The Ring of Silence in Ambisonics: Spectral Impairments in Loudspeaker and Binaural Reproduction

Filippo Maria Fazi and Jacob Hollebon

***Abstract*—When a sound field is reproduced by an Ambisonic system with a number of loudspeakers greater than the minimum required by the order of the input Ambisonic signals, a ring-shaped region (in 2D) or spherical shell (in 3D) is generated wherein the average sound pressure level is lower than the desired target. This phenomenon, already observed by Solvang in 2008, is further analysed in this work and explained in terms of truncation error, spatial aliasing (related to the chosen Ambisonic decoder), and energy of the radial functions. It is then shown that the same phenomenon is the cause of high-frequency energy loss occurring in the binaural reproduction of Ambisonic material when the HRTF is sampled with a high-density grid, as already observed by Bernschütz et al.**

***Index Terms*—Spatial Audio, Higher Order Ambisonics (HOA), Binaural Audio, HRTF, Spatial Aliasing.**

I. Introduction

HIGHER-Order Ambisonics (HOA) is an increasingly popular and flexible spatial audio technique, covering a full end-to-end audio capture, transmission and reproduction approach [1]. HOA relies on the representation of the sound field by means of a basis of orthogonal functions; in 3D these functions are the spherical harmonics [2], whilst in 2D they are the circular harmonics [3], [4]. The coefficients for the series expansion are referred to as Higher Order Ambisonics (HOA) signals, which are thus a scene-based representation of the sound field. In practice, the otherwise infinite expansions must be truncated to a finite order N, which leads to an error bound of validity for the representation. This follows the well-known $N = kr$ rule, where k is the wavenumber and r is the radial distance from the origin of the expansion [5].

A number of methods exist for reproduction of the HOA signals using either loudspeakers or headphones (binaural reproduction), although many similarities exist between these two categories. The reproduction over loudspeakers requires the definition of a *decoder* that processes the HOA input to generate a set of loudspeaker signals, which may be played back directly over a loudspeaker array. The problem of defining suitable decoders has been the topic of research for a considerable period of time. One commonly used approach for both loudspeaker arrays and binaural reproduction is the classic *Mode-Matching Decoder* [1] (also known as pseudoinverse decoder, and fundamentally equivalent to the *basic* decoder introduced in [6]), that utilises a pseudoinverse approach to ensure the expansion coefficients of the reproduced sound field match those of the target sound field (defined by the input HOA signals) [1, p.73] [2].

However, an issue arises with this decoder when considering the relationship between the number of reproduction loudspeakers L and the truncation order N. For an exact solution (that is the expansion coefficients of the reproduced sound field exactly match those of input HOA signals up to order N) the condition $L \geq (2N + 1)$ for the 2D scenario, and $L \geq (N + 1)^2$ for 3D, must be satisfied. In 2008 Solvang published an excellent piece of work that analysed the "spectral impairment" in Ambisonics reproduction using a mode-matching decoder for the 2D case [7]. The issue of spectral impairment emerges from using more loudspeakers than the minimum required for an exact solution. Solvang recognised three regions when using $L > (2N + 1)$:

- $kr < N$: the sound field is reproduced accurately.
- $N < kr < L - N$: spectral impairment will occur and the sound intensity will be reduced.
- $L - N < kr$: spatial aliasing will occur and cause a different kind of spectral impairment.

In the region where $N < kr < L - N$ the SPL drops due to the mismatch between the order of the input HOA signals and that maximal controllable order of the loudspeaker array. In the present work, this region is informally referred to as the *ring of silence*. Whilst not strictly silence, the name is coined from the fact that in this region the sound field is dominated by the reproduction of orders $N \to L - N$, whose energy is minimised by the decoder, as will be shown later.

A similar issue has been noted in the literature when using the psuedoinverse approach for binaural reproduction of HOA signals, for example in the work by Bernschütz et al. [8]. It is assumed that the Head-Related Transfer Function (HRTF) is sampled using a finite but dense grid of source positions. The binaural encoding process can be interpreted either as the decoding of the HOA signals to a virtual loudspeaker array (where the transducer positions correspond to the HRTF grid points) [9], [10], or as the convolution of the HOA signals with the spherical harmonic coefficients of the HRTF (estimated from the measurements on the grid discussed above) [10], [11]. It is also shown in this case that, when the HOA signal order is lower than that used for the HRTF representation, a reduction of energy in the high-frequency region is observed.

The mitigation approach to minimise this high frequency roll-off suggested by Bernschütz et al. is to spatially downsam-

Manuscript received December 12, 2024

[1]For special sampling schemes such as uniform sampling of a circle, or t-design sampling scheme of a sphere, the mode-matching decoder is equivalent to the sampling decoder, defined in [1, Sec. 4.9.1]

ple the HRTF to a lower density grid, that uses the minimum number of sampling points L matching the input HOA order ($N = (L+1)^2$ in this 3D case). This introduces spatial aliasing above order N. Whilst this is a form of spectral impairment itself due to incorrect reproduction of the HRTF, it is perceptually less problematic in terms of SPL than the alternative of a high frequency roll-off. A further problem with this approach is that the additional information from the dense sampling of the HRTF is discarded.

The high-frequency roll-off issue caused by order truncation is also analysed in the work by Ben-Hur et al. [12], wherein the authors provide a thorough analysis of the spatial aliasing artefacts and propose a solution based on diffuse-field equalisation. The roll-off depends on the direction of the virtual sound source and more strongly affects sources on the mid-sagittal plane. As illustrated, for example, in Fig. 5a below, the reproduced field energy is lowest along the direction perpendicular to the plane-wave propagation, corresponding to the listener's interaural axis when the source is in front of or behind the listener. This also implies that the high-frequency roll-off is less problematic for sources positioned to the far left or right of the listener.

Beyond the Mode-Matching Decoder, several other Ambisonics decoding approaches have been proposed in the literature. For Ambisonic audio reproduction over loudspeakers, these include HOA weighting schemes like the max-$\boldsymbol{r}_E$ and *in-phase* decoders [6], the Energy-Preserving Decoder [13], the All-Round Decoder [14], and data-dependent decoders based on sparse recovery strategies [15, p.49].

Specifically for binaural reproduction, methods have been developed to address the high-frequency roll-off while leveraging densely sampled HRTFs, as well as to mitigate the effects of spatial aliasing [16]–[18]. These approaches typically preprocess the phase of the HRTF data prior to spherical harmonic coefficient estimation. This is often based on the psychoacoustic assumption that high-frequency localisation is primarily determined by the HRTF magnitude, allowing the phase to be ignored or simplified at high frequencies. While this leads to incorrect phase reproduction at those frequencies, it improves the reproduction of magnitude and spatial fidelity. Consequently, such methods reduce the required spherical harmonic order by effectively compressing the relevant information into lower-order coefficients. Such approaches often perform time alignment of the HRTFs [19], [20], alter the phase, as in the Mag-LS approach [21], [22], remove information based on model HRTFs [23], or apply other HRTF preprocessing strategies [23]–[26].

In this work, a thorough mathematical analysis of the ring of silence is presented, focussing specifically on the Mode-Matching Decoder. The main contributions are:

- A mathematical formulation of the ring of silence is derived, generalising the expression proposed by Solvang for circular arrays of uniformly spaced loudspeakers [7] to more general two-dimensional and three-dimensional loudspeaker layouts. We explain the underlying causes of this phenomenon and demonstrate its connection to the pseudoinverse HOA decoder. Furthermore, we derive analytical expressions for the bounds of the ring of silence, which are exact for orthogonal loudspeaker configurations and approximate for sufficiently uniform non-orthogonal layouts. In the special case of a uniform circular array, our results reduce to those originally presented by Solvang.
- The expression of the binaural signals due to the HRTF-based rendering of a HOA signal is derived in the second part of this work. The proposed formulation shows that the high-frequency roll-off is equivalent to the ring of silence presented in the first part of the paper, even though the radial functions are different. This provides the mathematical explanation of the phenomenon observed by Bernschütz et al. [8].
- The consistent formulation used for the ring of silence for loudspeaker reproduction (first point above) and for binaural reproduction (second point) demonstrates that the HOA spectral impairment and the high-frequency roll-off observed in binaural reproduction are in fact due to the same causes, related to spatial aliasing and to the choice of a minimum-norm solution for the HOA decoder (mode-matching decoder) and for the estimation of the HRTF spherical harmonic coefficients.
- The results derived analytically are validated by experimental measurements carried out in an anechoic environment.
- Python code is provided to reproduce the results in this work[2].

