OPEN ACCESS Freely available online
E. Hoffbauer and M. Frank, "Four-Directional Ambisonic Spatial Decomposition Method With Reduced Temporal Artifacts"
J. Audio Eng. Soc., vol. 70, no. 12, pp. 1002–1014, (2022 December).
DOI: https://doi.org/10.17743/jaes.2022.0039

Four-Directional Ambisonic Spatial Decomposition Method With Reduced Temporal Artifacts

ELIAS HOFFBAUER, AND MATTHIAS FRANK, *AES Associate Member*
(elias.hoffbauer@alumni.tugraz.at) (frank@iem.at)

Institute of Electronic Music and Acoustics, University of Music and Performing Arts, Graz, Austria

For the creation of convincing virtual acoustics of existing rooms and spaces, it is useful to apply measured Ambisonic room impulse responses (ARIRs) as a convolution reverb. Typically, tetrahedral arrays offering only first-order resolution are the preferred practical choice for measurements, because they are easily available and processed. In contrast, higher order is preferred in playback because it is superior in terms of localization accuracy and spatial clarity. There are a number of algorithms that enhance the spatial resolution of first-order ARIRs. However, these algorithms may introduce coloration and artifacts. This paper presents an improvement of the Ambisonic Spatial Decomposition Method by using four directions simultaneously. The additional signals increase the echo density and thereby better preserve the diffuse sound field components during the process of enhancing measured first-order ARIRs to higher orders. An instrumental validation and a series of listening experiments compare the proposed Four-Directional Ambisonic Spatial Decomposition Method to other existing algorithms and prove its similarity to the best algorithm in terms of enhanced spatial clarity and coloration while producing the least artifacts.

0 INTRODUCTION

Ambisonics was first introduced as a 3D surround sound playback format and panning technology for positioning of virtual source objects in the 1970s [1, 2]. At this time, it was limited to first-order resolution that consists of the zeroth-order pattern and three first-order, orthogonal figure-eight directivity patterns. Later, it was named first-order Ambisonics (FOA). The FOA directivity functions drove the design of first-order tetrahedral microphone arrays that allowed to capture sound preserving its directional information and enabling the recording of Ambisonic material [3, 4].

Twenty years later, the evolution of digital signal processing and steadily increasing computing power made higher-order Ambisonics (HOA) a busy research area [5–9]. With its increased spatial resolution, HOA achieves better localization of the reproduced sound events on surrounding loudspeaker systems [10, 11]. Especially at off-center listening positions, the improvement is substantial [12–16], resulting in an increased sweet area [17]. Because of the decreased inter-channel crosstalk, decorrelation of signals can be largely preserved in playback, improving the perception of spatial depth [18].

In virtual acoustics, a typical approach of recreating the acoustical properties of existing places is to measure their room impulse responses (RIRs). When the measurement is done with a directional microphone array, the spatial information of direct sound, early reflections, and diffuse reverberation is preserved (acknowledging the physical limitations given by the array topology). The application of an FOA microphone array, e.g., a tetrahedral arrangement of four cardioid microphone capsules, is a practical solution. However, the limited spatial resolution of FOA RIRs results in a blurry image with reduced spatial clarity when decoded to loudspeakers directly. It can thus be favorable to rather work with its higher-order counterpart (HOA RIR). Higher-order microphone arrays with at least $(N + 1)^2$ microphones on a rigid sphere offer to capture HOA RIRs with a maximum order of N. These arrays can be more expensive, and their processing requires radial steering filters to boost higher orders at low frequencies. This processing step necessitates a high-enough signal-to-noise ratio [19, 20].

In practice, there is a trade-off between spatial resolution and noise boost, and the resulting spatial resolution will have to be successively reduced toward lower frequencies. Alternatively, HOA RIRs can be up-mixed from FOA RIR measurements through a subsequent algorithmic enhancement of their spatial resolution.

During the past years, a number of different resolution-enhancement algorithms have been presented for FOA

RIRs. Their designs are based on different sound field models: The Spatial Decomposition Method (SDM) [21] and Ambisonic SDM (ASDM) [22] assume that there is only a single direction of arrival (DOA) per time frame present in the measured impulse response.

In contrast, (Higher-Order) Spatial Impulse Response Rendering (SIRR) [23, 24] implies that there are simultaneously multiple DOAs in different frequency bands [of the equivalent rectangular bandwidth of human hearing], and, therefore, a narrow-band DOA analysis is conducted. Additionally, (HO-)SIRR tries to reproduce the diffuse part of the captured sound field based on a diffuseness estimate.

The so-called "Directional Enhancement By The 2 + 2 Directional Signal Estimator" (2DSE2) [25] allows the existence of one to two broadband DOAs plus two diffuse directions around each sample in order to improve the transition from direct sound to early and late reflections. The sound field model of the proposed Four-Directional–ASDM (4D-ASDM) relies on the same assumption as (A)SDM, i.e., a single DOA per time frame, but tries to represent the diffuse sound component with three additional directions that depend on the detected one.

Since the introduction of (A)SDM in 2013, multiple publications successfully employed it to accomplish reliable enhancement results for a wide variety of measured RIRs regarding the improved spatial resolution compared with FOA RIRs and the perceptual similarity to dummy head measurements [18, 22, 26–28]. Despite these results, a weakness of (A)SDM could also be observed: After the convolution with transient signals or by listening to the HOA RIRs directly, artifacts described as graininess, rattling, or roughness [29] can be perceived. This observation was recently discussed with enhanced FOA RIRs that were simulated [24]. Possible improvements for the binaural reproduction were proposed in [30].

This paper proposes and extensively tests the new algorithm variant, called 4D-ASDM, that is based on ASDM and reduces the graininess artifact in HOA RIRs up-mixed from FOA RIRs. In SEC. 1, first the properties of the artifacts and current findings in research are discussed (SEC. 1.1). After a concise presentation of the conventional ASDM (SEC. 1.2), the new algorithm is described regarding its differences (SEC. 1.3). In SEC. 2, the results of 4D-ASDM are technically validated by room acoustical measures, and a series of listening experiments in SEC. 3 using headphones (SEC. 3.3) and loudspeakers (SEC. 3.4) compares its performance to existing up-mixing algorithms and measured HOA RIRs. The experiments employ measured FOA RIRs from differently large rooms and evaluate artifacts, coloration, and spatial clarity. The test results are discussed in relation to the technical differences of the algorithms (SEC. 3.5). Finally, the new algorithm and most-prominent experimental findings are summarized in a short conclusion in SEC. 4.

1 TOWARD THE NEW ALGORITHM

The proposed up-mixing algorithm for measured FOA RIRs is based on ASDM. Firstly, this section explains the origin of spectral and temporal artifacts in broadband parametric methods, such as (A)SDM, and presents current approaches to minimize them. After an overview of ASDM, the extensions that lead to 4D-ASDM are presented. The signal flow of both algorithms is shown in Fig. 1.

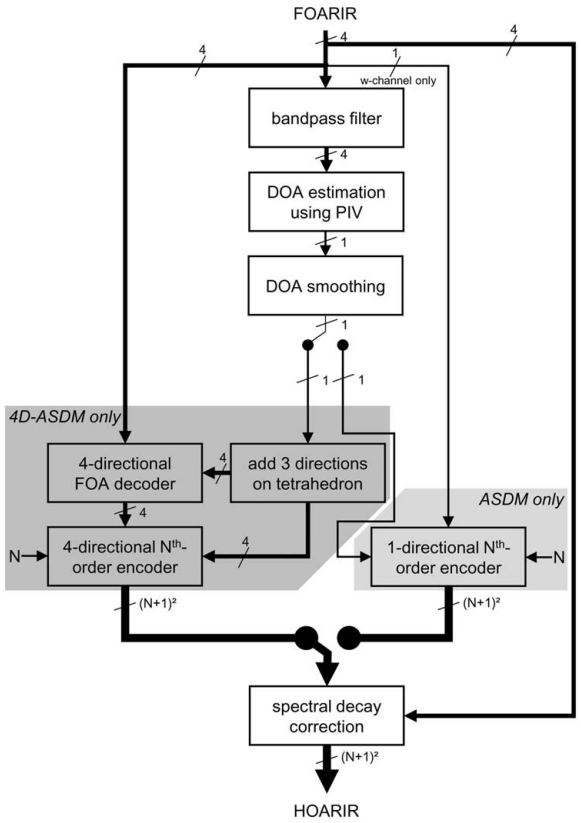

Fig. 1. Signal flow of Ambisonic Spatial Decomposition Method (ASDM) (right path) and Four-Directional–ASDM (4D-ASDM) (left path) to enhance a first-order Ambisonics (FOA) room impulse response (RIR) to a higher-order Ambisonics (HOA) of order N RIR. DOA = direction of arrival.

