DOI: 10.61782/fa.2023.1077

forumacusticum 2023

THE INFLUENCE OF THE NUMBER OF MICROPHONES CONSTITUTING A SPHERICAL MICROPHONE ARRAY ON THE PERCEPTION OF SPATIAL ALIASING.

Gauthier Berthomieu[1*] **François Salmon**[2] **Julian Palacino**[3]
Mathieu Paquier[1]
[1] University of Brest, CNRS, Lab-STICC UMR 6285,
6 avenue Victor Le Gorgeu, CS 93837, 29238 Brest Cedex 3, France
[2] Noise Makers, 3 Rue Brossay Saint-Marc, 35700 Rennes, France
[3] Feichter Audio, Espace de Broglie, 5 rue Louis de Broglie, 22300 Lannion, France

ABSTRACT

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 above which aliasing-induced artifacts occur is related to the number of microphones that constitute the array, along with the array radius and the sampling scheme according to which the microphones are positioned. In this study, ambisonics signals of order $N = 7$ are encoded from simulated recordings made by virtual spherical microphone arrays of varying characteristics (radius, sampling scheme, number of microphones...). These aliased stimuli are then compared to unaliased stimuli of order $N = 7$. An adaptive procedure reveals the minimum number of microphones required for aliased stimuli to become indistinguishable from unaliased stimuli.

Keywords: *Ambisonics, Spherical Microphone Array, Spatial Aliasing, diffuse-field equalization*

1. INTRODUCTION

Spherical Microphone Arrays (SMA) are commonly used for capturing sound fields in a physically precise way. Recent technology advances in this field allow to create spatial audio content that accurately copies the real-life audio environment. Sound fields recorded thanks to SMAs are commonly described in the Spherical Harmonics (SH) domain by the use of the ambisonics format. This formatting provides features such as rotating the sound field according to the listener's head orientation in real time before rendering it binaurally (dynamic binaural rendering).

1.1 Spatial resolution

When a SMA captures a sound field, the pressure that applies on the continuous surface elements (the surface of the sphere along which the microphones are distributed) is only captured at the discrete positions of the microphones. As a result, the SH representation of the sound field (which is theoretically composed of an infinite sum of discrete SH of different degrees and modes) is limited to a finite order. The physical accuracy of the represented sound field thus relies on the maximal SH order of the representation (referred to as the *ambisonic order*), which is limited by the number of microphones constituting the SMA.

Moreover, the restricted number of sensors leads to spatial ambiguities, where spatial information that belongs to higher-order harmonics are aliased into lower-order harmonics [1]. As higher orders harmonics contribute mostly to the representation of the higher frequencies (the contribution of harmonics of order n decay steeply for $n > kr$, where r is the SMA radius and k the wave number), spatial aliasing implies a corruption of the upper-frequency range of the recorded sound field. The

****Corresponding author***: *gauthier.berthomieu@univ-brest.fr.*

aliasing frequency f_A beyond which spatial aliasing is not negligible decreases as the distance between microphones increases [1]. Hence, for a SMA, f_A depends on the number of microphones, the radius and the sampling scheme according to which the microphones are arranged on the sphere.

1.2 Perceptual evaluation of the artifacts

The errors between the actual sound field recorded with a SMA and the reproduced sound field can lead to audible artifacts that degrade the quality of the listening experience. As an example, the truncation-induced error is known to reduce the localization accuracy as well as the externalization of binaural reproductions [2].

Some studies aimed at defining minimal SH orders beyond which the effect of encoding errors on the rendered audio content is negligible. Lübeck *et al.* [3] evaluated Points of Subjective Equality (PSE) between binaural auralizations synthetized from impulse responses captured through SMAs of varying ambisonic orders (inducing varying truncation and aliasing error) and a binaural reference synthetized from a SMA of order $N = 29$ (for which the introduced artifacts can be neglected). Moreover, they evaluated PSEs between the reference and binaurals signals synthetized from Binaural Room Impulse Response (BRIR) SH interpolations, which suffer from truncation error but for which each individual measurement has the maximum spatial resolution and thus do not exhibit artefacts due to spatial aliasing. Following an ABX three-interval/two-alternative forced choice (3I/2AFC) test design, the authors reported that orders $N = 17 - 20$ were required for the auralizations of SMA data to be indistinguishable from the reference, whereas orders $N = 9 - 13$ were required for the auralization of BRIR data. However, these PSEs were obtained using a baseline approach, and several methods have been developped in the last years to improve the binaural renderings of SMA captures, such as the diffuse-field equalization.

1.3 Diffuse-field equalization

As the aliasing error increases, the sound field is not correctly represented in the SH domain. The aliased components induce an increase of the high frequency energy, leading to an audible spectral coloration. The diffuse-field equalization step aims at correcting the aliased signal spectrum to that of a diffuse sound field beyond a given frequency.