The paper is organised as follows: the second section introduces the mathematical tools used in this work to describe sound fields and HOA. The concept of the ring of silence is presented in the following section. The roles of the HOA order, of the loudspeaker layout and of the spatial aliasing matrix are analysed, and lead to the formulation of the inner and outer bound of the ring of silence. The results are demonstrated by means of numerical simulations and experimental measurements. In the third section two methods for binaural reproduction of HOA material are presented and shown to be equivalent. In the light of this result, the ring of silence is investigated in the context of binaural audio reproduction with the support of numerical simulations based on experimental data. The conclusions of this work are drawn in the final section, together with some suggestions for future work.

II. Sound Field Model

We assume the sound field to be reproduced can be represented as a superposition of plane waves. In the frequency domain, with a $e^{j\omega t}$ time dependence, the field at a point $\boldsymbol{r}$ is given by [27, p. 46]

$$p(\boldsymbol{r}) = \int_{\hat{\boldsymbol{y}} \in S} e^{jk\boldsymbol{r}\cdot\hat{\boldsymbol{y}}}\, q(\hat{\boldsymbol{y}})\, dS \tag{1}$$

where k is the wave number and $\hat{\boldsymbol{y}}$ is the plane wave direction of arrival, identified by a point on S. The latter is the $\mathbb{S}^2$ unitary sphere for the three-dimensional case (3D) and the $\mathbb{S}^1$ unitary sphere (i.e. a circle with unitary radius) for the two-dimensional case (2D). $\hat{\boldsymbol{y}}$ is therefore identified by a single angle ϕ in 2D and by the elevation angle $\theta \in [-\pi/2, \pi/2]$

[2]https://github.com/jacobhollebon/ring-of-silence

and azimuth angle $\phi \in [0, 2\pi)$ in 3D. $e^{jk\boldsymbol{r}\cdot\hat{\boldsymbol{y}}}$ indicates the field due to a plane wave with unitary amplitude and arriving from direction $\hat{\boldsymbol{y}}$. $q(\hat{\boldsymbol{y}})$ is the plane wave density function (PWD), defined over S. The PWD can be represented in terms of a (generalised) Fourier series. In 2D this is

$$q(\phi) = \sum_{n=0}^{\infty} \sum_{m=0}^{1} \Psi_{nm}(\phi) b_{nm} \tag{2}$$

with

$$\Psi_{nm}(\phi) = \begin{cases} \cos(n\phi)/\sqrt{\pi} & \text{if } n > 0 \text{ and } m = 0 \\ \sin(n\phi)/\sqrt{\pi} & \text{if } n > 0 \text{ and } m = 1 \\ (1-m)/\sqrt{2\pi} & \text{if } n = 0 \end{cases} \tag{3}$$

whilst in 3D

$$q(\theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \Psi_{nm}(\theta, \phi) b_{nm} \tag{4}$$

with $\Psi_{nm}(\theta, \phi) = Y_n^m(\theta, \phi)$, the spherical harmonics, which may be defined using a real or complex representation [1], [28]. Hereafter we will use the following compact notation that encompasses both the 2D and 3D case:

$$q(\hat{\boldsymbol{y}}) = \sum_{n=0}^{\infty} \sum_{m} \Psi_{nm}(\hat{\boldsymbol{y}}) b_{nm} \tag{5}$$

wherein the limits of the second sum are implicitly defined and depend on whether the field is defined in 2D or 3D. The functions Ψ_{nm} are orthonormal on their domain of definition and are hereafter also referred to as the basis functions. The subscript nm can be treated as a single index when using it to define elements of a matrix or of a vector. In 3D, this indexing follows the common Ambisonics Channel Number (ACN) ordering [1, p. 68]. The coefficients b_{nm} are defined by the well-known Fourier analysis equation [28, p. 192]

$$b_{nm} = \int_{\hat{\boldsymbol{y}} \in S} \Psi_{nm}^*(\hat{\boldsymbol{y}}) q(\hat{\boldsymbol{y}}) dS \tag{6}$$

where $(.)^*$ is the conjugate operator. b_{nm} are therefore the spatial Fourier series coefficients that describe the PWD.

The sound field $p(\boldsymbol{r})$ can also be expressed in terms of the basis functions Ψ_{nm}. To that end, we use the Jacobi-Anger expansion of a plane wave [29, p. 32], that is

$$e^{jk\boldsymbol{r}\cdot\hat{\boldsymbol{y}}} = \sum_{n=0}^{\infty} \sum_{m} R_n(kr) \Psi_{nm}^*(\hat{\boldsymbol{y}}) \Psi_{nm}(\hat{\boldsymbol{r}}) \tag{7}$$

where $r = ||\boldsymbol{r}||$, $\hat{\boldsymbol{r}} = \boldsymbol{r}/r$, and R_n are the so-called radial functions, given by

$$R_n(kr) = \begin{cases} j^n \, 2\pi \, J_n(kr) & \text{in 2D} \\ j^n \, 4\pi \, j_n(kr) & \text{in 3D} \end{cases} \tag{8}$$

where $J_n(kr)$ and $j_n(kr)$ are Bessel functions and spherical Bessel functions, respectively. Combining this equation with (1) and using the orthogonality of the basis functions yields

$$p(\boldsymbol{r}) = \sum_{n=0}^{\infty} \sum_{m} R_n(kr) \Psi_{nm}(\hat{\boldsymbol{r}}) b_{nm}. \tag{9}$$

It is noteworthy that each spatial Fourier coefficient of the PWD, b_{nm}, affects one and only one of the sum terms of equation (9). This means that the (generalised) spatial Fourier transform diagonalises the plane wave expansion integral (1). This can be regarded as an extension of the well-known convolution theorem. Plots of the radial functions are available in the literature, for example [28, pp. 119, 195].

The energy of the radial functions of different orders is greater around the first maximum of the function, occurring when $n \approx kr$. This spatial selectivity of the radial functions implies that low order contribute mainly to the field energy for small kr values, and higher order to greater values of kr. This will be crucial to explain the existence of the ring of silence.

A. Target Sound Field and Truncation Error

In the context of sound field reproduction we define the target sound field, p_T, and the reproduced field p_R. Both fields can be defined in terms of the spatial Fourier series coefficients of the corresponding PWD.

We define the HOA signals (in the frequency domain) as the coefficients b_{nm} that describe the target sound field and that are fed as input to the sound reproduction system. The number of HOA signals must be finite for obvious practical reasons, meaning that the target field can only be described by a finite amount of information. The choice is usually made to represent the target sound field with all coefficients b_{nm} up to a maximum order N, which defines the HOA order, above which the HOA signals are not defined.

Assuming that the original (subscript O) sound field is given by an infinite series of the form

$$p_O(\boldsymbol{r}) = \sum_{n=0}^{\infty} \sum_{m} R_n(kr) \Psi_{nm}(\hat{\boldsymbol{r}}) b_{nm} \tag{10}$$

its representation p_T in terms of a finite number of HOA signals provides a truncated series, given by

$$p_T(\boldsymbol{r}) = \sum_{n=0}^{N} \sum_{m} R_n(kr) \Psi_{nm}(\hat{\boldsymbol{r}}) b_{nm} = p_O(\boldsymbol{r}) - \epsilon_T(\boldsymbol{r}) \tag{11}$$

where ϵ_T is the *truncation error*

$$\epsilon_T(\boldsymbol{r}) = \sum_{n=N+1}^{\infty} \sum_{m} R_n(kr) \Psi_{nm}(\hat{\boldsymbol{r}}) b_{nm}. \tag{12}$$

As the radial functions $R_n(kr)$ are proportional to Bessel functions, the truncation error is small for $kr < N$, which is a well-known rule-of-thumb for HOA [5]. The energy, and hence the SPL, of p_T drops when $kr > N$ as a consequence of the truncation error (see also [12]). This generates the inner boundary of the ring of silence, due to the fact that the number of HOA signals is limited. It can also be simply deduced that the averaged energy of the truncated field, $E_T(r)$, is the energy of the original field minus the energy of the truncation error.