1.1 Algorithmic Artifacts and Current Research

When listening to HOA RIRs that were created by up-mixing with variants of the SDM, such as (A)SDM or 2DSE2, temporal and spectral artifacts can be perceived in the late reverberation of the signal. In this part of the impulse response, the sound field model of the SDM fundamentally fails to describe the physical state of the sound field. Although the assumption of a single DOA per time instance holds adequately for the direct sound and early reflections, it is clearly violated in the diffuse part, where sounds from multiple directions reach the microphone array simultaneously. The DOA estimation based on the pseudo-intensity vector (PIV) starts to fluctuate highly between neighboring samples and induces artifacts in two ways.

In the re-encoding procedure in HOA, the fluctuation yields a spill of typically long low-frequency reverberation tails into higher frequencies, resulting in an unnatural increase of the reverberation time of higher frequencies at higher orders [31, 18, 22]. There are already some approaches to eliminate these spectral artifacts: In order to

maintain the temporal and spectral features of the impulse response, Tervo et al. applied a nine-band filter bank in a post-equalization stage for every decoded loudspeaker signal in their loudspeaker-based reproduction approach [31]. The Ambisonic variant ASDM also has a spectral decay correction implemented in the signal flow (see Fig. 1), which is presented in more detail in SEC. 1.2.

The second type of artifacts arises by estimating only a single DOA per time instance in the late part. The lack of diffuse information yields a sparse HOA RIR, which is perceived as rough and grainy. A listening experiment conducted by Meyer-Kahlen et al., in which Gaussian noise was randomly assigned to different numbers of loudspeakers, indicated that roughness artifacts are connected to encoding the late reflections into different directions. In a further binaural listening test, they could show that mainly level differences caused by head-shadowing are the origin of the perceived roughness. Modifying the presented stimuli by constraining the spatialization by distributing the impulses more evenly and thereby making the reproduced impulse response less sparse, yielded a less strong perception of roughness by the participants [29].

Based on similar observations, Amengual Garì et al. proposed to enrich the echo density of the sparse impulse response by adding artificial reverberation using three cascaded Schroeder all-pass filters after the spectral equalization stage. It was reported on informal listening experiments, which suggests that the perceived graininess decreases by applying this method [30]. The same goal is pursued by the application of Ambisonic widening [32, 33], used in the 2DSE2 by Gölles and Zotter [25].

The above-presented findings were published during the work on this publication. Whereas the other approaches try to reduce the temporal artifacts by synthetically adding reflections or widening [25, 30], 4D-ASDM only uses information that is already contained in the measured FOA RIR. Thereby, the main focus was to preserve temporal and spectral features of the FOA RIR. Regarding these features, the above-mentioned approaches for reducing the graininess, such as all-pass cascades or a constraint assignment algorithm, still need to be formally and perceptually evaluated. (HO-)SIRR does not exhibit this kind of temporal artifacts because of its different sound field model that employs multiple DOAs in frequency domain and estimates the diffuseness to synthesize the diffuse part of the RIR by spatializing decorrelated copies of the omnidirectional impulse response around the listener [34].

1.2 ASDM

The original SDM by Tervo et al. [21, 31] assumed sound arrived only from a single direction within a small time frame and used pairwise time-delays between the microphones of a compact array to estimate a broadband DOA for each time frame. In contrast, its Ambisonic version [18], which later became the ASDM [22, 27], directly uses the FOA RIR $\boldsymbol{h}(t) = [w(t), x(t), y(t), z(t)]$ of a coincident microphone array to calculate the direction of the pseudo-intensity vector as DOA, cf. Fig. 1:

$$\boldsymbol{\theta}_{\mathrm{DOA}}(t) = \frac{\tilde{\boldsymbol{\theta}}_{\mathrm{DOA}}(t)}{||\tilde{\boldsymbol{\theta}}_{\mathrm{DOA}}(t)||}. \tag{1}$$

This is similar to (HO-)SIRR [23]; however, it calculated broadband for every sample of the impulse response.[1] For the DOA estimation, a zero-phase band-pass filter $\mathcal{F}_{\mathrm{bp}}$ is applied to minimize low-frequency disturbance and ambiguities above the spatial aliasing frequency of the array. For typical FOA microphone arrays, the pass-band is set to 200 Hz and 4 kHz. Note that this filter is only applied to the signals used for DOA estimation but will not affect the full-band microphone signals that will be up-mixed to the final HOA RIR. Finally, the fluctuation of the DOA is smoothed by a median filter $\mathcal{F}_{\mathrm{smth}}$ with a length of about 10 samples at a sampling rate of 48 kHz (≈0.2 ms), which is also related to the typical geometry of FOA microphone arrays. None of the filtering and smoothing operations introduces a delay, therefore the DOA,

$$\tilde{\boldsymbol{\theta}}_{\mathrm{DOA}}(t) = \mathcal{F}_{\mathrm{smth}}\left\{\mathcal{F}_{\mathrm{bp}}\{w(t)\}\mathcal{F}_{\mathrm{bp}}\left\{\begin{bmatrix} x(t) \\ y(t) \\ z(t) \end{bmatrix}\right\}\right\}, \tag{2}$$

stays synchronized with the FOA RIR signals.

The spatially enhanced impulse response results from encoding the unfiltered omnidirectional impulse response $w(t)$ into the direction $\boldsymbol{\theta}_{\mathrm{DOA}}(t)$ for each sample at arbitrarily high orders, see right path in Fig. 1. A crucial component of ASDM is a correction of the spectral decay in the HOA RIR so that it matches the original RIR. In the first implementation [18], this was done by estimating the reverberation time of the original omnidirectional RIR $w(t)$ and the HOA RIR in third-octave bands and correcting the latter by multiplying with decaying exponentials. The later versions, starting from [22], used a simpler spectral correction that equalizes the third-octave energetic envelope $\mathcal{E}\{\cdot\}$ within each spherical harmonic order rather than correcting the reverberation times by exponential slopes. For every time instance t and third-octave band b, a correction factor is calculated and applied to the up-mixed impulse response $\tilde{\boldsymbol{h}}_N(t)$:

$$\hat{h}_{nm}(t,b) = \tilde{h}_{mn}(t,b)\sqrt{\frac{(2n+1)\mathcal{E}\{|h_{\mathrm{ref}}(t,b)|^2\}}{\sum_{m=-n}^{n}\mathcal{E}\{|h_{mn}(t,b)|^2\}}}. \tag{3}$$

Although in previous implementations the reference signal $h_{\mathrm{ref}}(t,b)$ was the omnidirectional channel only, a mixture of all first-order channels is considered in order to achieve a timbre that is similar to the cardioid capsules of the FOA microphone, cf. Fig. 1:

$$\mathcal{E}\{|h_{\mathrm{ref}}(t,b)|^2\} = |w(t,b)|^2 + \frac{|x(t,b)|^2 + |y(t,b)|^2 + |z(t,b)|^2}{3}. \tag{4}$$

[1] The HO-SIRR implementation by L. McCormack et al. [24] allows for analyzing the first peak as a broadband signal, which was the setting used in the listening experiments in SEC. 3.

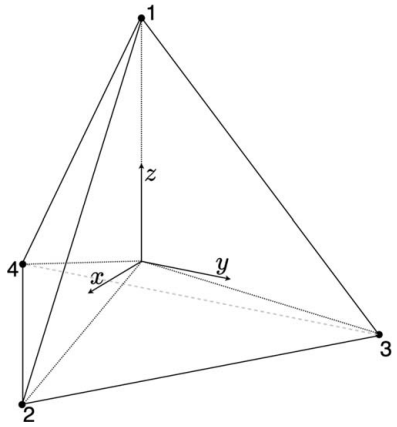

Fig. 2. Prototype tetrahedron in initial orientation; the positions of the vertices equal Eq. (8).

