delta\mathbf{b}_\Delta\|}{\left\|\mathbf{b}_{N_g}\right\|}. \tag{11}$$

Fig. 1 shows the relation between the aliasing error, the sphere radius, and frequency for different numbers of microphones that cover the sphere with a Gauss-Legendre

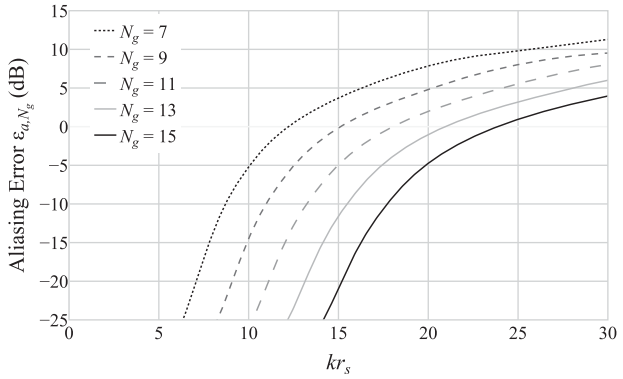


Fig. 1. The aliasing error as a function of kr_s using an SMA with a Gauss-Legendre sampling scheme for different sampling grid orders N_g.

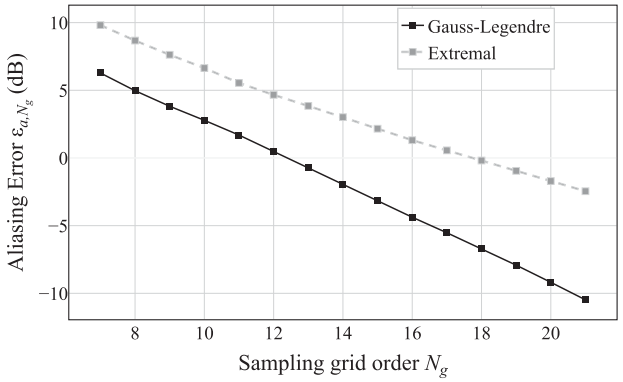


Fig. 2. The aliasing error averaged over frequency as a function of the sampling grid order N_g for an 8.5-cm–radius SMA using Gauss-Legendre and extremal sampling schemes and $N = 33$.

sampling scheme [21]. To compute the aliasing error, a diffuse sound field composed of plane waves coming from 5,200 directions in accordance with a spherical t-design sampling scheme was considered [33, 34]. As mentioned earlier, the aliasing error increases with frequency and sphere radius, while it decreases with the microphone density over the sphere. Fig. 2 illustrates the relation between aliasing error and the number of microphones involved for two different sampling schemes, highlighting the dependency of the spatial aliasing on the sampling scheme.

1.4 Equalization of the Estimated Components

In order to lessen the aliasing-induced spectral coloration, the estimated components can be equalized in the same manner as the time-invariant ambisonic encoding described in McCormack et al. [26] and as presented in SEC. 0.3. Here, the equalization is applied beyond the spatial aliasing frequency defined in Eq. (2). It is performed on the encoding matrix using an equalization matrix $\mathbf{G}$, composed of real positive values:

$$\mathbf{E}_{N_g}^{(eq)} = \mathbf{G}\mathbf{E}_{N_g}, \tag{12}$$

where $\mathbf{G} = \text{Diag}\,[\mathbf{g}]$ and Diag [.] denotes a diagonal matrix whose elements correspond to the enclosed vector.

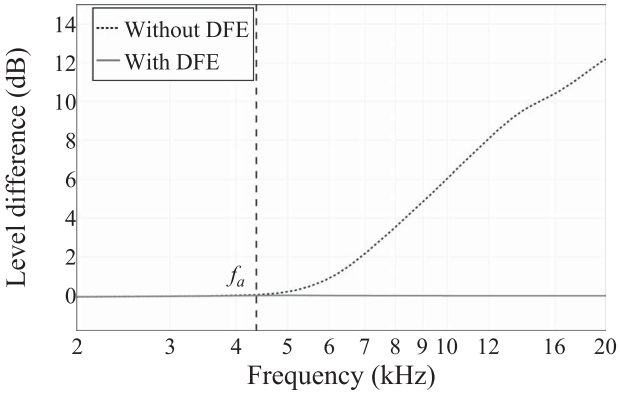


Fig. 3. Mean level difference over all directions between the ideal and simulated components for an 8.5-cm–radius SMA with a Gauss-Legendre sampling scheme.

To determine the vector $\mathbf{g}$, a diffuse sound field represented by matrix $\mathbf{B}$ was considered. As shown by Epain and Jin [27], in a perfectly diffuse sound field, SH signals are mutually uncorrelated and have equal power. Hence, the corresponding covariance matrix is proportional to the identity matrix. $\mathbf{B}$ is then defined as $D \geq (N+1)^2$ plane waves coming from uniformly distributed directions such as $\frac{1}{D}\mathbf{B}^T\mathbf{B} = \mathbf{I}_N$. Let $\mathbf{B}_{N_g}$ be the up-to-order-N_g SH components of the diffuse sound field and $\hat{\mathbf{B}}_{N_g}$ the estimated ones:

$$\hat{\mathbf{B}}_{N_g} = \mathbf{E}_{N_g}^{(eq)}\mathbf{T}_N\,\mathbf{B}_N = \mathbf{G}\boldsymbol{\Gamma}\mathbf{B}_N. \tag{13}$$

Here, the DFE is intended to equalize the spectral densities of the SH components to match those of a diffuse sound field. Since only a spectral correction is intended, the cross-correlations between the SH components were left unchanged. In order to achieve equal power of the SH signals, individual equalization of each channel is required so that the following equation is verified:

$$\text{diag}\left[\hat{\mathbf{B}}_{N_g}\hat{\mathbf{B}}_{N_g}^H\right] = \text{diag}\left[\mathbf{B}_{N_g}\mathbf{B}_{N_g}^H\right], \tag{14}$$

$$\text{diag}\left[\mathbf{G}\boldsymbol{\Gamma}\mathbf{B}_N\mathbf{B}_N^H\boldsymbol{\Gamma}^H\mathbf{G}^H\right] = \text{diag}\left[\mathbf{B}_{N_g}\mathbf{B}_{N_g}^H\right], \tag{15}$$

$$\mathbf{g}^2\,\text{diag}\left[\boldsymbol{\Gamma}\boldsymbol{\Gamma}^H\right] = \mathbf{1}_{(N_g+1)^2}, \tag{16}$$

$$\mathbf{g} = \text{diag}\left[\boldsymbol{\Gamma}\boldsymbol{\Gamma}^H\right]^{-1/2}, \tag{17}$$

where diag [.] denotes the diagonal vector of the enclosed square matrix.