B. Reproduced Sound Field

The Fourier coefficients of the reproduced field are denoted by the symbol $\tilde{b}_{nm}$ and should ideally match to the corresponding target field coefficients b_{nm}, but will in general be

different. For a loudspeaker array with L transducers, where the sound field generated by the ℓ-th loudspeaker can be modelled as a single plane wave arriving from direction $\hat{\boldsymbol{y}}_\ell$ and driven by signal g_ℓ, the plane wave density function for the reproduced field is given by

$$q(\hat{\boldsymbol{y}}) = \sum_{n=0}^{\infty} \sum_{m} \Psi_{nm}(\hat{\boldsymbol{y}}) \tilde{b}_{nm} = \sum_{\ell=1}^{L} \delta(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_\ell)\, g_\ell \qquad (13)$$

where $\delta(\hat{\boldsymbol{y}})$ is the Dirac delta. Combining the equation above with (6) we obtain

$$\tilde{b}_{nm} = \int_S \sum_{\ell=1}^{L} \delta(\hat{\boldsymbol{y}} - \hat{\boldsymbol{y}}_\ell)\, g_\ell \Psi^*_{nm}(\hat{\boldsymbol{y}}) dS = \sum_{\ell=1}^{L} g_\ell \Psi^*_{nm}(\hat{\boldsymbol{y}}_\ell). \qquad (14)$$

In the case of HOA, the loudspeaker signals, vector $\boldsymbol{g}$, are given by the product of the HOA signal vector $\boldsymbol{b}$ of the target sound field with a decoding matrix $\boldsymbol{D}$, namely

$$g_\ell = \sum_{n=0}^{N} \sum_{m} b_{nm} D_{\ell,nm} \qquad (15)$$

where N is the HOA signal order used for the representation. In vector notation this is

$$\boldsymbol{g} = \boldsymbol{D}\boldsymbol{b} \qquad (16)$$

The PWD Fourier coefficients of the reproduced field are given by combing the results above as

$$\tilde{b}_{nm} = \sum_{\nu=0}^{N} \sum_{\mu} A_{nm,\nu\mu} b_{\nu\mu}. \qquad (17)$$

or, with matrix notation, $\tilde{\boldsymbol{b}} = \boldsymbol{A}\boldsymbol{b}$. $A_{nm,\nu\mu}$ is an element of the $\infty \times \mathcal{N}$ spatial aliasing matrix $\boldsymbol{A}$, defined as

$$A_{nm,\nu\mu} = \sum_{\ell=1}^{L} \Psi^*_{nm}(\hat{\boldsymbol{y}}_\ell) D_{\ell,\nu\mu}. \qquad (18)$$

$\mathcal{N}$ is the total number of HOA signals, which is $2N+1$ for the 2D case and $(N+1)^2$ for 3D. Matrix $\boldsymbol{A}$ will play an important role in what follows. The coefficients $\tilde{b}_{nm}$ are the aliased version of the HOA signals b_{nm}. In an ideal case we would have that $\tilde{b}_{nm} = b_{nm}$. However, whilst any practical system will always deal with a finite number of HOA signals b_{nm}, the Fourier coefficients representing the reproduced field, $\tilde{b}_{nm}$, are not bounded to a finite order because the loudspeaker distribution is not continuous and therefore the PWD is not a smooth function.

As above, combining equations (1) and (7), and using the orthogonality property of the basis functions Ψ_{nm} we obtain

$$p_R(\boldsymbol{r}) = \sum_{n=0}^{\infty} \sum_{m} R_n(kr) \Psi_{nm}(\hat{\boldsymbol{r}}) \tilde{b}_{nm}. \qquad (19)$$

The aliasing error is

$$\epsilon_A(\boldsymbol{r}) = \sum_{n=0}^{\infty} \sum_{m} R_n(kr) \Psi_{nm}(\hat{\boldsymbol{r}}) \sum_{\nu=0}^{N} \sum_{\mu} (\delta_{nm,\nu\mu} - A_{nm,\nu\mu}) b_{\nu\mu} \qquad (20)$$

where $\delta_{nm,\nu\mu}$ is the Kronecker delta. Combining all the results above we obtain the expression of the reproduced field

$$p_R(\boldsymbol{r}) = p_O(\boldsymbol{r}) - \epsilon_T(\boldsymbol{r}) - \epsilon_A(\boldsymbol{r}). \qquad (21)$$

III. Aliasing Error and the Ring of Silence

We will now study the effect of the structure of the spatial aliasing error. This depends on the spatial aliasing matrix $\boldsymbol{A}$, which is in turn defined by the number and arrangement of the loudspeakers and by the Ambisonics decoding matrix $\boldsymbol{D}$. We start by defining matrix $\boldsymbol{Y}$, whose columns are the basis functions Ψ_{nm} evaluated at the L loudspeaker positions. Thus, for the 3D case,

$$\boldsymbol{Y}_N = \begin{bmatrix} Y_{00}(\theta_1, \phi_1) & \dots & Y_{NN}(\theta_1, \phi_1) \\ \vdots & \ddots & \vdots \\ Y_{00}(\theta_L, \phi_L) & \dots & Y_{NN}(\theta_L, \phi_L) \end{bmatrix}. \qquad (22)$$

An analogous definition holds for the 2D case. The subscript N indicates the maximum order of the basis functions included in the matrix.

We make two important simplifying assumptions on the decoding matrix $\boldsymbol{D}$:

1) The number of HOA signals is smaller than, or equal to the number of loudspeakers L. That is $L \geq (2N+1)$ or $L \geq (N+1)^2$ in 2D and 3D, respectively.
2) The Ambisonics decoder is a Mode-Matching Decoder, designed as the pseudoinverse of matrix $\boldsymbol{Y}_N^H$, namely

$$\boldsymbol{D} = (\boldsymbol{Y}_N^H)^\dagger := \boldsymbol{Y}_N (\boldsymbol{Y}_N^H \boldsymbol{Y}_N)^{-1}. \qquad (23)$$

This is consistent with the definition provided in reference [1, Sec. 4.9.2], apart from a normalisation factor [3]. The choice was made not to include any regularisation, for the sake of simplicity.

Crucially, if the two assumptions above are satisfied, the decoder equation corresponds to the minimum-norm solution of an under-determined system of linear equations. In practice, this means that the decoder will ensure the array will

1) provide an accurate reconstruction of the terms b_{nm} up to order N and, therefore, of the corresponding terms of equation (9);
2) minimise the energy of the higher order components of the PWD and of the sound field.

This means that, in 3D, the Mode-Matching Decoder will minimise the energy of $L-(N+1)^2$ linear combinations of basis functions of order greater than N. These are the *degrees of freedom* of the system that the array could control but is not asked to control. As a consequence of the minimum norm solution, the energy of these unused degrees of freedom is forced to zero. Mathematically speaking, the decoder will generate a PWD that has zero projection on the null-space of matrix $\boldsymbol{Y}_N^H$, which is indeed a space of dimension $L-(N+1)^2$ (in 3D, and assuming the matrix is full-rank) [4]. This concept is the fundamental cause of the ring of silence.

It should be noted that the above considerations would no longer be valid if a different decoder design strategy were used,

[3]The apparent difference between eq.(23) and the correspodning definition in [1, Eq.(4.40)] is due to the different definition of matrix $\boldsymbol{Y}_N$. The one used in this work is consistent with that provided in [27, Eq.(3.29)]

[4]The nullspace of $\boldsymbol{Y}_N^H$ is given by all possible combinations $\boldsymbol{g}_0$ of loudspeaker signals that generate a sound field with no energy in the coefficients $\tilde{b}_{nm}$ corresponding to the first N orders, i.e. $\boldsymbol{Y}_N^H \boldsymbol{g}_0 = 0$.

for example approaches that promote sparsity of the solution (such as compressive sensing) instead of its energy.

We have established that the Ambisonic Mode-Matching Decoder will minimise the energy of some components of the reproduced sound field, associated to the nullspace of $\boldsymbol{Y}_N^H$. We now want to demonstrate that most of the energy of these components is contained in a bounded region of space (indeed, the ring of silence). In order to do that, we shall investigate the representation of the nullspace of $\boldsymbol{Y}_N^H$ in terms of the basis functions Ψ_{nm}. A rigorous way to achieve that would be to calculate a basis for the nullspace and then compute the spatial Fourier analysis, equation (6), of each basis vector. As exciting as this analysis may be, it is not part of this work and is recommended as a task for future research. We will instead focus the analysis on the simple but representative case of orthogonal sampling schemes, and will then demonstrate numerically that the conclusions extend qualitatively to more general scenarios.

We define the orthogonality matrix as

$$\boldsymbol{B} = \boldsymbol{Y}_\infty^H \boldsymbol{Y}_N. \tag{24}$$

Each element of this matrix defines the inner product between two vectors corresponding to a pair of basis functions $\Psi_{nm,\nu\mu}$ sampled at the loudspeaker positions $\hat{\boldsymbol{y}}_\ell$. We then consider simple cases of loudspeaker arrangements that ensure the sampled basis functions Ψ_{nm} are orthogonal up to a given order. Under these conditions, and using the N3D convention for the spherical harmonic functions [1, p. 120],

$$B_{nm,\nu\mu} = \begin{cases} L/(\alpha 2\pi) & \text{if } nm = \nu\mu \text{ and } n \leq N_{max} \\ 0 & \text{if } nm \neq \nu\mu \text{ and } n \leq N_{max} \\ \beta_{nm,\nu\mu} & \text{if } n > N_{max} \end{cases} \tag{25}$$

where $\alpha = 1$ in 2D and $\alpha = 2$ in 3D, $\beta_{nm,\nu\mu}$ are complex scalars and N_{max} a natural number that depends on the array and on the HOA order N. N_{max} is the order above which higher-order basis functions are aliased down into lower orders. Examples of these orthogonal sampling schemes in 2D are circular arrays with $L \geq 2N+1$ loudspeakers and uniform angular separation (UCA) or, in 3D, a spherical t-design array [27, p. 68].