The final HOA RIR consists of the sum of all spectrally corrected components of the up-mixed impulse response:

$$\hat{\boldsymbol{h}}_N(t) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \hat{h}_{nm}(t). \tag{5}$$

As already mentioned before, results from [24, 29] indicate that (A)SDM suffers from a perceived graininess, when convolving the resulting HOA RIR with transient signals. The first idea to smooth or blur this artifact was to introduce additional artificial sound events to increase the temporal density by applying an Ambisonic widening effect [32, 33]. Informal listening indeed revealed a reduction of the artifact, although it was at the cost of reduced spatial clarity and increased coloration.

1.3 4D-ASDM

The idea behind the proposed algorithm is to increase the temporal density of the resulting impulse response, though without adding artificial components. This is achieved by extracting and re-encoding four directions that are arranged in the constellation of a regular tetrahedron, cf. Fig. 2. To accomplish a resolution enhancement, the tetrahedron is rotation-aligned so that its first vertex matches the DOA estimated by the PIV. Hence, it is called 4D-ASDM. The signals for the four directions are extracted by sampling/decoding the FOA RIR at the directions of the tetrahedron, cf. Fig. 1. In any orientation, the regular tetrahedron has multiple advantages for decoding and encoding: The decoding can yield completely decorrelated signals $s_1(t)$ to $s_4(t)$ so that comb filters and similar effects are avoided. Because the tetrahedron itself is an optimal spherical t-design of degree $t = 2$ [35], the orthogonality constraint $t \geq 2N$ is fulfilled for the order $N = 1$. Thus, the FOA RIR is preserved throughout the enhancement process, which is different to ASDM.

The most efficient method to construct a tetrahedron based on the estimated DOA $\boldsymbol{\theta}_{\mathrm{DOA}} = \boldsymbol{\theta}_1$ with three additional directions $\boldsymbol{\theta}_2$ to $\boldsymbol{\theta}_4$ is to define a rotation matrix $\boldsymbol{R}_{\varphi\vartheta}$, which rotation-aligns the regular tetrahedron with the estimated direction $\boldsymbol{\theta}_1$:

$$\boldsymbol{\Theta} = \boldsymbol{R}_{\varphi\vartheta}\boldsymbol{\Theta}_0 \tag{6}$$

$$= \begin{bmatrix} \cos(\varphi)\cos(\vartheta) & -\sin(\varphi) & \cos(\varphi)\sin(\vartheta) \\ \sin(\varphi)\cos(\vartheta) & \cos(\varphi) & \sin(\varphi)\sin(\vartheta) \\ -\sin(\vartheta) & 0 & \cos(\vartheta) \end{bmatrix} \boldsymbol{\Theta}_0, \tag{7}$$

where $\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_4]$ is the set of the four aligned direction vectors and $\boldsymbol{\Theta}_0$ the direction vector set of the prototype tetrahedron, cf. Fig. 2,

$$\boldsymbol{\Theta}_0 = \begin{bmatrix} 0 & \sqrt{\frac{8}{9}} & -\sqrt{\frac{2}{9}} & -\sqrt{\frac{2}{9}} \\ 0 & 0 & \sqrt{\frac{2}{3}} & -\sqrt{\frac{2}{3}} \\ 1 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{bmatrix}. \tag{8}$$

From Eqs. (7) and (8) follows that the third column vector of the rotation matrix $\boldsymbol{R}_{\varphi\vartheta}$ has to align $\boldsymbol{\theta}_1$ with $\boldsymbol{\theta}_{\mathrm{DOA}}$. Hence, the azimuth angle φ and zenith angle ϑ can be computed from the relation

$$\boldsymbol{\theta}_1 = [\cos(\varphi)\sin(\vartheta) \quad \sin(\varphi)\sin(\vartheta) \quad \cos(\vartheta)]^{\mathrm{T}}. \tag{9}$$

The values for the azimuth and zenith angles φ and ϑ and therefore the rotation matrix $\boldsymbol{R}_{\varphi\vartheta}$ are updated sample-wise for every new estimated direction $\boldsymbol{\theta}_{\mathrm{DOA}}$, and an efficient alignment of the tetrahedral arrangement is achieved.

The decoding of the four signals $\boldsymbol{s}(t) = [s_1(t), \ldots, s_4(t)]$ is done by multiplying the measured FOA RIR $\boldsymbol{h}(t)$ with the transposed weighting matrix $\boldsymbol{Y}_{N=1}$:

$$\boldsymbol{s}(t) = \boldsymbol{Y}_1^T \boldsymbol{h}(t)\,. \tag{10}$$

$\boldsymbol{Y}_1$ consists of the weights of the spherical harmonics Y_n^m evaluated in the directions $\boldsymbol{\theta}_1$ to $\boldsymbol{\theta}_4$ of the tetrahedron:

$$\boldsymbol{Y}_N(\boldsymbol{\Theta}) = [\boldsymbol{y}_N(\boldsymbol{\theta}_1), \ldots, \boldsymbol{y}_N(\boldsymbol{\theta}_4)] \tag{11}$$

where $\boldsymbol{y}_1(\boldsymbol{\theta}) = [Y_0^0, \ldots, Y_1^1]^T(\boldsymbol{\theta})$ for $N = 1$.

The directivity pattern of the decoding procedure equals a hyper-cardioid, which has its zeros at approximately $\pm 109.47^\circ$. Therefore, the three directions always lie in the zeros of the neighboring patterns, and entirely decorrelated signals can be decoded.

The spatial enhancement $\tilde{\boldsymbol{h}}_N(t)$ of the RIR is achieved by re-encoding the signals in a higher-order N with a weighting matrix evaluated in the directions of $\boldsymbol{\Theta}$:

$$\tilde{\boldsymbol{h}}_N(t) = \boldsymbol{Y}_N(\boldsymbol{\Theta})\boldsymbol{s}(t). \tag{12}$$

The spectral decay correction used in 4D-ASDM is the same as for ASDM, see Eqs. (3) and (4).

In comparison with 4D-ASDM, 2DSE2 by Gölles and Zotter [25] also uses a tetrahedral arrangement, yet they differ in several details. 2DSE2 can only decode four explicit directions (two directive signals and two diffuse signals) when two linearly independent directions are detected, otherwise it employs regularization. Moreover, because the tetrahedron is generally irregular and optimally changes its angles depending on the distribution of the two detected directions, its beamforming matrix has a time-varying condition number.

Table 1. RIRs used in this study and their size, reverberation time RT_{60}, calculated critical distance d_c, measurement distance d_m, microphone type, encoded order, and location.

Room	Size (m^3)	RT_{60} (s)	d_c (m)	d_m (m)	Microphone	Order	Location
Lecture hall	500	0.5	1.8	2.0	ZM-1	1	IEM CUBE, Graz, Austria
Small church	2,600	1.4	2.5	11.5	SPS422B	1	St. Andrew's Church, Lyddington, UK
Concert hall	5,700	2.1	3.0	3.0	EM32	1, 4	St. Paul's Concert Hall, Huddersfield, UK
Minster	140,000	7	8.1	23.7	SPS422B	1	York Minster, York, UK

RIR = room impulse response.

Subsequently, crosstalk cancellation is applied on the resulting signals, and the up-mixed HOA RIR is finally treated with Ambisonic widening [32, 33] in order to reduce possible temporal artifacts. By decoding four directions of a regular tetrahedron at all times, the conditioning is always unity so that these post-processing stages are not necessary for 4D-ASDM. In the listening experiment on 2DSE2 [25], speech and music signals were convolved with the enhanced impulse responses for assessing the naturalness of the stimuli. Although this would not reveal strong artifacts for transient signals, it can already be concluded from the results that widening was necessary to mitigate temporal artifacts.

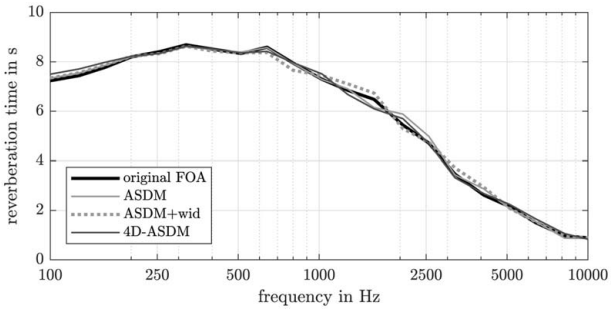

Fig. 3. Reverberation time of original first-order Ambisonics (FOA) room impulse responses (RIRs) and higher-order Ambisonics (HOA) RIRs. ASDM = Ambisonic Spatial Decomposition Method; ASDM+wid = ASDM with widening; 4D-ASDM = Four-Directional ASDM.

