Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data

Citation for the original published paper (version of record):
Lübeck, T., Helmholz, H., Arend, J. et al (2020)
Perceptual Evaluation of Mitigation Approaches of Impairments due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data
AES: Journal of the Audio Engineering Society, 68(6): 428-440
http://dx.doi.org/10.17743/jaes.2020.0038

(article starts on next page)

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 June.).
DOI: https://doi.org/10.17743/jaes.2020.0038

Perceptual Evaluation of Mitigation Approaches of Impairments Due to Spatial Undersampling in Binaural Rendering of Spherical Microphone Array Data

TIM LÜBECK,[1, 2] *AES Student Member*, (tim.luebeck@th-koeln.de)
HANNES HELMHOLZ,[3] *AES Student Member*, (hannes.helmholz@chalmers.se)
JOHANNES M. AREND,[1, 2] *AES Student Member*, (johannes.arend@th-koeln.de)
CHRISTOPH PÖRSCHMANN,[1] *AES Associate Member*, (christoph.poerschmann@th-koeln.de)
AND
JENS AHRENS,[3] *AES Member* (jens.ahrens@chalmers.se)

[1]*TH Köln – University of Applied Sciences, Cologne, Germany*
[2]*Technical University of Berlin, Berlin, Germany*
[3]*Chalmers University of Technology, Gothenburg, Sweden*

Spherical microphone arrays (SMAs) are widely used to capture spatial sound fields that can then be rendered in various ways as a virtual acoustic environment (VAE) including headphone-based binaural synthesis. Several practical limitations have a significant impact on the fidelity of the rendered VAE. The finite number of microphones of SMAs leads to spatial undersampling of the captured sound field, which, on the one hand, induces spatial aliasing artifacts and, on the other hand, limits the order of the spherical harmonics (SH) representation. Several approaches have been presented in the literature that aim to mitigate the perceptual impairments due to these limitations. In this article, we present a listening experiment evaluating the perceptual improvements of binaural rendering of undersampled SMA data that can be achieved using state-of-the-art mitigation approaches. In particular, we examined the Magnitude Least-Squares algorithm, the Bandwidth Extraction Algorithm for Microphone Arrays, Spherical Head Filters, SH Tapering, and a newly proposed equalization filter. In the experiment, subjects rated the perceived differences between a dummy head and the corresponding SMA auralization. We found that most mitigation approaches lead to significant perceptual improvements, even though audible differences to the reference remain.

0 INTRODUCTION

The increasing number of virtual and augmented reality applications creates the demand for high-fidelity virtual acoustic environments (VAEs). These can be created based on either simulations or captured data. A common method for capturing and auralizing spatial sound fields is the measurement of impulse responses with a dummy head. Such impulse responses represent the acoustic path from the sound source to the ears of a listener and are referred to as either head-related impulse responses (HRIRs), when representing anechoic conditions, or binaural room impulse responses (BRIRs), when representing nonanechoic conditions. Interactive VAEs that adapt to the listener's head orientation can be realized with head tracking based on sequential dummy head measurements on adequately high-resolution grids of head orientations.

However, this technique of sound field capture makes it impossible to realize auralizations of dynamic scenarios such as music concerts. An alternative to the time-consuming sequential dummy head measurements is a continuous capture of the sound field, including all dynamic changes. By means of a distribution of sensors in the region of interest such as a spherical microphone array (SMA), the original sound field can be reconstructed.

VAEs can be rendered to a listener with different loudspeaker-based reproduction methods such as Ambisonics [1] or wave-field synthesis [2]. In this paper, we focus on headphone-based binaural reproduction. Binaural reproduction computes the signals that would arise at the listener's ears if he/she were exposed to the sound field that the microphone array captured. This is performed by virtually exposing the listener's head to the sound field that impinges on the SMA. The method utilizes a spherical harmonics (SH) representation of the sound field as well as of a set of HRIRs (see, e.g., [3, 4]).

The physical accuracy that can be achieved with SMAs is limited, mainly due to the employment of a finite number of microphones as opposed to the continuous distribution that the theory assumes. This leads to spatial undersampling of the captured sound field, which induces spatial aliasing and limits the order of the SH representation of the captured sound field that can be obtained. The order of the SH presentation directly corresponds to the spatial resolution of the sound field. Both phenomena can lead to audible impairments.