For a UCA matrix Y_N is strictly related to the Discrete-Time Fourier Transform (DTFT) [30]. Because of the periodicity of the Fourier coefficients, it holds that $N_{max} = L - N - 1$. For the 3D t-design, the inner product of two sampled spherical harmonics of order n and n', respectively, satisfies the orthogonality relation if $n + n' \leq t$ [31]. For this reason, $N_{max} = t - N$. This theoretical result is confirmed by numerical simulations.

We then assume that $N < N_{max}$ (where N is the order of the decoder). Then the spatial aliasing error matrix is

$$\boldsymbol{A} = \boldsymbol{Y}_\infty^H \boldsymbol{D} = \boldsymbol{Y}_\infty^H \boldsymbol{Y}_N (\boldsymbol{Y}_N^H \boldsymbol{Y}_N)^{-1}. \tag{26}$$

For orthogonal sampling schemes such as UCA's or t-designs with suitably large L, we see that $\boldsymbol{Y}_N^H \boldsymbol{Y}_N = L/(\alpha 2\pi)\mathbf{I}$, where $\mathbf{I}$ is the identity matrix, and the aliasing matrix is thereofore proportional to the orthogonality matrix, namely $\boldsymbol{A} = \boldsymbol{B}\ \alpha 2\pi/L$.

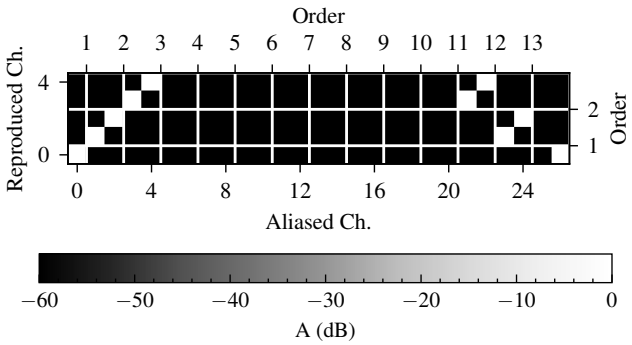

Fig. 1: Aliasing matrix for a Uniform Circular Array (UCA) with $L = 13, N = 2, N_{max} = 10$.

This results indicates that the spatial aliasing matrix can be subdivided in three sub-matrices:

1) The first $\mathcal{N}$ rows, up to order N ($2N + 1$ rows in 2D and $(N + 1)^2$ in 3D), define an identity matrix.
2) The rows from order $N+1$ to order N_{max} are all zeros. This is the central result that explains the existence of the ring of silence.
3) Rows corresponding to $n > N_{max}$ define spatial aliasing relations with a more complex pattern, caused by the lack of orthogonality of the sampled basis functions Ψ_{nm} of order $n \leq N$ with sampled basis functions of order $n > N_{max}$.

These three regions of matrix $\boldsymbol{A}$ can be clearly observed in Figs. 1 and 2, illustrating examples of spatial aliasing matrices for a UCA with $L = 13$ (Fig. 1) and for a spherical t-design [27] with $t = 12$ and $L = 84$ (Fig. 2a), all with $N = 2$. Note that the matrices have been rotated by $90°$ counterclockwise for better visualisation.

As a consequence of the structure of the spatial aliasing matrix, equation (17) yields

$$\tilde{b}_{nm} = \begin{cases} b_{nm} & \text{if } n \leq N \\ 0 & \text{if } N < n \leq N_{max} \\ \sum_{\nu=0}^{N} \sum_{\mu} A_{nm,\nu\mu} b_{\nu\mu} & \text{if } n > N_{max} \end{cases}. \tag{27}$$

This means that the sound field components of the reproduced field will be identical to those of the target field up to order N, will thereafter be zero up to order N_{max}, and finally above order N_{max} they will be aliased versions of the first N orders.

These three regions of matrix $\boldsymbol{A}$ are rigorously defined for orthogonal sampling schemes. Fig. 2b represents matrix $\boldsymbol{A}$ for a spherical equal area sampling scheme with $L = 49$ and $N = 2$ [32]. This configuration is nearly uniform but not orthogonal. It can be clearly observed that the first region ($n \leq N$) is well defined. However, since N_{max} is not defined for this sampling scheme, there is no sharp boundary between the second and third regions of the ring. Instead, a gradual onset of spatial aliasing artefacts is observed as n increases. This behaviour highlights the critical role of the sampling scheme in the occurrence of the ring of silence and suggests the potential for designing specialised sampling strategies to minimise this phenomenon.

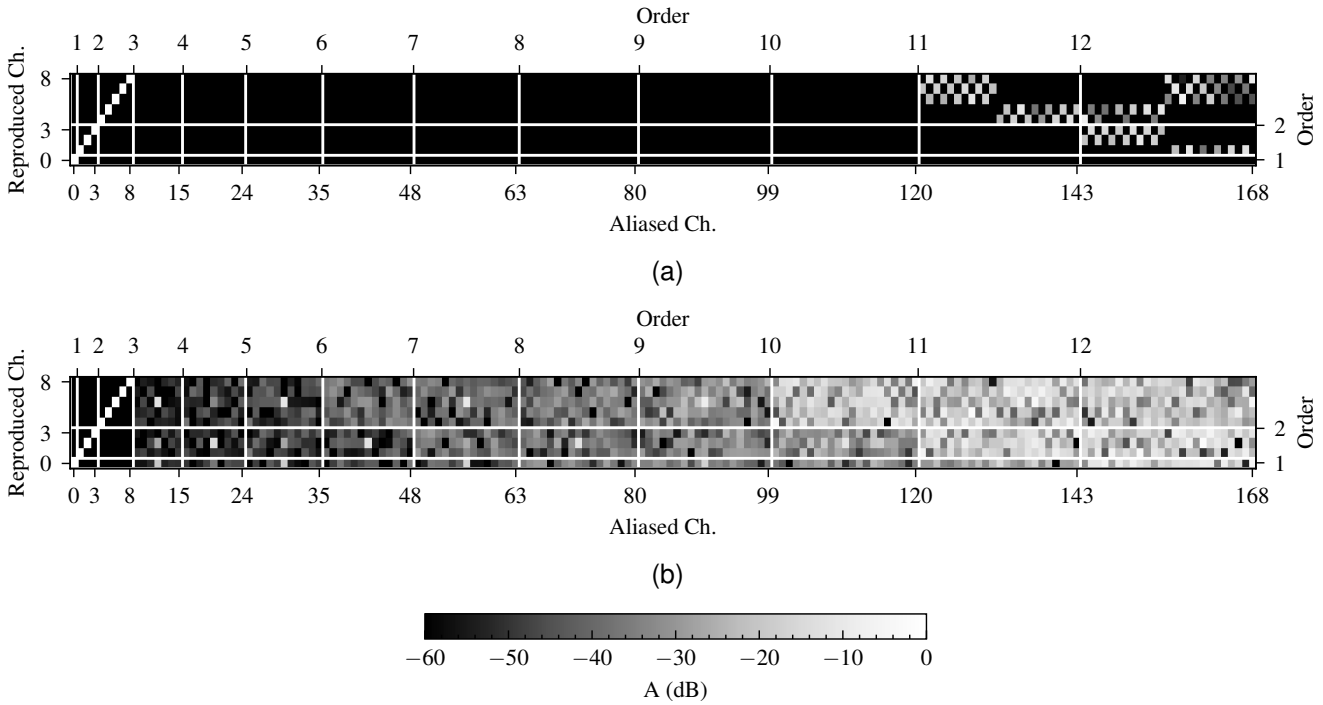

Fig. 2: Aliasing matrices for (a) Spherical t-design sampling with $t = 12,\ L = 84,\ N = 2,\ N_{max} = 10$ (b) Equal area spherical sampling with $L = 49,\ N = 2$.