2 INSTRUMENTAL VALIDATION

This section compares the different ASDM variants based on technical measures in order to show similarities and differences in the HOA RIRs and validate the improvements of the proposed 4D-ASDM algorithm. As in the subsequent listening experiment, the FOA RIR was enhanced to a fourth-order HOA RIR. All impulse responses were decoded to binaural signals using a magnitude least-squares decoder (magLS) [36, 37] with the respective order. The binaural decoding was chosen because results in [29] indicated that listeners were most sensitive to the graininess artifacts for binaural playback. Because typical measures from room acoustics can only be calculated from a single-channel RIR, left and right channels were summed up before the calculation.

The comparison employs reverberation time, spectrum of the first 80 ms and the later part, and clarity C80 in third-octave bands, as well as the broadband echo density over time as a possible estimate for graininess. For RIR, the FOA measurement of the minster with a very long reverberation time and a measurement distance outside the critical distance was chosen, cf. Table 1, which was also used in the listening experiment.[2,3,4]

The original FOA RIR was included as a reference because, apart from an increase of the spatial resolution, the temporal and spectral shape of the RIR should be unaltered by the up-mixing process. ASDM and the proposed 4D-ASDM were used with the settings as described above. For both, the spectral decay correction was smoothed with a Hann window of 1,024 samples length. ASDM with widening (ASDM+wid) treated the output of ASDM with an Ambisonic widening effect [32, 33]. The effect employed a time constant of 1 ms, an angular spread to 100°, and five taps of the casual-sided impulse response because these settings delivered, overall, the best suppression of graininess artifacts for all four RIRs from Table 1 during informal listening by the authors.

Fig. 3 shows the frequency-dependent reverberation times in third-octave bands for the spatially enhanced HOA RIRs using the three ASDM variants and original FOA RIR. The reverberation times of all HOA RIRs only slightly differ from the values of FOA. The maximum relative deviation between 100 Hz and 10 kHz is 10% for ASDM and 4D-ASDM and 12% for ASDM+wid, which is in the range of the just-noticeable difference [39] and therefore is likely to be inaudible with continuous signals. These small differences indicate that the spectral decay correction works properly in all compared ASDM variants.

The third-octave levels of the first 80 ms indicate the similarity of ASDM and 4D-ASDM to the FOA RIR, except for an increase of up to 2 dB between 800 Hz and 3 kHz, cf. Fig. 4. The spectral differences to ASDM+wid are larger. The first notch frequency of this variant at 500 Hz is directly related to the time constant of 1 ms in the widening effect. Note that the deviation of ASDM+wid is

[2]This is freely available as part of the Open AIR library at https://www.openair.hosted.york.ac.uk/?page_id=683.

[3]This is freely available as part of the 3D Microphone Array Recording Comparison (3D-MARCo) database [38] at https://zenodo.org/record/3477602.

[4]This is freely available as part of the Open AIR library at https://www.openair.hosted.york.ac.uk/?page_id=797.

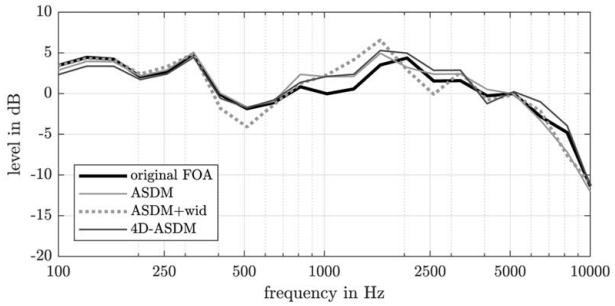

Fig. 4. Early spectrum of original first-order Ambisonics (FOA) room impulse responses (RIRs) and higher-order Ambisonics (HOA) RIRs. ASDM = Ambisonic Spatial Decomposition Method; ASDM+wid = ASDM with widening; 4D-ASDM = Four-Directional ASDM.

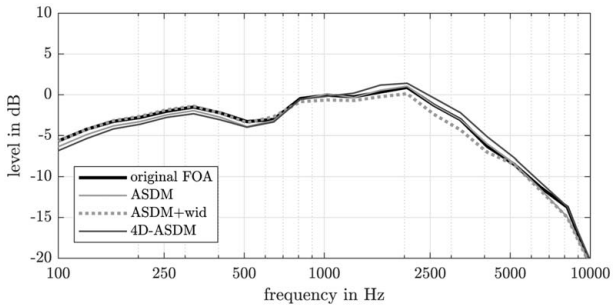

Fig. 5. Late spectrum of original first-order Ambisonics (FOA) room impulse responses (RIRs) and higher-order Ambisonics (HOA) RIRs. ASDM = Ambisonic Spatial Decomposition Method; ASDM+wid = ASDM with widening; 4D-ASDM = Four-Directional ASDM.

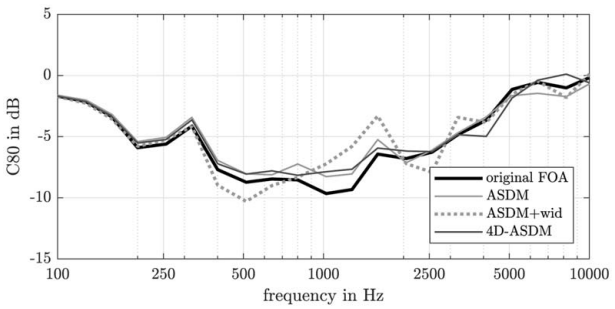

Fig. 6. Clarity C80 of original first-order Ambisonics (FOA) room impulse responses (RIRs) and higher-order Ambisonics (HOA) RIRs. ASDM = Ambisonic Spatial Decomposition Method; ASDM+wid = ASDM with widening; 4D-ASDM = Four-Directional ASDM.

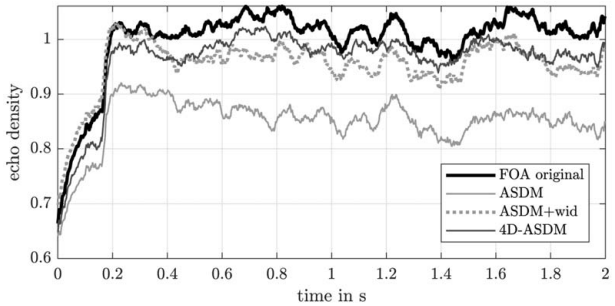

Fig. 7. Echo density of original first-order Ambisonics (FOA) room impulse responses (RIRs) and higher-order Ambisonics (HOA) RIRs. ASDM = Ambisonic Spatial Decomposition Method; ASDM+wid = ASDM with widening; 4D-ASDM = Four-Directional ASDM.

more clearly visible when analyzed in narrower frequency bands.

In the late reverberation, the spectra are more similar and stay within a 1.3-dB absolute difference to the FOA RIR for any ASDM variant, cf. Fig. 5. These values are close to the just-noticeable difference for third-octave band levels [40] and thus are perceptually less relevant than the differences in the early part.

The relation between the early and late parts of the RIR can be visualized by the clarity C80 measure, cf. Fig. 6. Because the spectra of the late parts in all ASDM variants were found to be similar to the original FOA RIR, the differences in C80 follow those from the early part: ASDM and 4D-ASDM yield an increase of up to 1.8 dB between 500 Hz and 2 kHz, whereas the deviation reaches 3.5 dB for ASDM+wid. Although the increase by ASDM and 4D-ASDM lies in the range of or below the just-noticeable difference for C80 [41, 42], the difference to ASDM+wid might be perceptually relevant. However, the known just-noticeable differences are defined for broadband signals and not in third octaves.

The echo density in Fig. 7 has been calculated as proposed in [43] for the first 2 s of the RIRs. Although after 0.2 s, the original FOA RIR has an echo density of around 1 and therefore resembles a Gaussian distribution, the density of the HOA RIR created by ASDM is reduced. Compared to the latter, applying the widening effect in ASDM+wid or 4D-ASDM results in an increase of the echo density more similar to the original FOA RIR.