Fig. 3 displays the mean level difference across all directions of the sampling sphere without (dotted line) and with (solid line) applying the DFE to the encoding matrix. The flat solid line at $y = 0$ dB shows the uniformity of the equalized field beyond f_a.

It is noteworthy that while this process reduces the spectral coloration of the estimated components by reducing the energy of the estimated components at high frequency, the aliasing-induced spatial artifacts remain because interchannel correlations are still present.

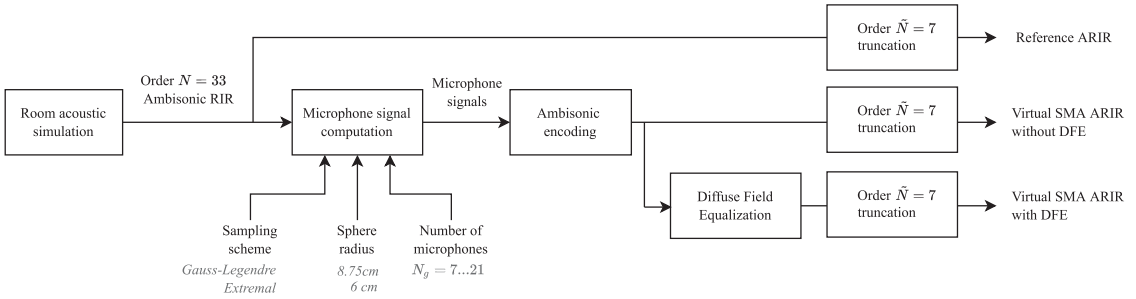


Fig. 4. Processing structure for generating the ARIRs that were used to generate the stimuli of the listening test.

2 LISTENING TEST

The listening test was conducted to evaluate the grid order N_g, beyond which the spatial aliasing error is no longer perceptible, for two radii and two sampling schemes. An adaptive procedure consisting of pairwise comparisons was carried out between spatially aliased stimuli and reference signals that did not exhibit spatial aliasing. The aliased stimuli were computed using different virtual SMAs and reproduced through headphones following a binaural decoding.

2.1 Ambisonic Room Impulse Responses

A small sound space modeled as a shoebox was considered in the listening test. The room dimensions were 5.3 × 4.1 × 2.3 m (width × depth × height), and the reverberation time was equal to 0.24 s. Five sound sources were used and located with an azimuth angle of ±40°, ±15°, and 0°. They had a cardioid directivity and were placed 1.15 m above the floor and 2 m away from the listening position. The receiver position was set to 1.15 m above the floor to fit the position of the listener head during the listening test. The resulting direct-to-reverberant energy ratio was −0.56 dB. It was calculated from omnidirectional room impulse responses and averaged over all sound sources. The integration time window used to calculate the direct sound energy was set to 2.5 ms.

Ambisonic room impulse responses (ARIRs) were simulated to model the incident sound field in the vicinity of the virtual SMA position. In accordance with Eq. (1), an order $N = 33$ is required to approximate the sound pressure at the microphone positions over a sphere with a radius as large as 8.75 cm and up to 20 kHz. Hence, ARIRs were simulated at this order using the MCRoomSim software [35].

To create virtual SMA signals, the microphone signals were first computed using different sampling grid orders N_g, sphere radii, and sampling schemes. The computation was performed from the ARIRs of order $N = 33$ as expressed in Eq. (3). The microphone signals were then encoded in ambisonics using Eq. (6). The regularization parameter was set to $\beta = 1.12e^{-12}$ (rather than the theoretical value $\beta = 0$) to prevent any numerical instability. This limits the maximum amplification gain of the encoding filters at 233 dB. The virtual SMA signals were processed with and without DFE so that the perceptual benefits of the spectral correction described in SEC. 1.4 could be evaluated.

Table 1. Number of microphones involved in the listening test for each sampling scheme (left) along with the corresponding aliasing frequency for each radius (right).

	Sampling scheme		Radius	
	Extremal	Gauss-Legendre	6 cm	8.75 cm
	N_g		Aliasing frequency (kHz)	
7	64	128	6.4	4.4
8	81	162	7.3	5.0
9	100	200	8.2	5.6
10	121	242	9.1	6.2
11	144	288	10.0	6.9
12	169	338	10.9	7.5
13	196	392	11.8	8.1
14	225	450	12.7	8.7
15	256	512	13.6	9.4
16	289	578	14.6	10.0
17	324	648	15.5	10.6
18	361	722	16.4	11.2
19	400	800	17.3	11.9
20	441	882	18.2	12.5
21	484	968	19.1	13.1

The ambisonic reference stimuli were generated by truncating the signals of the full simulation order $N = 33$ to $\tilde{N} = 7$, and the stimuli under test were generated by using the SMA processing according to the different sampling grids of orders $N_g \geq \tilde{N}$, with subsequent truncation to $\tilde{N} = 7$. A summary of this process is depicted in Fig. 4.

A total of 60 different virtual SMAs were created by combining two radii (6 and 8.75 cm), 15 sampling grid orders (7–21), and two sampling schemes (Gauss-Legendre [20] and extremal [21]). For these sampling schemes, a minimum number of sampling points $Q = (N_g + 1)^2$ and $Q = 2^*(N_g + 1)^2$ are required for extracting SH coefficients at a given sampling grid order N_g, respectively. Extremal is most efficient in terms of minimum number of microphones, because Gauss-Legendre requires twice as many microphones for reaching the same grid order N_g. The virtual SMA properties and corresponding aliasing frequencies computed from Eq. (2) are shown in Table 1.

The resulting ARIRs were then convolved with anechoic sound sources to create the stimuli. As a result, the reference

stimuli were unaliased ambisonic signals of order $\tilde{N} = 7$ and the aliased stimuli were ambisonic signals of order $\tilde{N} = 7$ with varying amount of spatial aliasing, rendered either with or without diffuse field equalization.

2.2 Sound Sources

Two audio contents were considered in the experiment:

- A 2-s–long pink noise and
- A 4.5-s–long musical excerpt of "Don't Mean a Thing" from Duke Ellington, played by a quartet comprising a violin, tenor saxophone, guitar, and double-bass. The sound sources were recorded in an anechoic chamber by Thery and Katz [36].

These contents were chosen so that the set contained a critical signal (noise) and a more realistic signal (a musical excerpt).

Sound sources from the quartet were located at -40°, -15°, 15°, and 40° in the horizontal plane. The audio signals were convolved with the corresponding ARIR, and the resulting signals were summed to synthesize the seventh-order ambisonic scene. The pink noise signal was convolved with ARIRs computed for a sound source located at 0 in the horizontal plane.