We now introduce the expression of the sound field energy averaged over a circle or sphere with radius r, which is

$$\begin{aligned} E(kr) &= \frac{1}{\alpha 2\pi r^{\alpha}} \int_{\hat{\mathbf{r}} \in S} |p(\hat{\mathbf{r}})|^2 r^{\alpha} dS \\ &= \frac{1}{\alpha 2\pi} \sum_{n=0}^{\infty} \sum_{m} |R_n(kr) b_{nm}|^2 \end{aligned} \tag{28}$$

The above formula is a result of the orthogonality of the basis functions and is, in fact, the Parseval relation. The Fourier coefficients appearing in the average energy formula will be $\tilde{b}_{nm}$ in the case of the reproduced field.

In light of this formula, of the spatial selectivity of the radial functions $R_n(kr)$ explained above, and of the result illustrated by equation (27), we deduce that the reproduced sound field will have an energy dip in the kr interval dominated by the contribution of the orders $N < n \leq N_{max}$. This is the ring of silence.

Fig. 3 shows the average energy of the sound field reproduced by the same loudspeaker arrays as presented in Figs. 1 and 2 (UCA, spherical t-design, and spherical equal area sampling). In all cases, the target field is a plane wave sound field impinging from direction $\phi_t = 0°,\ \theta_t = 0°$, represented by an $N = 2$ order HOA signal. The continuous horizontal line represents the average original sound field, which is 0dB for the considered plane wave. The dash-dotted line corresponds to the original sound field truncated to the N-th order. It can be clearly seen that the energy progressively reduces at $kr = N$ (first vertical dashed line), which defines the inner boundary of the ring of silence. The dotted line show the average energy of the sound field reproduced by the array under consideration. It can be clearly seen that, for the circular array and for the t-design, the energy increases again at $kr = N_{max}+1$ (second vertical dashed line). This defines the outer boundary of the ring of silence.

For the circular array, the results are consistent with the analytical results presented by Solvang [7]. For the array with spherical equal area sampling, N_{max} is not defined, but a behaviour similar to the spherical t-design array can be observed: the average energy decreases after $kr = N$ and then increases again after approximately $kr = 10$. This demonstrates that the ring of silence concept can be extended to arrays that do not exactly satisfy the orthogonality condition defined by equation (27).

A. Measurements

Measurements in anechoic conditions were performed to verify the results presented above. For simplicity, the analysis was restricted to the 2D scenario using the UCA with $L = 13$. The array consisted of Genelec 8020C loudspeakers, the impulse responses of which were measured using a linear microphone array. The microphone array consisted of 15 B&K type 4189 omnidirectional microphones with 0.037 m separation between each microphone, shown in Fig. 4a. This microphone spacing results in a spatial aliasing frequency of approximately 4600 Hz, corresponding to the point at which the microphone spacing equals a half wavelength. The array was rotated through 360 degrees to capture impulse responses from each loudspeaker to all positions of the sound field

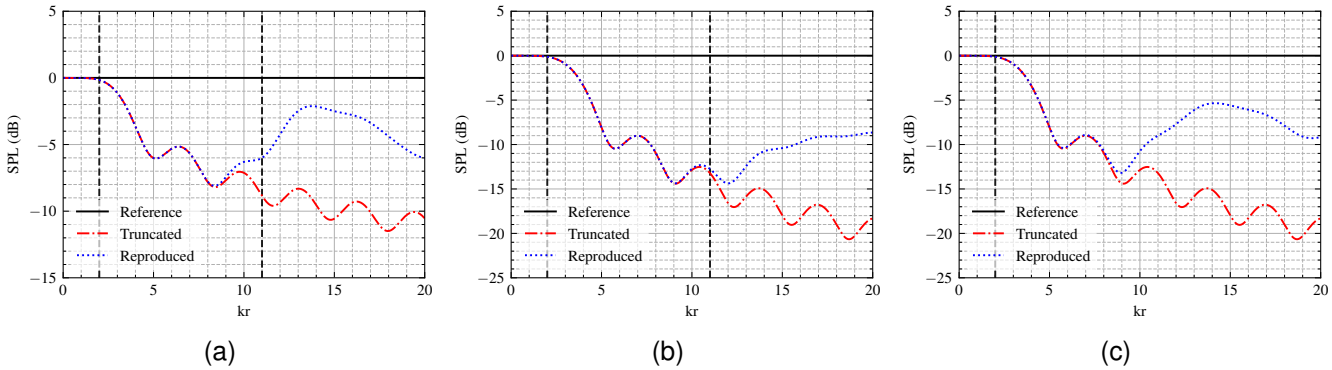

Fig. 3: Average energy for arrays with $N = 2$. (a) UCA, $L = 13,\ N_{max} = 10$ (b) Spherical t-design, $L = 84,\ t = 12,\ N_{max} = 10$ (c) Spherical equal area sampling, $L = 49$. The vertical dashed lines indicate N and $N_{max} + 1$ (not defined for (c)).

over a circular region. Post-processing including frequency-dependent windowing as described in [33] was performed to ensure fully anechoic conditions.

As previously explained, the target field due to a plane wave impinging from direction $\phi_T = 0°,\ \theta_T = 0°$, represented by an $N = 2$ order HOA signal, was utilised to define the loudspeaker signals. The sound field at the measured positions was then simulated using the measured impulse responses. Fig. 4b and 4c show the average energy across the circle at 3000 Hz and 4000 Hz, respectively, and as a function of the radial distance from the centre of the coordinate system (coinciding with the centre of both loudspeaker and microphone arrays). Notably the x axis differs from the previous results, as it is now across radial distance, for a fixed frequency and therefore fixed wave number. These frequencies were chosen to ensure both the radial boundaries of the ring of silence, defined by the $N = kr$ rule of thumb for both N and $N_{max} + 1$, fell within the area sampled by the microphone array. A similar trend to the simulations is observed, where the energy of the field is approximately correct before the limit defined by $r = N/k$ and then the energy drops within the ring of silence region, defined by $N/k \leq r \leq (N_{max} + 1)/k$. At larger radii, the energy grows again due to spatial aliasing, as expected.

It is should be emphasised that, for the case of a single plane wave, the ring of silence is observed when considering the radial variation of the acoustic field *averaged* over a circle or a sphere. In this specific case, when observing the field on a plane perpendicular to the wavefront (e.g. $x - y$ plane if the plane wave propagates in the x direction), the energy is not uniform across a circle of given radius, therefore in practice the region of space with an energy dip that we refer to as the ring of silence will not be an annulus or spherical shell. The actual energy measured radially out from the array centre varies due to a number of factors, most importantly the angle between the observation point and the plane wave direction of propagation. The effect on the energy (not averaged over a circle) due to varying this angle using the measured data at 4000 Hz is shown in Fig. 5a. Here it is notable that, for an incident single plane wave arriving from 0° (direction of x axis), when increasing the observation angle from 0° to 90° (y axis) the drop in energy due to the ring of silence is progressively more pronounced. Due to symmetry, the effect of the ring of silence will then decrease again transitioning from 90° to 180°. In contrast, when analysing the sound field on a plane parallel to the wavefront (e.g. $y - z$ plane for a wave propagating in the x direction) the ring of silence will appear as an annular region (not shown in the figure).

Notably, however, most practical sound fields (particularly those measured using microphone arrays) do not consist of a single plane, but instead of a summation of multiple and often incoherent plane waves. This is the case, for example, for a diffuse sound field, which is of much relevance in practical 3D audio applications. Under this scenario, if these plane waves sum incoherently (i.e multiple independent sources) a low-energy region will be observed, the shape of which resembles an annulus or spherical shell, thus justifying the choice of the name "ring of silence". This is demonstrated in Fig. 5b, where the sound field now consists of 6 incoherent plane waves incident with angular directions equally distributed across a circle. With comparison to Fig. 5a this sound field exhibits the ring of silence in a considerably more uniform manner.

IV. Binaural Reproduction

We now consider the case of binaural reproduction of HOA signals. For a frequency ω, the binaural signals are

$$p_{L/R} = \int_{\hat{\boldsymbol{y}} \in S} H_{L/R}(\hat{\boldsymbol{y}}) q(\hat{\boldsymbol{y}}) dS \tag{29}$$

where $H_{L/R}(\hat{\boldsymbol{y}})$ is either the left-ear or right-ear HRTF for a plane wave arriving from direction $\hat{\boldsymbol{y}}$. The subscript L/R is hereafter omitted to simplify the notation.