The results of the instrumental validation of binaurally decoded HOA RIRs showed that the spectral decay correction in all ASDM variants achieves a similar frequency-dependent reverberation time as the original FOA RIR. The same is true for the spectrum of the late reverberation. Although the spectrum of the early part of the RIRs is similar to the FOA RIR for ASDM and 4D-ASDM, the comb filters of the widening effect in ASDM+wid become apparent. This finding is also evident in the results for C80. In contrast, the positive impact of the widening effect becomes obvious for the echo density: There, widening increases the reduced values of ASDM and yields similar results as 4D-ASDM. To summarize, the instrumental validation showed that 4D-ASDM could increase the echo density in comparison to ASDM without the spectral drawbacks of ASDM+wid.

3 LISTENING EXPERIMENTS

A series of listening experiments was conducted to compare the new 4D-ASDM algorithm to existing algorithms

that are intended to enhance the spatial resolution of FOA RIRs. The experiments used measured impulse responses from different rooms, cf. Table 1, and employed playback on headphones and loudspeakers. In the first part, artifacts and coloration were compared to the original FOA RIR for headphone playback. The second part investigated the improvements in spatial clarity in comparison to the original FOA RIR and was played back over loudspeakers. The three attributes have been chosen to cover the perceptual target qualities of spatial enhancements: increase of spatial clarity while avoiding artifacts and coloration. Note that localization accuracy, as a similarity between perceived and target direction, has not been evaluated explicitly. This was decided, because all tested SIRR and ASDM variants share the broadband PIV estimator for the DOA of the first peak in the RIR and thus result in similar spatialization of the direct sound.

Moreover, the FOA RIR measurement was defined as reference for the listeners in the experiment, because the motivation for 4D-ASDM is to create an algorithm that enhances the spatial definition of the RIR without adding artifacts or introducing spectral and temporal changes.

3.1 Experimental Method

For each quality attribute, the algorithms were compared to each other, with the first-order reference in a MUSHRA-like [44] multi-stimulus comparison on a quasi-continuous scale. Note that in the case of artifacts and coloration, first-order playback was the upper limit of the scale (no stimulus could have less artifacts than the first-order playback and a more similar timbre), whereas it marked the center of the scale for spatial clarity, so that it was possible to judge the conditions to have more and less spatial clarity. Audio was played in an endless loop and switching between the stimuli was performed by a fast crossfade.

3.2 Conditions

In total, the conditions of the experiments consisted of four different up-mixing algorithms and a direct HOA RIR measurement with a higher-order microphone array. Because the latter provided a maximum order of four, the algorithms were also set to generate HOA RIR with the same order of four. All signals were rendered in a resolution of 24 bits with a sampling rate of 48 kHz. Binaural examples and an implementation of 4D-ASDM are available online.[5]

ASDM, 4D-ASDM, and ASDM with an Ambisonic widening effect ASDM+wid were used with the same settings as in the instrumental validation of SEC. 2. ASDM+wid treated the output of ASDM with an Ambisonic widening effect [32, 33]. Independent of the individual rooms, the time constant was set to 1 ms, an angular spread to 100°, and 5 taps of the casual-sided impulse response were used. These settings delivered the best suppression of graininess artifacts for all tested RIRs during informal listening by the authors.

SIRR employed the HO-SIRR plug-in[6] with the default settings of a wet/dry ratio of 1, a length of the analysis window of 128 samples, and the activated feature for a broadband first peak. Because SIRR is a channel-based algorithm, the output signals were rendered for a 60-point *t*-design and subsequently encoded into fourth-order Ambisonics using the MultiEncoder plug-in[7] for the headphone experiment. For the second experiment on loudspeakers, the output was directly generated for the 12 loudspeakers of the playback setup.

Fourth-order mic represented an HOA RIR measurement directly done with a fourth-order microphone array. The signals from the 32 capsules were converted into HOA using the Array2SH plug-in.[8] The filters were designed as "Z-style" (corresponding to a loudness-normalization [45] without any further frequency-dependent Ambisonic weighting [46, 47]) with a maximum gain of 15 dB and activated diffuse equalization.

The comparison was done for the RIRs of multiple rooms to cover a wide range of reverberation times and direct-to-reverberant energy ratios, cf. Table 1. In order to keep the number of conditions constant throughout the different rooms, the ASDM+wid condition was not tested for timbral similarity and spatial clarity in the concert hall, because only for this room the fourth-order microphone measurement was available.

Informal listening before the experiment showed that 2DSE2 yielded perceptually inferior results when applied to long FOA RIRs. These were most likely because of its time-varying condition number and regularization. It was therefore excluded to keep the experiment focused to comparison with the most established methods.

3.3 Experiment I: Artifacts and Coloration

The first listening experiment was conducted in November and December 2020 and compared the artifacts and coloration of HOA RIRs up-mixed by the algorithms and a direct HOA RIR measurement to the FOA RIR reference. For the evaluation of artifacts, the most sensitive audio signal was used: the participants had to listen to the pure impulse responses. They were given the definition of artifacts as a sound event or group of sound events that was not perceivable in the reference FOA RIR. For the coloration experiment, the impulse responses had been convolved with continuous pink noise. In order to suppress spatial and temporal cues during the build-up of the impulse responses, the first 10 seconds of the convolved audio signals were cut off.

Because of the pandemic situation, the experiments employed static headphone playback and were done at each participant's home. Binaural playback employed a fourth-order magLS decoder [36, 37] for all HOA conditions and its first-order version for the FOA reference. The magLS

[5]https://phaidra.kug.ac.at/o:126998.

[6]V1.0.3beta, freely available at: https://leomccormack.github.io/sparta-site/docs/plugins/hosirr/.

[7]This is freely available at: https://plugins.iem.at.

[8]V1.6.6, freely available at https://leomccormack.github.io/sparta-site/docs/plugins/sparta-suite/.

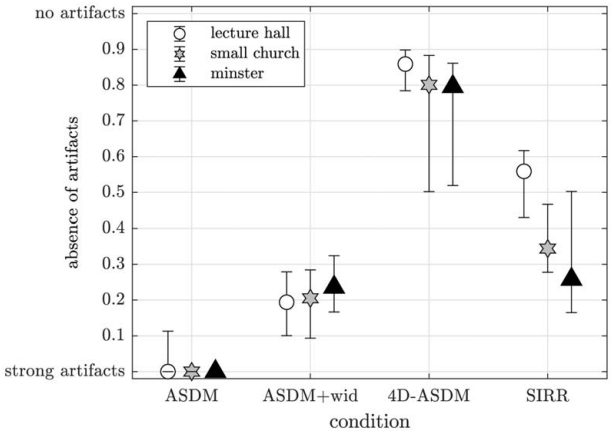

Fig. 8. Median and 95% confidence intervals of results from headphone experiment I on artifacts. ASDM = Ambisonic Spatial Decomposition Method; ASDM+wid = ASDM with widening; 4D-ASDM = Four-Directional ASDM; SIRR = Spatial Impulse Response Rendering.

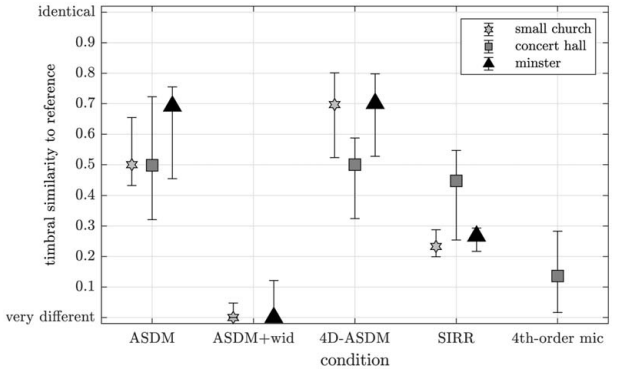

Fig. 9. Median and 95% confidence intervals of results from headphone experiment I on coloration. ASDM = Ambisonic Spatial Decomposition Method; ASDM+wid = ASDM with widening; 4D-ASDM = Four-Directional ASDM; SIRR = Spatial Impulse Response Rendering.

decoder provides a perfect playback of the diffuse part already for first order [48] because of its covariance filters. Static playback was chosen because of no availability of low-latency head trackers for all participants. Nevertheless, informal listening revealed no reduction in sensitivity to artifacts and coloration in comparison to dynamic playback.

A total of 15 listeners participated in the experiment and their ages ranged from 25 to 40 years (mean: 30). All of them were male students or members of the institute, affiliated with audio for at least four years, and experienced in listening experiments on spatial audio.