2.3 Dynamic Binaural Rendering

Stimuli were presented over open, circumaural headphones (Sennheiser HD 650). The ambisonic signals were converted into binaural headphone signals using the Ambi Head HD VST plug-in from Noise Makers [37] with HRTFs derived from measurements of a Neumann KU 100 mannequin head [38]. The binaural filters implemented the binaural rendering approach of Schörkhuber et al. [12] and Zaunschirm et al. [10]. Head tracking was performed using the T3 head tracker from Feichter Audio [39] along with the Ambi Head HD VST plug-in.

2.4 Procedure

The aliasing frequency f_a of various SMAs was shifted up or down by reducing or increasing their sampling grid order N_g following an adaptive procedure that led to a PSE [40]. Beyond this sampling grid order PSE, aliased stimuli were considered to be indistinguishable from unaliased references.

The procedure consisted in a $2 \times 2 \times 2 \times 2$ repeated-measures design gathering 16 runs [2 radii $\times$ 2 schemes $\times$ 2 DFE (on/off) $\times$ 2 audio contents], as each of these parameters are likely to affect the way spatial aliasing–induced artifacts are perceived. Each run started at a grid order $N_g = 7$ (i.e., with the lowest number of microphones for the corresponding sampling scheme given in Table 1), and the grid order was either increased or decreased after each trial of the run following a one-up-one-down staircase method.

Each trial of the run consisted of a three-interval/three-alternative forced-choice test. During each trial, three sounds (*A*, *B*, and *C*) were presented to the listener. Two sounds were identical (the aliased version or unaliased reference), and the third sound (named the *oddball*) was different (the unaliased reference or aliased version, respectively), randomly leading to one of the six possible assignations depicted in Table 2. The sounds *A*, *B*, and *C* were played once during each trial. Participants were instructed to designate which of *A*, *B*, or *C* was the oddball on a graphical user interface or using a keyboard.

Table 2. Each of the six possible assignments of the sounds *A*, *B*, and *C*.

xxR	*xRx*	*Rxx*
RRx	*RxR*	*xRR*

Note. *x* means that the sound is the aliased version; *R* means that the sound is the unaliased reference.

If the listener correctly identified the oddball (that is, by correctly identifying which of *A*, *B,* or *C* was assigned a different stimulus than the other two), the grid order of this run was increased to the next step, whereas it was decreased to the previous step if the listener gave the wrong answer. Each run ended after 12 reversals (a reversal is defined as a correct answer followed by a wrong answer or vice versa). In order to minimize learning effects, consecutive trials could belong to different runs [that is, different combinations of radius, sampling scheme, DFE (on/off), and audio content]. This way, the test appeared to the listener as a series of unrelated trials.

The experiment consisted of two 30-min–long sessions, one gathering the eight runs that concerned the ARIRs computed from Gauss-Legendre sampling grids and one gathering the eight runs that concerned the ARIRs computed from extremal sampling grids. The two sessions were carried out in a random order by each participant and were separated by a break. A training phase allowed subjects to become thoroughly familiar with the test environment and the grading process at the beginning of each session. Playback, commands, and data capture were controlled by a software implemented in Max 8.

2.5 Participants

A total of 15 participants, including five women and ten men with a mean age of 26.5 years ($\sigma = 5.6$ years), took part in the listening test and were remunerated for their participation. They were all related to the Sound Master's degree of the University of Brest and already had strong experience in critical listening.

3 RESULTS

A repeated-measures analysis of variance (ANOVA) was performed on the grid-order PSEs N_g. This analysis included four factors: the SMA radius (two levels), SMA sampling scheme (two levels), audio content (two levels), and DFE (on/off; two levels). The residuals of the linear model were approximately normally distributed. The results of the ANOVA are presented in Table 3. These results comply with and extend the preresults presented in a brief conference paper [41].

Table 3. Results of the ANOVA conducted on the grid-order PSEs.

Cases	DF	F	Sig.p	η^2_G
AC	1	12.850	0.003	0.065
SS	1	39.172	< 0.001	0.098
R	1	228.213	< 0.001	0.437
DFE	1	113.911	< 0.001	0.489
AC × SS	1	2.830	0.1150	0.015
AC × R	1	0.084	0.777	< 0.001
SS × R	1	0.369	0.553	0.001
AC × DFE	1	0.022	0.884	< 0.001
SS × DFE	1	17.454	0.001	0.067
R × DFE	1	71.776	< 0.001	0.181

Note. F depicts the F-ratio of the model, which is the ratio of the model to its error [46]. Sig.p depicts the p value associated with the given F-ratio, which is the probability to obtain results at least as extreme as the experimental results under the assumption that the null hypothesis is correct. η^2_Gdepicts the generalized eta squared, which is an effect size estimate used in repeated-measures designs [47]. For the sake of readability, only main effects and simple interactions are displayed (no three or four-way interaction was revealed to have a significant effect). AC = audio content; DF = degrees of freedom [48]; R = SMA radius; SS = SMA sampling scheme.

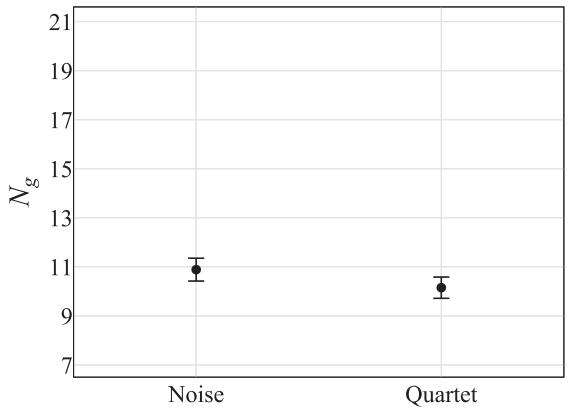


Fig. 5. *Ng* PSEs as a function of the audio content with 95% confidence intervals.

3.1 Simple Effects

The simple effect of the audio content factor [$F(1, 14) = 12.85$; $p = 0.003$] is rather small ($\eta^2_G = .065$). As depicted in Fig. 5, the pink noise exhibited slightly more aliasing than the jazz quartet, because a higher grid order N_g was required for the aliased stimuli to be more undistinguished from nonaliased references when the audio content was the pink noise ($\bar{N}_g = 10.9$) than when it was the jazz quartet ($\bar{N}_g = 10.2$).