The HRTF can be represented by its Fourier series

$$H(\hat{\boldsymbol{y}}) = \sum_{n=0}^{\infty} \sum_{m} \Psi^*_{nm}(\hat{\boldsymbol{y}}) H_{nm} \tag{30}$$

$$H_{nm} = \int_{\hat{\boldsymbol{y}} \in S} H(\hat{\boldsymbol{y}}) \Psi_{nm}\, dS. \tag{31}$$

Note that, without loss of rigour or generality, the Fourier series is expressed in terms of the complex conjugate of the basis functions Ψ_{nm}, in order to simplify the following mathematical manipulation. The binaural signal can also be

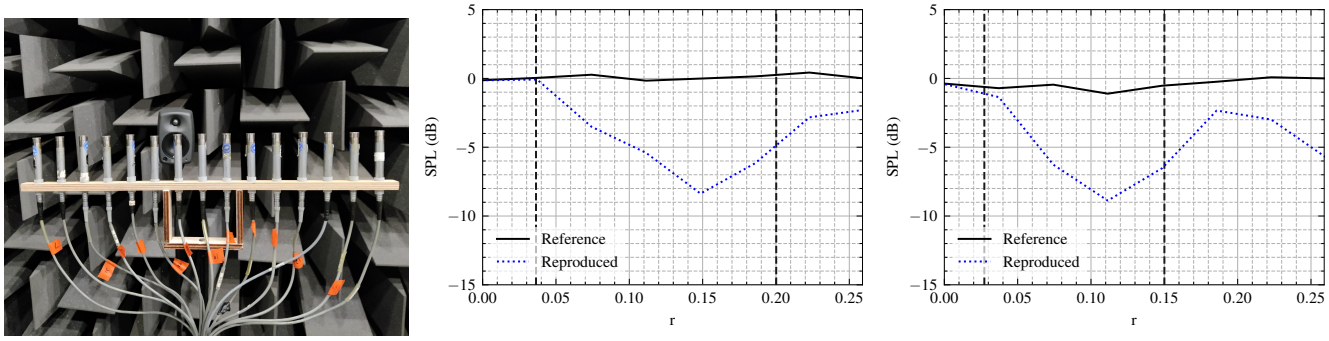

Fig. 4: Average energy from measured data for a UCA with $L = 13$, $N_{max} = 10$ and $N = 2$. The vertical dashed lines indicate the radial distances corresponding to N and $N_{max} + 1$, using the $N = kr$ rule. (a) Picture of the microphone array (b) 3000 Hz (c) 4000 Hz.

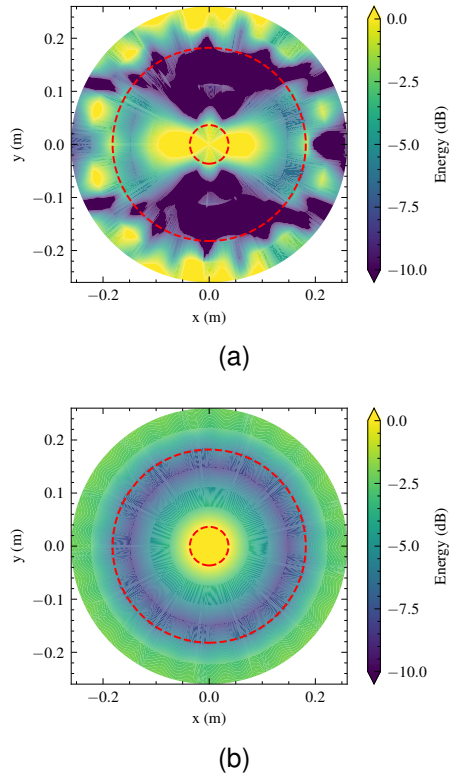

Fig. 5: Measured energy at 4000 Hz reproduced by a UCA with $L = 13$, $N_{max} = 10$ and $N = 2$. The dashed lines indicate the circles with radii corresponding to N and $N_{max}+1$, using the $N = kr$ rule. (a) Single plane wave incident from 0° (b) 6 incoherent plane waves incident with a uniform circular distribution.

expressed in terms of the spatial Fourier coefficients of the PWD, as

$$p = \sum_{n=0}^{\infty} \sum_{m} H_{nm} b_{nm}. \tag{32}$$

This result is derived from equation (29), analogously to equation (9) with H_{nm} instead of $R_{nm}(kr)\Psi_{nm}(\hat{\boldsymbol{r}})$.

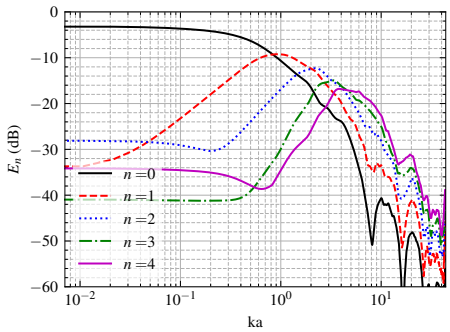

Fig. 6: Energy of the KU100 HRTF Fourier coefficients, summed over all m's for each given n.

Fig. 6 shows the energy of the coefficients H_{nm} as a function of ka, where $a = 0.1$ m is the notional radius of the head. The HRTF is that of a Neumann KU100 dummy head, as provided by the TH Köln 2702 point Lebedev grid measurement [34]. Each line shows the sum of the energy of all orders with same n and different m, i.e. $E_n = \sum_m |H_{nm}|^2$, normalised by the total HRTF energy. It can be observed that, similarly to the free-field radial functions R_{nm}, high-order coefficients do not contribute to small ka values, hence a truncation error can be expected if the series in equation (30) is truncated, as reported in the relevant scientific literature [8], [10]. The relative energetic contribution of each order decays with ka shortly after it has reached its peak, with lower orders decaying before higher orders, but differently than R_{nm}. This has consequences on the outer boundary of the ring of silence, as will be shown later.

A. Two HOA-to-Binaural Decoding Methods

We will refer here to two established methods from the literature for reproducing HOA signals over headphones, both of which, as will be shown later, are directly or indirectly based on the Mode-Matching Decoder. As mentioned in the introductory section of this paper, other, more modern HOA-to-binaural decoding strategies have been developed, but these

are not included in this work. The first method consists in decoding the HOA signals to obtain the driving signals of a virtual array of L loudspeakers arranged in the far field at angual positions $\{\hat{\boldsymbol{y}}_1, \dots, \hat{\boldsymbol{y}}_L\}$, and then convolving these signals with the HRTF corresponding to each loudspeaker [9], [10]. In this case, the PWD Fourier coefficients are identical to those in equation (17). The binaural signal is therefore

$$p = \sum_{n=0}^{\infty} \sum_{m} H_{nm} \sum_{\nu=0}^{N} \sum_{\mu} A_{nm,\nu\mu} b_{\nu\mu} \tag{33}$$

which in matrix form reads

$$p = \boldsymbol{h}^T \boldsymbol{A} \boldsymbol{b}. \tag{34}$$

The second method is derived from equation (32) and consists of multiplying (convolving in time domain) each HOA signal with the corresponding HRTF Fourier coefficient, then summing the results [10], [11]. In mathematical terms:

$$p = \sum_{n=0}^{N} \sum_{m} \tilde{H}_{nm} b_{nm}. \tag{35}$$

In this case, it is assumed that the Fourier coefficients are estimated up to the HOA signal order N. The sum is therefore truncated to order N and the HRTF Fourier coefficients $\tilde{H}_{nm}$ are estimated and may be corrupted by spatial aliasing if the sampling scheme is not sufficiently dense.

Measured HRTF data is often provided as samples $H(\hat{\boldsymbol{y}}_\ell), \ \ell = 1, \dots, L$ over a grid of L points arranged over a sphere or circle. To show the equivalence between the two methods discussed here, we will now assume that these sampling positions $\hat{\boldsymbol{y}}_\ell$ are the same as those of the virtual loudspeaker positions used in the first method introduced above. If the H_{nm} are estimated from this sampled data, the coefficients may be computed with a least-squares estimation (again avoiding the use of regularisation) as [11]

$$\tilde{\boldsymbol{h}}^T = [(\boldsymbol{Y}_N^*)^\dagger \boldsymbol{x}]^T = \boldsymbol{x}^T \boldsymbol{Y}_N (\boldsymbol{Y}_N^H \boldsymbol{Y}_N)^{-1} \tag{36}$$

where $\boldsymbol{x}^T = [H(\hat{\boldsymbol{y}}_1), \dots, H(\hat{\boldsymbol{y}}_L)]$. Coefficients up to order N are estimated and it is assumed that their number is equal to, or smaller than the number of sampling points L - hence the finite order of matrix $\boldsymbol{Y}_N$. Again, if the sampling scheme is not sufficiently dense, the estimated $\tilde{H}_{nm}$ may be corrupted by spatial aliasing.