This equalization was performed on the encoding matrix (which allows to convert the microphonic signals into SH signals) in the samme manner as in McCormack et al. [4] for $f > f_A$.

2. EXPERIMENT

The present experiment aims at evaluating the minimal number of microphones constituting a SMA beyond which the spatial aliasing error is no longer perceptible. An adaptive procedure was carried out, consisting in pairwise comparisons between spatially aliased stimuli and reference signals that do not exhibit spatial aliasing. SMAs of different radii and sampling schemes were investigated.

The aliased stimuli were computed using different virtual SMAs and reproduced through headphones following a dynamic binaural decoding. In order to evaluate the effect of the diffuse-field equalization on the reproduced signals, the experiment was carried out with both equalized and non-equalized encoding matrices.

2.1 Sound sources

Spatial Room Impulse Responses (SRIR) were simulated to model the incident sound field reaching the SMA. A small ($5.3\,\mathrm{m} \times 4.1\,\mathrm{m} \times 2.3\,\mathrm{m}$ – width $\times$ depth $\times$ height) room was modeled as a shoe-box, in which five sound sources were placed with an azimuth angle of $\pm 40°$, $\pm 15°$ and $0°$ from the recording position. The reverberation time of the room was equal to $0.24\,\mathrm{s}$. These SRIR were simulated at the order $N = 33$ so that the the incident sound field was physically accurate.

To create virtual SMA signals, the microphone signals were first computed using different sampling grid orders N_g (which defines the number of microphones), sphere radii and sampling schemes depending on the SMA configuration from the SRIRs of ambisonic order $N = 33$, and encoded in ambisonics at the order $\tilde{N} = 7$.

The unaliased reference stimuli were created by directly truncating the simulated SRIRs of order $N = 33$ at the order $\tilde{N} = 7$. The resulting SRIRs were then convolved with anechoic sound sources to create the stimuli. In summary, aliased ambisonics signals of order $\tilde{N} = 7$ exhibiting different amount of spatial aliasing induced error (depending on the characteristics of the SMA and the diffuse-field equalization step) were compared to unaliased ambisonic signals of order $\tilde{N} = 7$. The order $\tilde{N} = 7$ was chosen since it has been reported to provide a reasonable trade-off between technical effort and perceptual quality [5].

In total, 60 virtual SMAs were created by combining 2 radii ($6\,\mathrm{cm}$ and $8.75\,\mathrm{cm}$), 15 sampling grid orders (7 to 21) and 2 sampling schemes (Gauss-Legendre [6] and Extremal [7]). 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. The virtual SMA properties and the corresponding aliasing frequency are shown in Table 1.

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 grid		Radius	
	Extremal	Gauss-L.	6 cm	8.75 cm
N_g	Microphones		Aliasing freq. (Hz)	
7	64	128	6368	4367
8	81	162	7278	4991
...	...	...	...	...
21	484	968	19106	13101

For these 60 different SMAs, two encoding matrices were created: one being diffuse-field equalized and one not being diffused-field equalized.

Two audio contents were considered in the experiment: a $2\,\mathrm{s}$-long pink noise and a $4.5\,\mathrm{s}$-long musical excerpt of *Don't mean a thing* from Duke Ellington, played by a quartet comprising a violin, a tenor saxophone, a guitar and a double-bass.

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 SRIRs and the resulting signals were summed to synthesize the 7th-order ambisonic scene. The pink noise signal was convolved with SRIRs computed for a sound source located at 0° in the horizontal plane.

Stimuli were presented over open, circumaural headphones (Sennheiser HD650). The ambisonic signals were converted into binaural headphone signals using the *Ambi Head HD* VST plug-in from Noise Makers [8] with HRTFs derived from measurements of a Neumann KU 100 dummy head [9]. This decoder implements the binaural rendering approach of Schörkhuber *et al.* [10] and Zaunschirm *et al.* [11]. Head-tracking was performed using the *T3* head-tracker from Feichter Electronic [12] along with the *Ambi Head HD* VST plug-in.

2.2 Procedure

This experiment intended to highlight to what extent the aliasing artifacts were perceptible in different SMA configurations. Here, the aliasing frequency f_A was shifted up or down by increasing or reducing the grid order N_g of the SMA, respectively. This was done in an adaptive procedure which aimed at revealing a PSE, in terms of grid order, beyond which the induced aliasing artifacts were not detected by the listeners.

PSEs were evaluated independently for each combination of the SMA characteristics and for equalized and non-equalized encoding matrices. In total, the procedure consisted in a repeated-measures design gathering 16 runs (2 radii $\times$ 2 grids $\times$ 2 equalization conditions $\times$ 2 audio contents).

Each run started at a grid order $N_g = 7$ (i.e. with the lowest number of microphones for the corresponding sampling scheme given Table 1), and the grid order was either increased or decreased after each trial following a one-up-one-down staircase method.