The ANOVA also revealed a significant effect of the SMA radius [$F(1, 14) = 228.213$; $p < 0.001$] of the sampling scheme [$F(1, 14) = 39.172$; $p < 0.001$] and of the DFE [on/off; $F(1, 14) = 113.911$; $p < 0.001$]. The effects of these three factors on the PSEs will be analyzed in the significant interactions in which they are involved.

3.2 Sampling Scheme × DFE

The ANOVA revealed a significant interaction between the SMA sampling scheme and DFE [on/off; $F(1, 14) = 17.454$; $p = 0.001$]. As depicted in Fig. 6, the Gauss-Legendre sampling scheme led to lower N_g PSEs than the extremal sampling scheme when the encoding matrices were not diffuse-field equalized. Moreover, the DFE appeared to lower the PSEs for both sampling schemes.

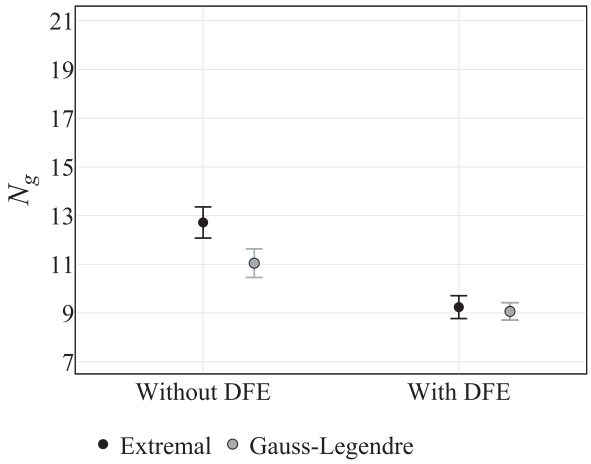


Fig. 6. N_g PSEs as a function of the DFE condition for extremal and Gauss-Legendre sampling scheme with 95% confidence intervals.

Table 4. Bonferroni-corrected pairwise comparisons of the PSEs obtained with different DFE and SS conditions.

Case 1		Case 2			
SS	DFE	SS	DFE	Sig.p*	d
Extremal	On	Extremal	Off	<0.001	1.35
G-L	On	G-L	Off	<0.001	0.97
Extremal	Off	G-L	Off	<0.001	1.01
Extremal	On	G-L	On	1.0	0.09

Note. Sig.p is the p value of the post hoc test. d is the Cohen's d measure of the magnitude of the observed effect. SS = sampling scheme; G-L = Gauss-Legendre.
*The p value was adjusted with a Bonferroni correction accounting for four pairwise comparisons.

In order to statistically evaluate these observations, four Bonferroni-corrected pairwise comparisons [42] were run. The results of these comparisons are given in Table 4. The observations made on Fig. 6 are confirmed, because every pairwise comparison led to a significant difference, except for the comparison between the two sampling schemes when DFE was applied.

3.3 Radius × DFE

The ANOVA also revealed a significant interaction between the SMA radius and DFE [on/off; $F(1, 14) = 71.776$; $p < 0.001$]. As depicted in Fig. 7, the SMAs with a radius of 6 cm led to lower PSEs than the SMAs with a radius of 8.75 cm, and all of these PSEs were lowered by a DFE of the encoding matrices.

The results of Bonferroni-corrected pairwise comparisons are given in Table 5. These comparisons support that the DFE significantly lowered the PSEs for both radii. Moreover, while the radius significantly affected the PSEs

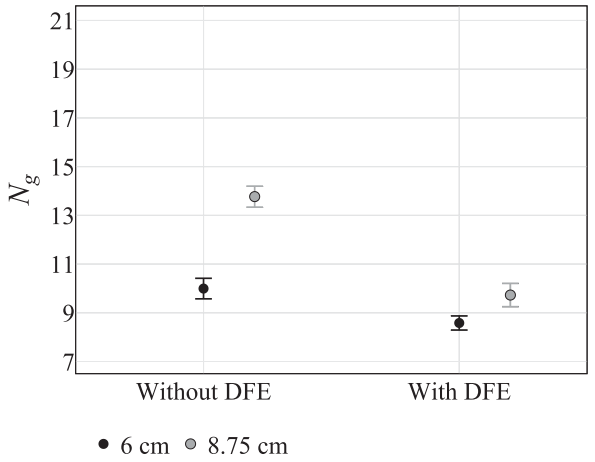


Fig. 7. N_g PSEs as a function of the DFE condition for SMAs with a radius of 6 and 8.75 cm along with the associated 95% confidence intervals.

Table 5. Bonferroni-corrected pairwise comparisons between the PSEs obtained with different DFE and radius conditions.

Case 1		Case 2			
Radius	DFE	Radius	DFE	Sig.p*	$d^{\dagger}$
6 cm	On	6.0 cm	Off	<0.001	0.77
8.75 cm	On	8.75 cm	Off	<0.001	1.8
6 cm	Off	8.75 cm	Off	<0.001	2.7
6 cm	On	8.75 cm	On	<0.001	0.62

Note. Sig.p is the p value of the post hoc test. d is the Cohen's d measure of the magnitude of the observed effect.
*The p value was adjusted with a Bonferroni correction accounting for four pairwise comparisons.

both with and without DFE, its effect was stronger without DFE ($d = 2.7$, depicting a large effect) than with DFE ($d = 0.62$, depicting a moderate effect).

4 DISCUSSION

The current study gathered PSEs, in terms of grid order N_g, beyond which aliasing-induced artifacts are no longer perceptible. It was expected that the more the SMA signals are likely to exhibit spatial aliasing, the higher PSEs should be for the resulting artifacts to become imperceptible.

4.1 SMA Properties

In this study, aliasing-induced artifacts were investigated for SMAs differing in radii (6 and 8.75 cm) and sampling schemes (extremal or Gauss-Legendre). For an SMA of a given grid order N_g, spatial aliasing artifacts are expected to increase with the SMA radius, because the distance between microphones increases accordingly. The results of the present study are in accordance with this expectation, as shown by the significant simple effect of the SMA radius (see Table 3) along with Fig. 7, which displays larger required N_g PSEs for the larger SMA radius (8.75 cm) than for the smaller one (6 cm).

The two sampling schemes were chosen so that the aliasing frequencies computed from Eq. (2) were equal for a given grid order N_g. However, the aliasing pattern according to which higher-than-order-N_g SH components are aliased into lower-order SH components depends on the sampling scheme [16]. This leads to different aliasing-induced artifacts for SMAs designed according to an extremal or Gauss-Legendre scheme. As shown in Fig. 2, a lower amount of energy is aliased in Gauss-Legendre grids than in extremal grids of same order. The results of the present study are in line with this expectation, because the significant simple effect of the sampling scheme, along with Fig. 6, shows that extremal SMAs led to overall higher PSEs than Gauss-Legendre SMAs.