Fig. 8 shows median values and corresponding confidence intervals of the perceived artifacts for all conditions and three spatial RIRs. Pairwise Wilcoxon signed-rank tests with Bonferroni-Holm correction reveal that for the lecture hall and the small church, all conditions are significantly different ($p \leq 0.013$). The strongest artifacts were perceived using ASDM, followed by ASDM+wid and SIRR. The ranking is the same for the large minster; however, ASDM+wid and SIRR are not significantly different ($p = 0.73$). For all rooms, 4D-ASDM was perceived to produce least artifacts.

Fig. 9 shows the results for the timbral similarity to the FOA reference. For the small church, all conditions were perceived significantly different ($p \leq 0.010$) with the ranking from strongest to least coloration: ASDM+wid, SIRR, ASDM, and 4D-ASDM. In case of the concert hall, the coloration of ASDM, SIRR, and 4D-ASDM was not perceived differently ($p \geq 0.32$); however, all these conditions produced less coloration than the fourth-order microphone ($p \leq 0.017$). Note that the ASDM+wid condition was not tested for the concert hall in order to keep the number of conditions constant. ASDM and 4D-ASDM were not perceived differently for the minster ($p = 0.19$), whereas both produced less coloration than SIRR ($p \leq 0.009$). Again, ASDM+wid was perceived most different ($p \leq 0.0017$).

3.4 Experiment II: Spatial Clarity

The second listening experiment was conducted in March 2021 and evaluated the spatial clarity of HOA RIRs upmixed by the different algorithms and a direct HOA RIR measurement. Because the results of the previous experiment showed that the diffuse equalization in the plug-in attenuated the high frequencies of the higher-order microphone too much, the authors tried to compensate for this with a simple high-shelf filter with +6 dB at 2.8 kHz. Informal listening sessions of the authors showed that the lack of high frequencies reduced the perceived spatial clarity and depth. The term "spatial clarity" is defined similarly to Francombe et al.'s definition of "ease of localization of individual sources" [49]; however, it is also including distinct reflections.

The participants were asked to rate the conditions based on the localizability and compactness of the direct sound and its separation to the early reflections and the late, diffuse reverberation that envelopes them. This experiment was done with loudspeaker playback, because preliminary results indicated that the evaluation of spatial clarity could not be done as well on headphones. Playback used the loudspeaker setup at IEM production studio with 12 loudspeakers: a 7.0 basis, four loudspeakers at 45° elevation, and a voice-of-god loudspeaker directly above the listeners, all at a radius of 2.5 m. Decoding employed fourth-order AllRAD [50] without any additional weighting, except for the SIRR conditions that were directly mapped to the 12 loudspeakers. Because speech signals seemed to be appropriate to evaluate spatial clarity [18], the first 8 s of the EBU's male speech recording [51] convolved with the different RIRs was used.

A total of 13 listeners participated in the experiment and their ages ranged from 24 to 47 years (mean: 31). All of them were male students or members of the institute, with similar listening experience as participants in the first experiment (five of them also took part in the first experiment).

Fig. 10 shows median values and corresponding confidence intervals of the tested conditions with regard to the perceived spatial clarity. For the lecture hall and small

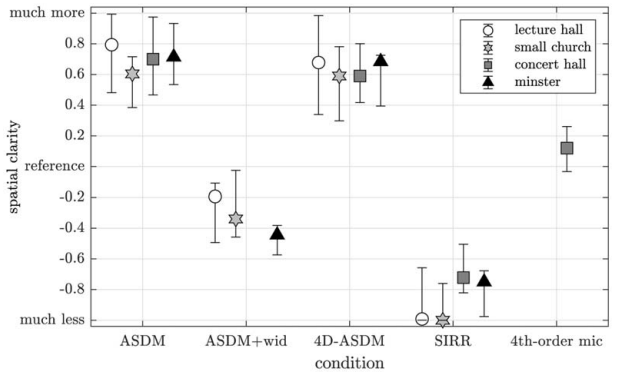

Fig. 10. Median and 95% confidence intervals of results from loudspeaker experiment II on spatial clarity. ASDM = Ambisonic Spatial Decomposition Method; ASDM+wid = ASDM with widening; 4D-ASDM = Four-Directional ASDM; SIRR = Spatial Impulse Response Rendering.

church, all conditions were perceived significantly different ($p \leq 0.041$), except for ASDM and 4D-ASDM ($p \geq 0.54$). These two conditions provided the most spatial clarity, followed by ASDM+wid. ASDM and 4D-ASDM were again not different ($p = 0.3394$) and the best conditions in the concert hall. All other conditions were significantly different ($p \leq 0.0342$). The spatial clarity of the fourth-order microphone was perceived to be between SIRR and ASDM+wid. Note that without the additional high-shelf filter, the results for the fourth-order microphone might have been worse. For the large minster, all conditions were perceived to be significantly different ($p \leq 0.0339$). Although significant, the differences between the medians of ASDM and 4D-ASDM are rather small (0.684 vs. 0.714). Some participants reported that ASDM sounded a little bit more tidy than 4D-ASDM with a slightly more compact direct sound.

3.5 Discussion

The results from the listening experiments show that widening (ASDM+wid) could reduce the artifacts of ASDM, though at the cost of strong coloration and reduced spatial clarity. Coloration is caused by the comb filters of the widening effect. The effect also creates two cluster points at $\pm 50^\circ$ around each DOA. These additional directions distract from the actual direction, resulting in reduced spatial clarity. This is particularly noticeable for the direct sound. One idea to decrease these drawbacks could be the restriction of widening to the later part of the HOA RIR only and letting the direct sound untouched.

Based on these results, SIRR was even more effective in terms of avoiding artifacts. Although the timbral similarity to the reference was much better compared with ASDM+wid, spatial clarity was less. These results seem to partly disagree with the results in [24], whereas SIRR was perceived better than SDM. However, the experiments in the latter paper were different because they were done in an anechoic chamber with a higher number of loudspeakers, employed no measured but simulated FOA RIRs, and were a less advanced version of SDM.

The new 4D-ASDM produced the least artifacts in these experiments. At the same time, coloration was equal or even more identical than for ASDM. Similarly, spatial clarity was the same for both algorithms. The only exception was the large minster, in which ASDM was perceived a little bit, but still significantly, more clear than 4D-ASDM. The new algorithm can be seen as a kind of hybrid between FOA and the original ASDM. By decoding four directions in tetrahedral arrangement, the first order is completely transferred to the up-mixed RIR, whereas ASDM only encodes the omnidirectional component. This yields a more similar timbre at the cost of slightly reduced spatial clarity. The reduction is caused by the three additional, simultaneous directions that preserve the spatial aliasing from the original FOA RIR in the up-mixed RIR.

For the concert hall, HOA RIRs directly measured in the fourth order were compared to the algorithms that employ the FOA RIR measurement. Except for the comb-filter–like ASDM+wid, all algorithms exhibited more timbral similarity to the reference than the direct fourth-order measurement. Participants mainly reported a loss of high frequencies for the latter. Below the aliasing frequency of the microphone array ($\approx$5 kHz), the energy-preserving bandpass approach for the radial filters should ensure a flat frequency response. Above, the diffuse equalization applied in the plug-in seems to have attenuated the high frequencies too much.

With respect to spatial clarity, the fourth-order measurement is outperformed by both ASDM and 4D-ASDM. The poorer clarity can be explained by the limited boost of high orders in the radial filters that produces fourth-order resolution only above 3 kHz. Below 1 kHz, the resolution does not even exceed first order. These frequency limits could be shifted toward lower frequencies by increasing the filter gain at the cost of boosting noise. In contrast, resolution enhancement by up-mixing algorithms produces fourth order over the entire frequency range.

The ranking of the algorithms in the results of the listening experiment was generally independent of the specific RIR. An exception was the similarity in coloration for ASDM, 4D-ASDM, and SIRR in the concert hall. This might be because of the stronger direct sound in the RIR, because the distance between microphone and loudspeaker was within the critical distance.

In terms of coloration, the perceptual differences between ASDM, ASDM+wid, and 4D-ASDM could be well predicted by the early spectra of the RIRs in SEC. 2. Moreover, the echo density could estimate the reduction of the artifacts for ASDM+wid and 4D-ASDM in comparison to ASDM. However, the stronger reduction for 4D-ASDM could not be explained by the echo density indicating that perceived graininess requires other or additional predictors.