Previous studies such as [3, 4] compared binaural auralizations based on SMA data to a reference based on dummy head measurements of the exact same scenario. It was shown that, evidently, higher-order renderings yield more similarity to the dummy head auralizations. In direct comparison, renderings with representations below order 8 were perceived as noticeably different to the synthesis with dummy head data. Furthermore, listening experiments [4] showed that these differences are evoked mainly by high-frequency components, which are those that are primarily affected by spatial undersampling.

In recent years, several approaches to mitigate such impairments in binaural rendering of undersampled SMA data have been proposed. Although most of these approaches have been evaluated independently, up to now, no comparative listening experiment of all these methods has been made. We present a listening experiment comparing the perceptual improvements that can be achieved with the state-of-the-art undersampling mitigation approaches.

The paper is organized as follows: Sec. 1 presents the fundamentals of analyzing spatial sound fields by means of SMAs and binaural rendering. We describe the artifacts introduced by spatial undersampling and the state-of-the-art rendering approaches to mitigate these artifacts. Sec. 2 introduces the materials utilized in the comparative instrumental and perceptual evaluation in Sec. 3 and Sec. 4. The results are further discussed in Sec. 4 and completed with conclusions in Sec. 5.

1 THEORY

This section presents a conceptual overview on the binaural rendering of a sound field captured by an SMA. We refer the reader to [3, 5–7] for more detailed treatments.

1.1 Binaural Rendering of Spherical Microphone Array Data

Let $S(r, \phi, \theta, \omega)$ be the sound pressure distribution on a spherical surface Ω (for example, an SMA) with respect to the radius r, the azimuth angle ϕ ranging from 0 to 2π, the colatitude θ ranging from 0 to π, and the angular frequency $\omega = 2\pi f$, whereby f denotes the temporal frequency. Any sound pressure distribution whose mathematical representation fulfills the wave equation can be transformed into the SH domain using the spatial Fourier transform (SFT) [6]

$$S_{nm}(r, \omega) = \int_{\Omega} S(r, \phi, \theta, \omega)\, Y_n^m(\theta, \phi)^*\, \mathrm{d}A_{\Omega}\,, \tag{1}$$

where $Y_n^m(\theta, \phi)$ denote a set of SH basis functions, $(\cdot)^*$ the complex conjugate, and $\int_{\Omega}(\cdot)\, \mathrm{d}A_{\Omega} = \int_0^{2\pi}\int_0^{\pi}(\cdot)\, \sin\theta\, \mathrm{d}\theta\, \mathrm{d}\phi$ the integration over the surface of the sphere. The SHs are orthogonal basis functions of the sphere and form a complete set of solutions of the angular component of the Helmholtz equation. Furthermore, any sound field on the spherical surface can be described as a continuum of infinitely many plane waves impinging the sphere from all possible directions. The plane wave coefficients $D(\phi_d, \theta_d, \omega)$ can be computed from the SH coefficients $S_{nm}(r, \omega)$ of the sound field on the surface of an acoustically rigid sphere as [5]

$$D(\phi_d, \theta_d, \omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} d_n\, S_{nm}(r, \omega)\, Y_n^m(\phi_d, \theta_d)\,. \tag{2}$$

The term d_n denotes a set of radial filters that compensate for the scattering effects on the surface of the sphere. These filters can exhibit very high amplification gains that need to be restricted in practical implementations. The influence of the radial filters has been discussed extensively, e.g., [3, pp. 90–118], [8], and [5, pp. 34–38].

A head-related transfer function (HRTF) $H(\phi, \theta, \omega)$ can be interpreted as the spatiotemporal transfer function of a given broadband plane wave to the listeners' ears. Note that we omit differentiating between left-ear and right-ear HRTFs, as well as left-ear and right-ear binaural signals in the mathematical formulations for convenience. The resulting binaural signals $Y(\omega)$ can hence be calculated by weighting all plane wave coefficients $D(\phi_d, \theta_d, \omega)$ of the sound field with the corresponding HRTF $H(\omega)$ of that direction and integrating them over all possible propagation directions as

$$Y(\omega) = \frac{1}{4\pi} \int_{\Omega} H(\phi, \theta, \omega)\, D(\phi, \theta, \omega)\, \mathrm{d}A_{\Omega}\,. \tag{3}$$

Transforming the HRTFs into the SH domain as well and exploiting the orthogonality property of the SHs allows to resolve the integral in Eq. (3) and compute the binaural signals as

$$Y(\omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} d_n\, S_{nm}(\omega, r)\, H_{nm}(\omega)\,. \tag{4}$$

The exact formulation of Eq. (4) depends on the particular definition of the employed SH basis functions [7, pp. 7].