In view of equation (30), the estimated HRTF Fourier coefficients are related to the original ones by

$$\tilde{\boldsymbol{h}}^T = \boldsymbol{h}^T \boldsymbol{Y}_\infty^H \boldsymbol{Y}_N (\boldsymbol{Y}_N^H \boldsymbol{Y}_N)^{-1} = \boldsymbol{h}^T \boldsymbol{A} \tag{37}$$

where $\boldsymbol{h}$ and $\tilde{\boldsymbol{h}}$ are the vectors of true coefficients H_{nm} and estimated coefficients $\tilde{H}_{nm}$, respectively. The spatial aliasing matrix $\boldsymbol{A}$ introduced above is the same as in equation (26). This HOA decoder design is related to the process of estimating the HRTF Fourier coefficients, with the HRTF playing the role of the PWD and the HRTF sampling points the role of the loudspeakers. Combining equations (35) and (37) yields

$$p = \sum_{\nu=0}^{N} \sum_{\mu} b_{\nu\mu} \sum_{n=0}^{\infty} \sum_{m} A_{\nu\mu,nm} H_{nm} \tag{38}$$

and in matrix form

$$p = \boldsymbol{h}^T \boldsymbol{A} \boldsymbol{b}. \tag{39}$$

This is identical to equation (34), thus proving that the two methods are in fact equivalent if the HRTF is sampled at the same spatial positions as those of the loudspeakers of the virtual array. The equivalence of the two methods may be also understood by observing that equation [35] represents the sound field reproduced by an array of virtual spatially-low-passed loudspeakers, that is an array where each source generates a sound field limited to order N (and possibly including spatial aliasing artefacts).

The results presented above rely on the assumption that the maximum order of the estimated HRTF Fourier coefficients H_{nm} is the same as the order of the HOA signals to be reproduced, i.e. that the order N in eq. (35) is the same as in eq. (36). It may be the case, however, that the order N of the HRTF Fourier coefficients is greater than the order N' of the HOA signal (but the total number of HRTF coefficients is still smaller than L), in which case the coefficients H_{nm} with order $N' < n \leq N$ would be estimated but not used in equation (35). This situation is equivalent to imposing $b_{nm} = 0$ for $N' < n \leq N$, that is extending vector $\boldsymbol{b}$ by adding zero entries for the last $N - N'$ orders. The results presented above will therefore still be valid.

The equivalence between these two methods was already investigated in the literature, most notably with the joint sampling theory presented by Ben-Hur et al. in [10]. The results presented in that work are extended by the derivation presented in this section, which demonstrates that the two methods are equivalent even if the sampling scheme used to estimate $\tilde{H}_{nm}$ for the second method is not sufficiently dense to accurately represent the HRTF, thus breaking the joint sampling condition [10, eq. (13)][5], as long as this sampling scheme is consistent with the the virtual loudspeaker positions used in the first method. The reproduced binaural signal will be inaccurate due to spatial aliasing and truncation errors, but these errors will be the same in the two methods.

B. The Ring of Silence in Binaural Audio

We calculate the energy of p for a plane wave target, represented by N-th order HOA signals, and we average that over all possible incoming directions. The HOA signals for a target plane wave coming from direction $\hat{\boldsymbol{y}}$ are $b_{nm}(\hat{\boldsymbol{y}}) = \Psi_{nm}^*(\hat{\boldsymbol{y}})$. Considering equation (35), the reproduced binaural signal is

$$p(\hat{\boldsymbol{y}}) = \sum_{n=0}^{N} \sum_{m} \tilde{H}_{nm} \Psi_{nm}^*(\hat{\boldsymbol{y}}). \tag{40}$$

The averaged energy is thus

$$E_B = \frac{1}{\alpha 2\pi} \int_{\hat{\boldsymbol{y}} \in S} |p(\hat{\boldsymbol{y}})|^2 dS = \frac{1}{\alpha 2\pi} \sum_{n=0}^{N} \sum_{m} |\tilde{H}_{nm}|^2. \tag{41}$$

Note that this is proportional to the energy of the binaural signal for an isotropic diffuse field.

[5] $N_{\mathcal{L}} < N_h$ in the notation used in [10]

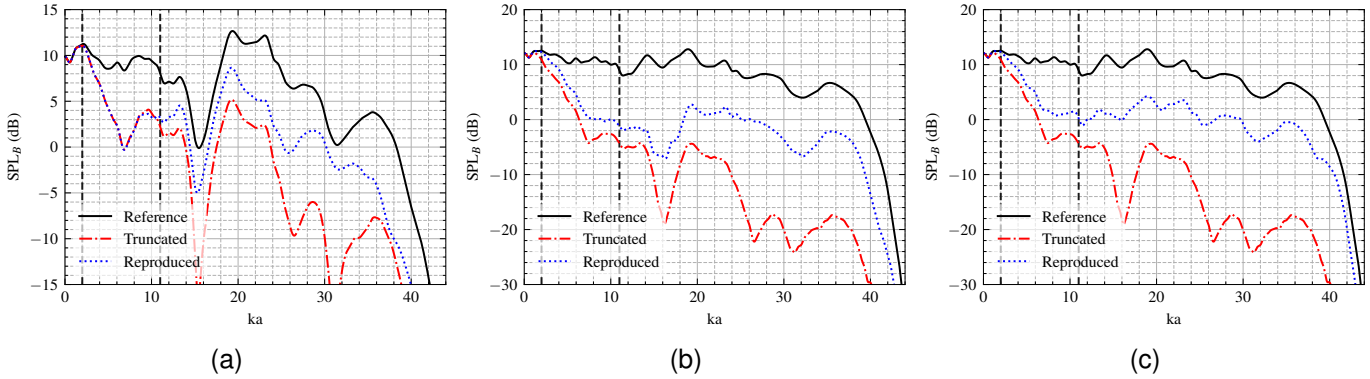

Fig. 7: Average HRTF energy for $N = 2$. The vertical dashed lines indicate N and $N_{max} + 1$ (not defined for (c)). (a) UCA with $L = 13,\ N_{max} = 10$ (b) Spherical t-design with $L = 84,\ t = 12,\ N_{max} = 10$ (c) Spherical equal area sampling with $L = 49$.

Fig. 7 shows the average energy of the binaural signal computed using the same three loudspeaker arrays (in this case virtual arrays) introduced for previous figures. Note that the ka axis is linear, for better visualisation. The plots show the energy of the reference, order-truncated, and reproduced signals. It can be clearly seen that the energy starts decaying when $ka = N$, for both the truncated and reproduced signals, which again sets the inner bound of the ring of silence. This is consistent with the high-frequency spectral impairment observed by Bernshutz et al. [8]. The energy of the reproduced signals increases again when $kr > N_{max} + 1$, but not as markedly as for the free-field case (Fig. 3) because of the different energetic decay of H_{nm} and R_{nm} (see Fig. 6). Note that, for the binaural reproduction case, the ring of silence is in fact a *frequency band of silence*, i.e. an interval on the domain ka, as opposed to a region of space, as in the free-field case.

These results clearly show therefore that the ring (or band) of silence phenomenon occurs also for binaural reproduction of HOA material, and it can be explained with the very same mathematical arguments introduced for the analogous phenomenon occurring in loudspeaker reproduction.

The results presented here bring to light the relation between the ring of silence, the order of the HOA signals to be reproduced, the order of the HRTF expansion, and its sampling scheme. More specifically, the bounds of the ring of silence discussed in Section III, and hence the "thickness" of the ring, are shown to be related to the difference between the order of the chosen HRTF sampling scheme and the order of the HOA signals. This result shares some relevant aspects with the joint sampling theory presented in [10].

V. CONCLUSIONS

It has been shown that, when reproducing HOA signals over a loudspeaker array with a Mode-Matching Decoder, a ring-shaped region with lower SPL is generated, informally referred to as the ring of silence. It has been shown mathematically that this phenomenon is a consequence of the structure of the spatial aliasing matrix, which depends on the loudspeaker arrangement and on the chosen HOA decoder. The ring of silence occurs because the energy of a range of "modes" of the reproduced sound field (from $n > N$ to $n \leq N_{max}$) is set to zero by the popular minimum-norm HOA decoder. Formulae for the inner and outer boundaries of the ring of silence have been derived, which are exact for orthogonal sampling schemes, and provide indicative bounds for approximately uniform layouts. For the special case of a uniform circular array, the results obtained coincide with those presented by Solvang [7] and are demonstrated theoretically and using measurements. The results presented in this work also encompass more generic loudspeaker arrangements, including three-dimensional setups.

A phenomenon analogous to the ring of silence has been shown to occur also when rendering HOA signal binaurally and using decoding strategy related to the Mode-Matching Decoder. A frequency has been identified above which the average energy decreases, and then increases again above a given higher frequency. Formulae for these two regions have been derived mathematically. The high-frequency energy roll-off of HOA signals decoded to binaural was previously reported in the literature, including in the work by Bernshutz et al. [8]. It has been shown that this phenomenon is the same as that of the ring of silence occurring in loudspeaker reproduction.