However, there is a compromise between the robustness against spatial aliasing and the number of microphones required for building the SMA. For reaching a given grid order N_g, Gauss-Legendre grids require twice as many sampling points as extremal grids.

4.2 DFE

As depicted in Figs. 6 and 7, the PSEs are lower, whatever the SMA properties, when the encoding matrices are equalized with the DFE. Moreover, both effects of the radius and sampling scheme are weaker when the DFE is applied.

According to the results of the present study, equalizing the encoding matrix of an extremal SMA allows one to reach the performances of a similar SMA using a Gauss-Legendre sampling scheme, in terms of spatial aliasing, while using half as many microphones. As an example, the PSEs found in the current study when DFE was applied were $\bar{N}_g = 9.24$ for SMAs using an extremal sampling scheme and $\bar{N}_g = 9.07$ for SMAs using a Gauss-Legendre sampling scheme (see Fig. 6). As depicted in Table 1, constituting an extremal SMA of grid order $N_g = 9$ requires 100 microphones, whereas 200 microphones are required for a Gauss-Legendre SMA of same grid order.

Furthermore, while considering roughly the same number of sampling points, Lecomte et al. [22] showed a greater aliasing performance of a Lebedev sampling scheme in comparison with a spherical t-design and Fliege sampling scheme. It could be hypothesized from the results of the present study that such differences could be perceptually lessened by applying a DFE on the encoding matrices.

Additionally, the results of the present study show that the PSEs differ less across SMA radii when the encoding matrix is equalized with a DFE. The advantages of increasing an SMA radius are twofold: i) positioning the same number of microphones on a larger sphere is less difficult than on a smaller one, and ii) larger SMAs are more efficient for capturing the lower-frequency content of a given sound field than smaller SMAs, as explained in SEC. 0.2. Spatial aliasing is the main downside when trying to increase the SMA array size or reduce the number of sampling points. The DFE method investigated in the present study reduces spatial aliasing-induced artifacts, thus allowing the use of larger arrays constituted of fewer microphones, with fewer

repercussions on the binaural rendering of the recorded content.

Further studies should consider a DFE that erases the interchannel correlations between the ambisonic components (as proposed by Schörkhuber and Höldrich [28]) to investigate a further reduction in spatial aliasing artifacts. Moreover, the DFE was applied beyond the aliasing frequency, but other cut-on frequencies could be used depending on the quantity of aliased energy in the estimated components.

4.3 Using a Full-Resolution Reference

As hypothesized by Ahrens and Andersson [6], spatial aliasing is suspected to be a major reason why low-order ambisonic content can be distinguished from full-resolution references. Hence, the DFE method might be an effective way of reducing the minimum ambisonic order for a given rendering to be indistinguishable from a full-resolution reference.

In the current study, only the spatial aliasing error varied over the test conditions because the reference stimuli had the same ambisonic order as the test stimuli ($\tilde{N} = 7$). However, truncation-induced artifacts also occur when comparing SMA-based binaural signals to mannequin-head auralizations. In a perceptual evaluation where binaural renderings of SMA signals were compared to mannequin-head signals as references, Lübeck et al. [7] found slightly higher PSEs using SMAs with a radius of 8.75 cm ($\bar{N}_g \simeq 17 - 20$) than those obtained in the present study without DFE ($\bar{N}_g = 13.8$; see Fig. 7). Apart from the differently used stimuli and apparatus, the authors used a three-interval/two-alternative forced-choice protocol, which gives the participants a 50% chance of correct answer when picking a random answer, potentially leading to slightly higher PSEs than the three-interval/three-alternative forced-choice protocol used in the present study.

Another source of discrepancy may be the use of the MagLS method in the current study for the computation of the binaural filters, whereas Lübeck et al. used the virtual loudspeaker method for the binaural rendering. It can be hypothesized that the perceptual evaluation of Lübeck et al. performed using the same protocol and rendering method would lead to a lower PSE than the one they reported. Reaching analogous PSEs using a seventh-order reference or full-resolution reference would mean that the truncation error is small beyond the seventh order when using MagLS. This is suggested by the numerical evaluations of Engel et al. [43] and Schörkhuber et al. [12].

In a second experiment, Lübeck et al. [7] investigated the minimum ambisonic order required for truncation-induced artifacts to be imperceptible. They used sparse mannequin-head auralizations instead of SMA-based stimuli so that only truncation errors and no aliasing errors were included in the stimuli. They identified this minimum ambisonic order N_{min} to be comprised between 9 and 13. As pointed out by Lübeck et al., this perceptual threshold is rather conservative and could be lowered depending on the decoding methods used. Consequently, it can be hypothesized that if aliasing artifacts are not perceptible or are removed by DFE, an SMA that allows the estimation of SH components up to $N_g \geq N_{min}$ might not be perceptually different from a full-resolution reference.

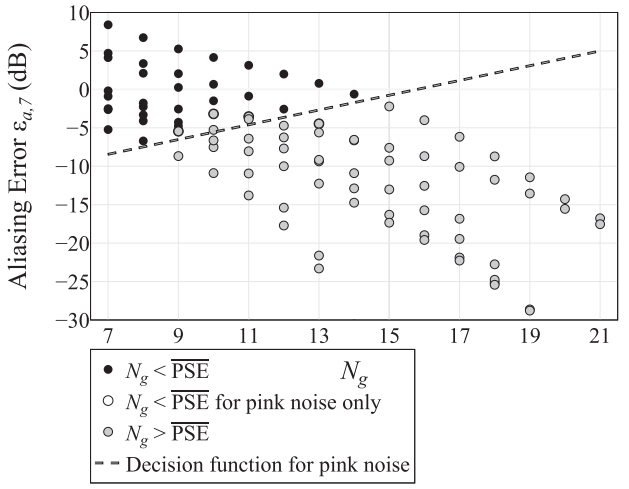


Fig. 8. Aliasing error $\epsilon_{a,7}$ as a function of the sampling grid order N_g for all the SMA configurations and DFE (on/off) comprised in the listening test.

4.4 Prediction of Audible Artifacts Based on the Aliasing Error

Fig. 8 shows the aliasing error $\epsilon_{a,7}$ defined in Eq. (11) with respect to the sampling grid order N_g. The aliasing error was averaged over frequency up to 20 kHz and averaged over 5,200 directions in accordance with a spherical t-design sampling scheme. The norms defined in Eq. (11) were computed using the ambisonic components of orders lower than the seventh order. For each N_g, the errors corresponding to all SMA configurations and DFE (on/off) are represented, giving a total of eight test conditions [2 sampling schemes × 2 radii × 2 DFE (on/off)]. For each of these test conditions, the mean N_g PSE over the subjects was computed. Two classes are represented: the class of stimuli for which the N_g was lower than the mean PSE and the one for which the N_g was higher. In other words, the stimuli of the first class are more likely to contain audible artifacts, whereas in the second class, artifacts are less likely to be audible. The first class had four more stimuli when considering the pink noise condition because it was a more critical content.