Note that all experiments used enhancement of the spatial resolution up to the fourth order. Informal listening indicated that the artifacts of 4D-ASDM did not change noticeably for up-mixing to the third or fifth order.

Because of governmental regulations mitigating the spread of the pandemic, it was not possible to conduct the first listening experiments in a completely controlled en-

vironment. However, it was ensured that all listeners used high-quality headphones in a quiet room. Future listening experiments could also include low-latency head tracking to evaluate the influence of movements on phasing artifacts and fluctuations in timbre.

The FOA RIR was deliberately chosen as a reference, because FOA microphones are the practical choice that provides spatial information, low number of channels, and simple rotation based on a single measurement. This is in contrast to dummy heads and higher-order microphone arrays. However, because of its similarity to ASDM except for the artifacts, it can be expected that 4D-ASDM results of high-enough order could be indistinguishable from dummy head measurements as they were for seventh-order ASDM regarding distance, width, and diffuseness in [36].

4 CONCLUSION

This paper presented an algorithm to enhance the spatial resolution of FOA RIRs to arbitrarily high orders. Whereas the DOA estimation of the prominent direction uses the pseudo-intensity vector as in the ASDM, the new 4D-ASDM algorithm extends ASDM by encoding four directions/signals instead of a single one. The idea of the new algorithm was to reduce the graininess-like artifacts known from ASDM by decoding and encoding of the additional signals. These directions are arranged on a regular tetrahedron around the estimated DOA, and four hyper-cardioid beamformers are used to extract the signals from the FOA RIR.

An instrumental validation using a measured FOA RIR revealed that the spatially enhanced impulse responses of 4D-ASDM could increase the echo density in comparison with ASDM, which can be seen as an indicator for reduced graininess. At the same time, perceptual differences between 4D-ASDM and the original FOA RIRs were estimated to be mostly irrelevant for frequency-dependent reverberation time, spectra of early and late part of the RIR, and the clarity C80.

In two listening experiments, 4D-ASDM was compared to existing algorithms that up-mix measured FOA RIRs to HOA RIRs. The results for an enhancement to the fourth order revealed that 4D-ASDM produced the least artifacts for all evaluated RIRs, whereas the spatial clarity and coloration was comparable to the best of the existing algorithms. The experiments also showed that 4D-ASDM outperformed impulse responses directly measured with a fourth-order microphone array with regard to spatial clarity and timbral similarity to the FOA RIR reference.

5 ACKNOWLEDGMENT

The authors thank all listeners for their participation in the experiments and Franz Zotter for fruitful discussions during the development of the algorithm and his helpful suggestions improving the efficiency of the tetrahedron alignment method. Furthermore, the authors highly appreciate the valuable comments of the anonymous reviewers, which helped to illustrate the key findings of this paper.

6 REFERENCES

[1] D. H. Cooper and T. Shiga, "Discrete-Matrix Multichannel Stereo," *J. Audio Eng. Soc.*, vol. 20, no. 5, pp. 346–360 (1972 Jun.).

[2] P. B. Fellgett, "Ambisonic Reproduction of Directionality in Surround-Sound Systems," *Nature*, vol. 252, no. 5484, pp. 534–538 (1974 Dec.).

[3] 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.

[4] P. G. Craven and M. A. Gerzon, "Coincident Microphone Simulation Covering Three Dimensional Space and Yielding Various Directional Outputs," US Patent 4,042,779 (1977 Aug.).

[5] D. G. Malham and A. Myatt, "3D Sound Spatialization Using Ambisonic Techniques," *Comput. Music J.*, vol. 19, no. 4, pp. 58–70 (1995 Winter). http://dx.doi.org/10.2307/3680991.

[6] M. A. Poletti, "The Design of Encoding Functions for Stereophonic and Polyphonic Sound Systems," *J. Audio Eng. Soc.*, vol. 44, no. 11, pp. 948–963 (1996 Nov.).

[7] D. G. Malham, "Higher Order Ambisonic Systems for the Spatialisation of Sound," in *Proceedings of the International Comupter Music Conference*, vol. 1999, pp. 484–487 (Beijing, China) (1999 Oct.).

[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 AES 16th International Conference: Spatial Sound Reproduction* (1999 Mar.), paper 16-025.

[9] J. Daniel, J.-B. Rault, and J.-D. Polack, "Acoustic Properties and Perceptive Implications of Stereophonic Phenomena," in *Proceedings of the AES 16th International Conference: Spatial Sound Reproduction* (1999 Mar.), paper 16-008.

[10] E. Benjamin, A. Heller, and R. Lee, "Localization in Horizontal-Only Ambisonic Systems," presented at the *121th Convention of the Audio Engineering Society* (2006 Oct.), paper 6967.

[11] S. Braun and M. Frank, "Localization of 3D Ambisonic Recordings and Ambisonic Virtual Sources," in *Proceedings of the 1st International Conference on Spatial Audio*, pp. 2–11 (Detmold, Germany) (2011 Nov.).

[12] E. Bates, F. Boland, D. Furlong, and G. Kearney, "A Comparative Study of the Performance of Spatialization Techniques for a Distributed Audience in a Concert Hall Environment," in *Proceedings of the AES 31st International Conference: New Directions in High Resolution Audio* (2007 Jun.), paper 13.

[13] M. Frank, F. Zotter, and A. Sontacchi, "Localization Experiments Using Different 2D Ambisonics Decoders," in *Proceedings of the 25th International Conven-*

tion Tonmeistertagung (VDT), pp. 1–9 (Leipzig, Germany) (2008 Nov.).

[14] S. Bertet, J. Daniel, E. Parizet, and O. Warusfel, "Investigation on Localisation Accuracy for First and Higher Order Ambisonics Reproduced Sound Sources," *Acta Acust. united Acust.*, vol. 99, no. 4, pp. 642–657 (2013 Jul./Aug.). https://doi.org/10.3813/AAA.918643.

[15] P. Stitt, S. Bertet, and M. van Walstijn, "Perceptual Investigation of Image Placement With Ambisonics for Non-Centred Listeners," in *Proceedings of the 16th International Conference on Digital Audio Effects (DAFx)*, paper 32 (Maynooth, Ireland) (2013 Sep.).

[16] P. Stitt, S. Bertet, and M. van Walstijn, "Off-Centre Localisation Performance of Ambisonics and HOA for Large and Small Loudspeaker Array Radii," *Acta Acust. united Acust.*, vol. 100, no. 5, pp. 937–944 (2014 Oct.). https://doi.org/10.3813/AAA.918773.

[17] M. Frank and F. Zotter, "Exploring the Perceptual Sweet Area in Ambisonics," presented at the *142nd Convention of the Audio Engineering Society* (2017 May), paper 9727.

[18] M. Frank and F. Zotter, "Spatial Impression and Directional Resolution in the Reproduction of Reverberation," in *Proceedings of the Fortschritte der Akustik (DAGA)*, pp. 1304–1307 (Aachen, Germany) (2016 Mar.).

[19] J. Daniel and S. Moreau, "Further Study of Sound Field Coding With Higher Order Ambisonics," presented at the *116th Convention of the Audio Engineering Society* (2004 May), paper 6017.

[20] B. Rafaely, *Fundamentals of Spherical Array Processing*, Springer Topics in Signal Processing, vol. 16 (Springer, Cham, Switzerland, 2019).

[21] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial Decomposition Method for Room Impulse Responses," *J. Audio Eng. Soc.*, vol. 61, no. 1/2, pp. 17–28 (2013 Jan.).

[22] M. Zaunschirm, M. Frank, and F. Zotter, "BRIR Synthesis Using First-Order Microphone Arrays," presented at the *144th Convention of the Audio Engineering Society* (2018 May), paper 9944.

[23] V. Pulkki, J. Merimaa, and T. Lokki, "Reproduction of Reverberation With Spatial Impulse Response Rendering," presented at the *116th Convention of the Audio Engineering Society* (2004 May), paper 6057.

[24] L. McCormack, V. Pulkki, A. Politis, O. Scheuregger, and M. Marschall, "Higher-Order Spatial Impulse Response Rendering: Investigating the Perceived Effects of Spherical Order, Dedicated Diffuse Rendering, and Frequency Resolution," *J. Audio Eng. Soc.*, vol. 68, no. 5, pp. 338–354 (2020 May). https://doi.org/10.17743/jaes.2020.0026.