This work has also highlighted the relation between the ring of silence and the minimum-norm solution used for the HOA decoder's design and for the estimation of the HRTF spherical harmonic coefficients. This finding suggests that a different choice of decoder or spherical harmonics estimation method may prevent the occurrence of this phenomenon or alleviate its severity. Preliminary results obtained by the authors indicate that the ring of silence effect is mitigated when using a decoding strategy based on compressive sensing for Ambisonics rendering over loudspeaker, similar to the approach proposed by Jin at al. [15, p. 49]. Likewise, binaural Ambisonics rendering using the Mag-LS decoding method appears to significantly reduce the occurrence of the ring of silence. Further investigation into this topic is suggested as future work.

It has been shown that the ring of silence is intimately related to the structure of the spatial aliasing matrix, which

in turn depends on the chosen loudspeaker (or virtual loudspeaker) arrangement. This observation suggests that specific sampling arrangement (i.e., sampling schemes) could be designed to reduce the ring of silence. Further investigation into this possibility is proposed as future research.

Finally, all results presented in this work are based on theoretical derivations and validated by objective measurements; however, further validation via subjective listening experiments is recommended.

VI. Acknowledgments

The authors would like to thank Mr Yueheng Li for his assistance with numerical simulations.

References

[1] F. Zotter and M. Frank, *Ambisonics. A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer International Publishing, 2019, vol. 19.

[2] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *J. Audio Eng. Soc*, vol. 53, no. 11, pp. 1004–1025, 2005.

[3] J. Daniel and S. Moreau, "Further study of sound field coding with higher order Ambisonics," in *Audio Eng. Soc. Convention 116*, 2004.

[4] J. Bamford, "An analysis of ambisonic sound systems of first and second order," Ph.D. dissertation, University Of Waterloo, 1995.

[5] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Trans. Audio Speech Lang. Process.*, vol. 9, no. 6, pp. 697–707, 2001.

[6] J. Daniel, J.-B. Rault, and J.-D. Polack, "Ambisonics encoding of other audio formats for multiple listening conditions," in *Audio Eng. Soc. Convention 105*, no. 4795, 1998.

[7] A. Solvang, "Spectral impairment of two-dimensional higher order Ambisonics," *J. Audio Eng. Soc*, vol. 56, no. 4, pp. 267–279, 2008.

[8] B. Bernschütz, A. Giner, C. Pörschmann, and J. Arend, "Binaural reproduction of plane waves with reduced modal order," *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 972–983, 2014.

[9] M. Noisternig, A. Sontacchi, T. Musil, and R. Holdrich, "A 3d ambisonic based binaural sound reproduction system," in *Audio Eng. Soc. Conference: 24th International Conference: Multichannel Audio, The New Reality*, 2003.

[10] Z. Ben-Hur, J. Sheaffer, and B. Rafaely, "Joint sampling theory and subjective investigation of plane-wave and spherical harmonics formulations for binaural reproduction," *Appl. Acoust.*, vol. 134, pp. 138–144, 2018.

[11] B. Rafaely and A. Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," *J. Acoust. Soc. Am.*, vol. 127, no. 2, pp. 823–828, 2010.

[12] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely, "Spectral equalization in binaural signals represented by order-truncated spherical harmonics," *J. Acoust. Soc. Am.*, vol. 141, pp. 4087–4096, 2017.

[13] F. Zotter, H. Pomberger, and M. Noisternig, "Energy-preserving ambisonic decoding," *Acta Acust. united Acust.*, vol. 98, no. 1, pp. 37–47, 2012.

[14] F. Zotter and M. Frank, "All-round ambisonic panning and decoding," *J. Audio Eng. Soc.*, vol. 60, pp. 807–820, november 2012.

[15] V. Pulkki, S. Delikaris-Manias, and A. Politis, Eds., *Parametric Time–Frequency Domain Spatial Audio*. Wiley, 2017.

[16] F. Brinkmann and S. Weinzierl, "Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition," in *AES International Conference on Audio for Virtual and Augmented Reality (AVAR), Redmond, WA, USA*, august 2018.

[17] I. Engel, D. F. M. Goodman, and L. Picinali, "Assessing hrtf preprocessing methods for Ambisonics rendering through perceptual models," *Acta Acust.*, vol. 6, p. 4, 2022.

[18] 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.

[19] M. J. Evans, J. A. S. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *J. Acoust. Soc. Am.*, vol. 104, no. 4, pp. 2400–2411, 10 1998.

[20] J. M. Arend, C. Pörschmann, S. Weinzierl, and F. Brinkmann, "Magnitude-corrected and time-aligned interpolation of head-related transfer functions," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 31, pp. 3783–3799, 2023.

[21] C. Schörkhuber, M. Zaunschirm, and R. Holdrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *44th DAGA*, 2018.

[22] T. Deppisch, H. Helmholz, and J. Ahrens, "End-to-end magnitude least squares binaural rendering of spherical microphone array signals," in *I3DA: International Conference on Immersive and 3D Audio*, 2021, pp. 1–7.

[23] 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.

[24] 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.

[25] Z. Ben-Hur, D. L. Alon, R. Mehra, and B. Rafaely, "Efficient representation and sparse sampling of head-related transfer functions using phase-correction based on ear alignment," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 27, no. 12, pp. 2249–2262, 2019.

[26] ——, "Binaural reproduction based on bilateral Ambisonics and ear-aligned hrtfs," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 29, pp. 901–913, 2021.

[27] B. Rafaely, *Fundamentals of Spherical Array Processing*, 1st ed. Springer-Verlag Berlin Heidelberg, 2015, vol. 8.

[28] E. G. Williams, *Fourier Acoustics. Sound Radiation and Nearfield Acoustical Holography*, 1st ed. London: Academic Press, 1999.

[29] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer, 2013.

[30] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing - Principles, Algorithms and Applications (2. ed.)*. Macmillan, 1992.

[31] R. H. Hardin and N. J. A. Sloane, "Mclaren's improved snub cube and other new spherical designs in three dimensions." *Discrete Comput Geom*, vol. 15, p. 429–441, 1996.

[32] P. Dietrich, M. Guski, M. Pollow, M. Müller-Trapet, B. Masiero, R. Scharrer, and M. Vorländer, "Ita-toolbox – an open source matlab toolbox for acousticians," in *DAGA*, 2012.

[33] J. Hollebon and F. M. Fazi, "Experimental study of various methods for low frequency spatial audio reproduction over loudspeakers," in *I3DA: International Conference on Immersive and 3D Audio*, 2021.

[34] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the Neumann KU 100," in *AIA-DAGA Conference on Acoustics*, 2013.

Filippo Maria Fazi is Professor of Acoustics and Signal Processing at the Institute of Sound and Vibration Research of the University of Southampton, where he also serves as Head of the Signal Processing, Audio and Hearing Group, and leads the Virtual Acoustics and Audio Engineering Team. His research interests include acoustics, audio technologies, electroacoustics and digital signal processing, with special focus on acoustical inverse problems, multi-channel systems (including Ambisonics and Wave Field Synthesis), virtual acoustics, and microphone arrays. He is the author of more than 170 scientific publications and several patents. Prof Fazi was awarded a research fellowship by the Royal Academy of Engineering (2010) and the Tyndall Medal by the Institute of Acoustics (2018). He is a fellow of the Audio Engineering Society, a member of the Institute of Acoustics and is co-founder and chief scientist at Audioscenic, a start-up company that develops and commercialises 3D audio and loudspeaker array technologies.

Jacob Hollebon graduated from the University of Warwick in 2017 with a BSc in Physics. He then completed the MSc in Acoustical Engineering at the Institute of Sound and Vibration Research, University of Southampton in 2018, with a thesis on spatial audio reproduction for multiple listeners. Between 2018 and 2022 he joined the Virtual Acoustics and Audio Engineering team to begin a PhD working in 3D audio reproduction methods over loudspeaker arrays including working with sound field reproduction, Ambisonics and Crosstalk Cancellation. This led to a new spatial audio reproduction technique titled Higher-Order Stereophony. Jacob continued as a Research Fellow between 2022 and 2023 focusing on 3D audio capture and reproduction, with particular focus on array technology and binaural reproduction, and is now Research and Development Lead at Audioscenic, working on the commercialization of listener-adaptive spatial audio technologies. Jacob is the holder of the 2018 ISVR Elsevier prize, as well as being the AES Educational Foundation 2018 Emil Torick Scholar.