A linear support vector classification was used to obtain a decision function that would help to predict whether or not aliasing is more likely to be audible given a sampling grid order N_g and aliasing error $\epsilon_{a,\tilde{N}}$. The model was fitted over the classes related to the pink noise content. The decision function y_ϵ is visible in Fig. 8 and is defined by the following equation:

$$y_\epsilon = 0.96 \times N_g - 15.17 \approx N_g - 15 \text{ dB}. \tag{18}$$

This equation was obtained using a limited number of samples, and thus, no cross-validation was performed during the model fitting. A more robust decision function could be obtained using more SMA geometries as well as non-

spherical microphone arrays [44] in future listening evaluations.

However, it gives a simple criterion to estimate the perception of spatial aliasing artifacts. This equation relies on the aliasing error, which, contrary to the aliasing frequency f_a, takes into account the precise number of microphones and the sampling scheme of an SMA in addition to its radius. It makes the decision function of Eq. (19) a better predictor than the aliasing frequency.

This equation was derived using a critical sound source (i.e., pink noise). Audio contents that contain low energy at high frequency would not suffer as much from spatial aliasing. Thus, this equation needs to be put into perspective depending on the type of sound sources to be recorded. However, it is worth noting that the classification for the jazz quartet, which contains more common sound sources, does not appear to be very different from the other one (which is in line with the weak effect size of the sound source factor in the ANOVA). Moreover, the current study obtained PSEs using pairwise comparisons along with an adaptive procedure, which is rather conservative. It should be noted that quality-based comparisons (e.g., multiple stimulus with hidden reference and anchor procedures) might lead to lower N_g PSEs and that the PSEs found in the present study should cover for the most critical cases. Hence, the decision function y_ϵ that follows is then applicable for critical cases only.

4.5 Application to Loudspeaker Playback

It would be of interest to study the benefits of the DFE in the context of loudspeaker-based rendering. In such a scenario, loudspeaker placement, listening position, and room reflections can have an influence on the perceived differences between aliased and unaliased sound stimuli. As an example, the distribution of physical loudspeakers is often nonuniform, resulting in directional loudness and localization variations [45]. When rendering ambisonic scenes binaurally over headphones, preprocessing methods such as MagLS can be used to reduce perceptual artifacts resulting from spatial sampling in binaural filters. However, such rendering filters are not applicable for loudspeaker playback unless an alternative formulation for optimizing rendering filters is found in this context. Finally, aliasing errors increase outside the sweet spot, creating aliasing artifacts at the reconstruction stage [22] that are not present in a three-degrees-of-freedom binaural rendering context.

5 CONCLUSION

In this study, several recordings were simulated using virtual SMAs of different radii and sampling schemes to investigate the perception of spatial aliasing on binaural auralizations. A DFE that reduces the spectral coloration due to spatial aliasing was also investigated. The study used nonindividualized dynamic binaural rendering of seventh-order ambisonic signals using decoding filters optimized with the MagLS method.

As expected, a listening test showed that the perception of spatial aliasing artifacts increased with the SMA radius, decreased with the sampling grid order N_g, and depended on the sampling scheme. Additionally, it showed that the use of a DFE can significantly reduce the perception of spatial aliasing artifacts. This method made it possible to reach similar perceptual thresholds while considering larger radii and fewer microphones. This result can have implications for the design of ambisonic microphone arrays, because larger radii allow better performances in the lower end of the spectrum and the use of fewer microphones has practical and economical advantages.

A metric was introduced—the spatial aliasing error $\epsilon_{a,\tilde{N}}$—for predicting the perception of artifacts in $\tilde{N}$-order ambisonic signals depending on the SMA configuration. The results showed that spatial aliasing artifacts in seventh-order ambisonic scenes were less likely to be audible if the mean aliasing error averaged over frequency was lower than the quantity $y_\epsilon \approx N_g - 15$ dB. This value is rather conservative given the procedure employed in the listening test. This metric is a more reliable predictor than the aliasing frequency, because it takes more properties of the sampling grid into account.

6 ACKNOWLEDGMENT

This work was supported by the Region Bretagne under the HDAudio3D project. The authors thank the reviewers for their comprehensive comments, Nicolas Epain for his valuable insights on the ambisonic encoding and DFE, and all the subjects who took part in the subjective experiment.

7 REFERENCES

[1] B. Rafaely, "Acoustical Background," in *Fundamentals of Spherical Array Processing*, Springer Topics in Signal Processing, vol. 16., pp. 33–58 (Springer Nature Switzerland, Cham, Switzerland, 2019), 2nd ed. https://doi.org/10.1007/978-3-319-99561-8_2.

[2] B. Rafaely, "Analysis and Design of Spherical Microphone Arrays," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 135–143 (2005 Jan.). https://doi.org/10.1109/TSA.2004.839244.

[3] A. Avni, J. Ahrens, M. Geier, et al., â€œSpatial Perception of Sound Fields Recorded by Spherical Microphone Arrays With Varying Spatial Resolution," *J. Acoust. Soc. Am.*, vol. 133, no. 5, pp. 2711–2721 (2013 May). https://doi.org/10.1121/1.4795780.

[4] B. Bernschütz, A. V. Giner, C. Pörschmann, and J. Arend, "Binaural Reproduction of Plane Waves With Reduced Modal Order," *Acta Acust. united Acust.*, vol. 100, no. 5, pp. 972–983 (2014 Sep./Oct.). https://doi.org/10.3813/AAA.918777.

[5] B. Bernschütz, *Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording*, Ph.D. thesis, Technische Universitaet, Berlin, Germany (2016 Feb.). https://doi.org/10.14279/depositonce-5082.

[6] J. Ahrens and C. Andersson, "Perceptual Evaluation of Headphone Auralization of Rooms Captured With Spher-

ical Microphone Arrays With Respect to Spaciousness and Timbre," *J. Acoust. Soc. Am.*, vol. 145, no. 4, pp. 2783–2794 (2019 Apr.). https://doi.org/10.1121/1.5096164.