[25] L. Gölles and F. Zotter, "Directional Enhancement of First-Order Ambisonic Room Impulse Responses by the 2+2 Directional Signal Estimator," in *Proceedings of the 15th International Conference on Audio Mostly*, pp. 38–45 (Graz, Austria) (2020 Sep.). https://doi.org/10.1145/3411109.3411131.

[26] J. Ahrens, "Perceptual Evaluation of Binaural Auralization of Data Obtained From the Spatial Decomposition Method," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 65–69 (New Paltz, NY) (2019 Oct.). https://doi.org/10.1109/WASPAA.2019.8937247.

[27] M. Zaunschirm, M. Frank, and F. Zotter, "Binaural Rendering With Measured Room Responses: First-Order Ambisonic Microphone vs. Dummy Head," *Appl. Sci.*, vol. 10, no. 5, paper 1631 (2020 Feb.). https://doi.org/10.3390/app10051631.

[28] A. Pawlak, H. Lee, A. Mäkivirta, and T. Lund, "Subjective Evaluation of Spatial Analysis and Synthesis Methods Using Different Microphone Arrays," in *Proceedings of the International Conference on Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, pp. 1–7 (Bologna, Italy) (2021 Sep.). https://doi.org/10.1109/I3DA48870.2021.9610905.

[29] N. Meyer-Kahlen, S. J. Schlecht, and T. Lokki, "Perceptual Roughness of Spatially Assigned Sparse Noise for Rendering Reverberation," *J. Acoust. Soc. Am.*, vol. 150, no. 5, pp. 3521–3531 (2021 Nov.). https://doi.org/10.1121/10.0007048.

[30] S. V. Amengual Garí, J. M. Arend, P. T. Calamia, and P. W. Robinson, "Optimizations of the Spatial Decomposition Method for Binaural Reproduction," *J. Audio Eng. Soc.*, vol. 68, no. 12, pp. 959–976 (2020 Dec.). https://doi.org/10.17743/jaes.2020.0063.

[31] S. Tervo, J. Pätynen, N. Kaplanis, et al., "Spatial Analysis and Synthesis of Car Audio System and Car Cabin Acoustics With a Compact Microphone Array," *J. Audio Eng. Soc.*, vol. 63, no. 11, pp. 914–925 (2015 Nov.). https://doi.org/10.17743/jaes.2015.0080.

[32] F. Zotter, M. Frank, M. Kronlachner, and J.-W. Choi, "Efficient Phantom Source Widening and Diffuseness in Ambisonics," in *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, pp. 69–74 (Berlin, Germany) (2014 Apr.). https://doi.org/10.14279/depositonce-12.

[33] F. Zotter and M. Frank, "Phantom Source Widening by Filtered Sound Objects," presented at the *142nd Convention of the Audio Engineering Society* (2017 May), paper 9728.

[34] J. Merimaa and V. Pulkki, "Spatial Impulse Response Rendering I: Analysis and Synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 12, pp. 1115–1127 (2005 Dec.).

[35] R. H. Hardin and N. J. A. Sloane, "Spherical Designs," http://neilsloane.com/sphdesigns/ (accessed Jan. 7, 2021).

[36] 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.

[37] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in *Proceedings of the Fortschritte der Akustik (DAGA)*, pp. 339–342 (Munich, Germany) (2018 Mar.).

[38] H. Lee and D. Johnson, "3D Microphone Array Recording Comparison (3D-MARCo)," *Zenodo* (2019 Oct.). https://doi.org/10.5281/zenodo.3477602.

[39] Z. Meng, F. Zhao, and M. He, "The Just Noticeable Difference of Noise Length and Reverberation Perception," in *Proceedings of the IEEE International Symposium on Communications and Information Technologies*, pp. 418–421 (Bangkok, Thailand) (2006 Oct.). https://doi.org/10.1109/ISCIT.2006.339980.

[40] M. Karjalainen, E. Piirilä, A. Järvinen, and J. Huopaniemi, "Comparison of Loudspeaker Equalization Methods Based on DSP Techniques," *J. Audio Eng. Soc.*, vol. 47, no. 1/2, pp. 14–31 (1999 Feb.).

[41] F. Martellotta, "The Just Noticeable Difference of Center Time and Clarity Index in Large Reverberant Spaces," *J. Acoust. Soc. Am.*, vol. 128, no. 2, pp. 654–663 (2010 Aug.). https://doi.org/10.1121/1.3455837.

[42] M. C. Vigeant, R. D. Celmer, C. M. Jasinski, et al., "The Effects of Different Test Methods on the Just Noticeable Difference of Clarity Index for Music," *J. Acoust. Soc. Am.*, vol. 138, no. 1, pp. 476–491 (2015 Jul.). https://doi.org/10.1121/1.4922955.

[43] H. P. Tukuljac, V. Pulkki, H. Gamper, et al., "A Sparsity Measure for Echo Density Growth in General Environments," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1–5 (Brighton, UK) (2019 May). https://doi.org/10.1109/ICASSP.2019.8682878.

[44] ITU-R, "Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems," *ITU-R Recommendation BS.1534-3* (2015 Oct.).

[45] R. Baumgartner, H. Pomberger, and M. Frank, "Practical Implementation of Radial Filters for Ambisonic Recordings," in *Proceedings of the 1st International Conference on Spatial Audio*, paper 22199 (Detmold, Germany) (2011 Nov.).

[46] F. Zotter, M. Zaunschirm, M. Frank, and M. Kronlachner, "A Beamformer to Play With Wall Reflections: The Icosahedral Loudspeaker," *Comput. Music J.*, vol. 41, no. 3, pp. 50–68 (2017 Sep.). https://doi.org/10.1162/comj_a_00429.

[47] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*, Springer Topics in Signal Processing, vol. 19 (Springer, Cham, Switzerland, 2019).

[48] D. Perinovic and M. Frank, "Spatial Resolution of Diffuse Reverberation in Binaural Ambisonic Playback," in *Proceedings of the Fortschritte der Akustik (DAGA)*, pp. 1622–1624 (Vienna, Austria) (2021 Aug.).

[49] J. Francombe, T. Brookes, and R. Mason, "Evaluation of Spatial Audio Reproduction Methods (Part 1): Elicitation of Perceptual Differences," *J. Audio Eng. Soc.*, vol. 65, no. 3, pp. 198–211 (2017 Mar.). https://doi.org/10.17743/jaes.2016.0070.

[50] F. Zotter and M. Frank, "All-Round Ambisonic Panning and Decoding," *J. Audio Eng. Soc.*, vol. 60, no. 10, pp. 807–820 (2012 Oct.).

[51] EBU, "Sound Quality Assessment Material Recordings for Subjective Tests: EBU SQAM CD," https://tech.ebu.ch/publications/sqamcd (2008 Sep.).

THE AUTHORS

Elias Hoffbauer

Matthias Frank

Elias Hoffbauer completed his Diploma in electrical and audio engineering at the University of Music and Performing Arts (KUG) and the University of Technology (TU Graz) in 2021. During his Master's studies, his interests in research were focused on the reproduction of spatial audio and signal processing algorithms for Ambisonic applications. In his project thesis, he investigated possibilities to improve algorithms for spatial enhancement of Ambisonic impulse responses, which led to the 4D-ASDM approach presented in this paper. His Master's thesis dealt with the evaluation of surround setups by adapting classical quality measures used in room acoustics to multi-loudspeaker scenarios. Currently, he is working as a consultant in the room acoustics department of the engineering bureau Müller-BBM.

•

Matthias Frank studied electrical and audio engineering at University of Technology and University of Music and Performing Arts in Graz. After receiving his Diploma in 2009, he joined the Institute of Electronic Music and Acoustics in Graz as a teacher and researcher. In 2013, he finished his Ph.D. and received the Award of Excellence from the Austrian Federal Ministry of Science and Research. His research focuses on perception, recording, playback, and production of spatial audio. Besides his interest in spatial audio, he works on product sound quality and musical acoustics. He was involved in the organization of the 2010 International Conference on Digital Audio Effects (DAFx-10) and co-chaired International Conference on Spatial Audio (ICSA) 2015 and 2017. Starting from 2017, he has been one of the organizers of the annual Student 3D Audio Production Competitions. He regularly plays drums in bands and percussion in various wind and symphony orchestras. Matthias is a member of the Audio Engineering Society and German Acoustical Society.