Each trial of the run consisted in a three-interval/three-alternative forced choice (3I/3AFC) test. During each trial, three intervals (A, B and C) were presented to the listener. Each interval could be assigned either the stimulus which grid order varied accordingly to the procedure (the test stimulus) or the unaliased reference (the reference stimulus). During each trial, two intervals were assigned the same stimulus (either the test or reference stimulus) and the third interval (named the oddball) was assigned the other stimulus. During each trial, each of A, B and C was played once in this order. Participants were instructed to designate which of A, B or C was the oddball.

The grid order was increased by one for the next trial of this run if the listener correctly identified the oddball whereas it was decreased by one if the listener gave the wrong answer. Each run ended after 9 reversals (a reversal is defined as a correct answer followed by a wrong answer, or vice versa).

Playback, commands, and data capture were controlled by a software implemented in Max 8.

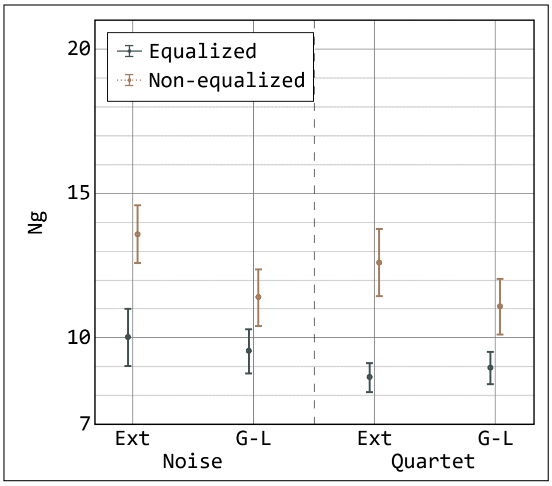
The experiment consisted of two $30\,\mathrm{min}$-long sessions (one per sampling scheme condition). A training phase allowed subjects to become thoroughly familiar with the test environment and the grading process prior to formal grading.

2.3 Pre-results

The pre-results presented in this paper were gathered on 10 expert participants who were related to the Sound Master's degree of the University of Brest.

Fig. 1 shows that grid order PSEs were lowered by the diffuse-field equalization applied to the encoding matrices for each SMA configuration and each recorded audio content. Grids of order as low as $N_g = 7 - 10$ could lead to auralizations of the music excerpt that were indistinguishable from unaliased references providing the encoding matrices were equalized. When the encoding matrices were not equalized, SMAs required $N_g > 10$ in each configuration for the aliased stimuli to be indistinguishable from the unaliased references. As Lübeck *et al.* reported that truncation error might be audible up to the orders $N = 9 - 13$ [3], correcting the aliasing-induced error thanks to a diffuse-field equalization might allow both truncation and aliasing induced artifacts to no longer be audible beyond these orders.

Figure 1. Grid order PSE of SMAs using an Extremal (Ext) or Gauss-Lengendre (G-L) sampling scheme, recording noise (left) and musical excerpt (right) with (blue) or without (brown) diffuse-field equalization.



3. REFERENCES

[1] B. Rafaely, "Spatial Aliasing in Spherical Microphone Arrays," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, p. 8, 2007.

[2] A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," *The Journal of the Acoustical Society of America*, vol. 133, pp. 2711–2721, May 2013.

[3] 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," *The Journal of the Acoustical Society of America*, vol. 151, pp. 467–483, Jan. 2022.

[4] L. McCormack, A. Politis, R. Gonzalez, T. Lokki, and V. Pulkki, "Parametric ambisonic encoding of arbitrary microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 2062–2075, 2022.

[5] B. Bernschütz, *Microphone arrays and sound field decomposition for dynamic binaural recording*. Technische Universitaet Berlin (Germany), 2016.

[6] I. Bogaert, "Iteration-free computation of gauss–legendre quadrature nodes and weights," *SIAM Journal on Scientific Computing*, vol. 36, no. 3, pp. A1008–A1026, 2014.

[7] I. H. Sloan and R. S. Womersley, "Extremal systems of points and numerical integration on the sphere," *Advances in Computational Mathematics*, vol. 21, pp. 107–125, 2004.

[8] Noise Makers, "AmbiHeadHD."

[9] B. Bernschütz, "A spherical far field HRIR/HRTF compilation of the neumann ku 100," in *Proc. of the 40th Italian (AIA) annual conference on acoustics and the 39th German annual conference on acoustics (DAGA)*, p. 29, AIA/DAGA, 2013.

[10] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural rendering of ambisonic signals via magnitude least squares," in *Proc. of the DAGA*, vol. 44, pp. 339–342, 2018.

[11] M. Zaunschirm, C. Schörkhuber, and R. Höldrich, "Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint," *The Journal of the Acoustical Society of America*, vol. 143, no. 6, pp. 3616–3627, 2018.

[12] Feichter Electronics, "T3 Head-tracker."