[7] T. Lübeck, J. M. Arend, and C. Pörschmann, "Binaural Reproduction of Dummy Head and Spherical Microphone Array Data—A Perceptual Study on the Minimum Required Spatial Resolution," *J. Acoust. Soc. Am.*, vol. 151, no. 1, pp. 467–483 (2022 Jan.). https://doi.org/10.1121/10.0009277.

[8] J.-M. Jot, V. Larcher, and J.-M. Pernaux, "A Comparative Study of 3-D Audio Encoding and Rendering Techniques," in *Proceedings of the 16th AES International Conference on Spatial Sound Reproduction* (1999 Mar.), paper 16-025.

[9] F. Brinkmann and S. Weinzierl, "Comparison of Head-Related Transfer Functions Pre-Processing Techniques for Spherical Harmonics Decomposition," in *Proceedings of the AES Conference on Audio for Virtual and Augmented Reality* (2018 Aug.), paper P9-3.

[10] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural Rendering of Ambisonic Signals by Head-Related Impulse Response Time Alignment and a Diffuseness Constraint," *J. Acoust. Soc. Am.*, vol. 143, no. 6, pp. 3616–3627 (2018 Jun.). https://doi.org/10.1121/1.5040489.

[11] C. Pörschmann, J. M. Arend, and F. Brinkmann, "Directional Equalization of Sparse Head-Related Transfer Function Sets for Spatial Upsampling," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 6, pp. 1060–1071 (2019 Jun.). https://doi.org/10.1109/TASLP.2019.2908057.

[12] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in *Proceedings of the German Annual Conference on Acoustics (DAGA)*, vol. 44, pp. 339–342 (Munich, Germany) (2018 Mar.).

[13] T. Lübeck, H. Helmholz, J. M. Arend, C. Pörschmann, and J. Ahrens, "Perceptual Evaluation of Mitigation Approaches of Impairments Due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data," *J. Audio Eng. Soc.*, vol. 68, no. 6, pp. 428–440 (2020 Jun.). https://doi.org/10.17743/jaes.2020.0038.

[14] J. M. Arend, F. Brinkmann, and C. Pörschmann, "Assessing Spherical Harmonics Interpolation of Time-Aligned Head-Related Transfer Functions," *J. Audio Eng. Soc.*, vol. 69, no. 1/2, pp. 104–117 (2021 Jan.). https://doi.org/10.17743/jaes.2020.0070.

[15] C. T. Jin, N. Epain, and A. Parthy, "Design, Optimization and Evaluation of a Dual-Radius Spherical Microphone Array," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 22, no. 1, pp. 193–204 (2014 Jan.). https://doi.org/10.1109/TASLP.2013.2286920.

[16] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial Aliasing in Spherical Microphone Arrays," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 1003–1010 (2007 Feb.). https://doi.org/10.1109/TSP.2006.888896.

[17] D. B. Ward and T. D. Abhayapala, "Reproduction of a Plane-Wave Sound Field Using an Array of Loudspeakers," *IEEE Trans. Speech Audio Process.*, vol. 9, no. 6, pp. 697–707 (2001 Sep.). https://doi.org/10.1109/89.943347.

[18] M. A. Poletti, "Three-Dimensional Surround Sound Systems Based on Spherical Harmonics," *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004–1025 (2005 Nov.).

[19] B. Rafaely, *Fundamentals of Spherical Array Processing*, Springer Topics in Signal Processing, vol. 16 (Springer Nature Switzerland, Cham, Switzerland, 2019), 2nd ed. https://doi.org/10.1007/978-3-319-99561-8.

[20] I. Bogaert, "Iteration-Free Computation of Gauss–Legendre Quadrature Nodes and Weights," *SIAM J. Sci. Comput.*, vol. 36, no. 3, pp. A1008–A1026 (2014 May). https://doi.org/10.1137/140954969.

[21] I. H. Sloan and R. S. Womersley, "Extremal Systems of Points and Numerical Integration on the Sphere," *Adv. Comput. Math.*, vol. 21, pp. 107–125 (2004 Jul.) https://doi.org/10.1023/B:ACOM.0000016428.25905.da.

[22] P. Lecomte, P.-A. Gauthier, C. Langrenne, A. Berry, and A. Garcia, "A Fifty-Node Lebedev Grid and Its Applications to Ambisonics," *J. Audio Eng. Soc.*, vol. 64, no. 11, pp. 868–881 (2016 Nov.). http://dx.doi.org/10.17743/jaes.2016.0036.

[23] M. A. Gerzon, "The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound," presented at the *50th Convention of the Audio Engineering Society* (1975 Mar.), paper L-20.

[24] A. Heller and E. M. Benjamin, "Calibration of Soundfield Microphones Using the Diffuse-Field Response," presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), paper 8711.

[25] T. McKenzie, D. Murphy, and G. Kearney, "Diffuse-Field Equalisation of Binaural Ambisonic Rendering," *Appl. Sci.*, vol. 8, no. 10, paper 1956 (2018 Oct.). https://doi.org/10.3390/app8101956.

[26] L. McCormack, A. Politis, R. Gonzalez, T. Lokki, and V. Pulkki, "Parametric Ambisonic Encoding of Arbitrary Microphone Arrays," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 30, pp. 2062–2075 (2022 Jun.). https://doi.org/10.1109/TASLP.2022.3182857.

[27] N. Epain and C. T. Jin, "Spherical Harmonic Signal Covariance and Sound Field Diffuseness," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 24, no. 10, pp. 1796–1807 (2016 Jun.). https://doi.org/10.1109/TASLP.2016.2585862.

[28] C. Schörkhuber and R. Höldrich, "Ambisonic Microphone Encoding With Covariance Constraint," in *Proceedings of the International Conference on Spatial Audio*, pp. 7–10 (Graz, Austria) (2017 Sep.).

[29] T. Deppisch, H. Helmholz, and J. Ahrens, "End-to-End Magnitude Least Squares Binaural Rendering of Spherical Microphone Array Signals," in *Proceedings of Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, pp. 1–7 (Bologna, Italy) (2021 Sep.). https://doi.org/10.1109/I3DA48870.2021.9610864.

[30] S. Moreau, *Étude et Réalisation d'Outils Avancés Dencodage Spatial Pour la Technique de Spatialisation Sonore Higher Order Ambisonics: Microphone 3D et Contrôle de Distance*, Ph.D. thesis, Université du Maine, Le Mans, France (2006 Jul.).

[31] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "Soft-Limiting der Modalen Amplituden-

verstärkung bei Sphärischen Mikrofonarrays im Plane Wave Decomposition Verfahren," in *Proceedings of the 37th German Annual Conference on Acoustics (DAGA)*, pp. 661–662 (Dusseldorf, Germany) (2011 Jan.).

[32] S. Moreau, J. Daniel, and S. Bertet, "3D Sound Field Recording With Higher Order Ambisonics–Objective Measurements and Validation of a 4th Order Spherical Microphone," presented at the *120th Convention of the Audio Engineering Society* (2006 May), paper 6857.

[33] P. Delsarte, J. M. Goethals, and J. J. Seidel, "Spherical Codes and Designs," in D .G. Corneil and R. Mathon (Eds.), *Geometry and Combinatorics: Selected Works of J. J. Seidel*, pp. 68–93 (Academic Press, San Diego, CA, 1991). https://doi.org/10.1016/B978-0-12-189420-7.50013-X.

[34] R. H. Hardin and N. J. Sloane, "McLaren's Improved Snub Cube and Other New Spherical Designs in Three Dimensions," *Discrete Comput. Geom.,* vol. 15, no. 4, pp. 429–441 (1996 Apr.). https://doi.org/10.1007/BF02711518.

[35] A. Wabnitz, N. Epain, C. Jin, and A. Van Schaik, "Room Acoustics Simulation for Multichannel Microphone Arrays," in *Proceedings of the International Symposium on Room Acoustics* (Melbourne, Australia) (2010 Aug.), paper P5d.

[36] D. Thery and B. Katz, "Anechoic Audio and 3D-Video Content Database of Small Ensemble Performances for Virtual Concerts," in *Proceedings of the 23rd International Congress on Acoustics (ICA)*, pp. 739–746 (Aachen, Germany) (2019 Sep.).

[37] Noise Makers, "Ambi Head HD," https://www.noisemakers.fr/ambi-head-hd/ (accessed Apr. 2, 2024).

[38] B. Bernschütz, "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," in *Proceedings of the 40th Italian Annual Conference on Acoustics (AIA) and 39th German Annual Conference on Acoustics (DAGA)*, vol. 29, pp. 592–595 (Merano, Italy) (2013 Mar.).

[39] Feichter Audio, "T3 Headphone Head-Tracker," http://feichter-audio.com/produits/diffusion/t3/ (accessed Apr. 2, 2024).

[40] T. S. Meese, "Using the Standard Staircase to Measure the Point of Subjective Equality: A Guide Based on Computer Simulations," *Percept. Psychophys.*, vol. 57, no. 3, pp. 267–281 (1995 Apr.). https://doi.org/10.3758/BF03213053.

[41] G. Berthomieu, F. Salmon, J. Palacino, and M. Paquier, "The Influence of the Number of Microphones Constituting a Spherical Microphone Array on the Perception of Spatial Aliasing," in *Proceedings of the 10th Convention of the European Acoustics Association, Forum Acusticum*, pp. 1149–1152 (Torino, Italy) (2023 Sep.). https://doi.org/10.61782/fa.2023.1077.

[42] R. Bender and S. Lange, "Adjusting for Multiple Testing—When and How?" *J. Clin. Epidemiol.*, vol. 54, no. 4, pp. 343–349 (2001 Apr.). https://doi.org/10.1016/S0895-4356(00)00314-0.

[43] I. Engel, D. F. Goodman, and L. Picinali, "Assessing HRTF Preprocessing Methods for Ambisonics Rendering Through Perceptual Models," *Acta Acust.*, vol. 6, paper 4 (2022 Jan.). https://doi.org/10.1051/aacus/2021055.

[44] J. Ahrens and Z. Hu, "Evaluation of Non-Spherical Scattering Bodies for Ambisonic Microphone Arrays," in *Proceedings of the AES International Conference on Audio for Virtual and Augmented Reality Conference* (2022 Aug.), paper 7.

[45] F. Zotter, M. Frank, and H. Pomberger, "Comparison of Energy-Preserving and All-Round Ambisonic Decoders," in *Proceedings of the International Conference on Acoustics AIA-DAGA* (Merano, Italy) (2013 Mar.), paper 412.

[46] A. Field, *Discovering Statistics Using IBM SPSS Statistics* (SAGE, Thousand Oaks, CA, 2018), 5th ed.

[47] S. Olejnik and J. Algina, "Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs." *Psychol. Methods*, vol. 8, no. 4, pp. 434–447 (2003 Dec.) https://doi.org/10.1037/1082-989X.8.4.434.

[48] H. M. Walker, "Degrees of Freedom," *J. Educ. Psychol.*, vol. 31, no. 4, pp. 253–269 (1940 Apr.). https://doi.org/10.1037/h0054588.

THE AUTHORS


François Salmon


Gauthier Berthomieu


Julian Palacino


Mathieu Paquier

François Salmon obtained an M.Sc. in electrical engineering from the Ecole Nationale Supérieure de l'Electronique et de ses Applications in 2013, followed by an M.Sc. in audio engineering from the Ecole Nationale Supérieure Louis-Lumière in 2016. Subsequently, he worked as a 3D audio engineer at the Applied Mathematics Research Laboratory of the Ecole Polytechnique before joining the Institute of Research and Technology b-com in Rennes, where he obtained a Ph.D. in acoustics from the University of Brest in 2021. Since then, he has worked as a research engineer in sound field recording and virtual acoustics at Noise Makers in Rennes, France.

•

Gauthier Berthomieu received an M.Sc. in sound engineering from the Image & Sound Department of the University of Brest in 2016 and then a Ph.D. in psychoacoustics from the University of Brest in 2019. He has been working as a postdoctoral fellow in the Sound Perception team of the Lab-STICC CNRS laboratory since 2022. His research interests are mainly related to spatial perception and the evaluation of 3D sound technologies.

•

Julian Palacino began his academic journey in sound engineering in Colombia. In 2008, he culminated his efforts by obtaining a master's degree in science from the University of Maine. After spending two years as an acoustic engineer, he transitioned into the field of doctoral studies in 3D audio at Orange Labs under the supervision of Rozenn Nicol. In 2015, Palacino joined the Psychoacoustics Laboratory at the University of Brest, focusing on research into the perception of 3D audio rendering. In 2017, his expertise earned him his current role at the helm of the Research and Development Audio Department at Feichter Audio, where he primarily focuses on advancing binaural and 3D audio systems.

•

Mathieu Paquier was born in Bordeaux, France, in 1974. He received an M.Sc. in sound engineering from the University of Brest in 1997 and then a Ph.D. in acoustics from the University of Lyon in 2002. He has been full professor and head of the Image and Sound Department of the University of Brest since 2016. He has directed the Sound Perception Team of the Lab-STICC CNRS laboratory since 2018, and his research interests mainly include spatial perception, evaluation of sound reproduction systems, music perception, and auditory risks related to sound reinforcement.