1.2 Spatial Undersampling

Sec. 1.1 assumed a continuous pressure distribution on the surface of the SMA. Real-world SMAs, on the other hand, employ a discrete and finite set of sound pressure sensors. This leads to spatial undersampling of the sound field and audible impairments in the synthesized binaural VAE. These impairments can be divided into two categories, namely spatial aliasing and SH order truncation.

1.2.1 Spatial Aliasing

When sampling continuous-time signals, components above the Nyquist frequency cannot be deduced reliably and are aliased to lower frequency components [9]. Analogously, when spatially sampling space-continuous sound fields at discrete locations, higher spatial modes cannot be deduced reliably and are mirrored into lower modes. This results in spatial ambiguities and changes in the time-frequency spectrum.

In contrast to continuous-time signals that can exhibit a limited bandwidth, sound fields are not band-limited in their modal order. Spatial aliasing is therefore apparent over the entire time-frequency spectrum. There is a temporal spatial aliasing frequency f_{A}

$$f_A = \frac{N_{grid}\, c}{2\pi r}\,, \tag{5}$$

above which the spatial aliasing artifacts increase rapidly [10]. Thereby, c denotes the speed of sound and N_{grid} the maximum resolvable SH order of the sampling scheme. In other words, spatial aliasing artifacts are very small in magnitude below f_{A}.

1.2.2 SH Order Truncation

The second fundamental impairment of undersampled SMA data is the truncation of the natural SH order. The integral in Eq. (1) has to be discretized to Q points, corresponding to the microphone locations Ω_q. This leads to the discrete SFT

$$S_{nm}(\omega) = \sum_{q=1}^{Q} w_q\; S(\Omega_q, \omega)\; Y_n^m(\Omega_q)^*\,. \tag{6}$$

The weights w_q ensure orthogonality of the SH basis functions. The coefficients $S_{nm}(\omega)$ can be obtained only for orders $n \leq N_{\mathrm{grid}}$.

1.2.3 Consequences of Spatial Undersampling

Spatial aliasing depends on the density of the SMA microphone sampling scheme, whereas order truncation solely depends on the SH order. Even though both phenomena affect similar time-frequency regions, they exhibit different and sometimes even contrary effects. The compound error of spatial aliasing and truncation was termed “sparsity error” in [11]. The authors investigated the sparsity error with a focus on binaural auralization, which is summarized in the following.

Fig. 1 illustrates the energy distribution of HRTFs in different SH modes as a function of time frequency. It can be seen that the higher SH modes contain a significant fraction of the energy at higher frequencies. Order truncation leads to a loss of spatial details in the according frequency range, which may result in an impairment of the interaural level differences (ILDs), among other things. Moreover, the hard truncation of the SH coefficients leads to side lobes of the plane wave components from Eq. (2), which can also impair ILDs [12]. A side effect is the circumstance that the order truncation attenuates the signal at high time-frequencies to a considerable extent. This can be seem in Fig. 2 (bottom), where we used the HRIRs of the Neumann KU100 dummy head to calculate binaural signals according to Eq. (4) resulting from a simulated broadband plane wave impinging on virtual SMAs from ($\phi = 0°,\ \theta = 90°$). Both ear signals depicted in Fig. 2 (bottom) are based on a 2,702-node grid so that they exhibit a negligible amount of spatial aliasing. The attenuation of the magnitude is apparent at frequencies above 4 kHz.

Spatial aliasing constitutes spatial ambiguities, as information from higher modal orders appears in lower-order modes. This may likewise impair interaural cues. As a side effect, it results in an increase of the level at higher time-frequencies and therefore produces a high-shelf effect on the time-frequency response, as illustrated in Fig. 2 (top). The black curve depicts the left-ear binaural room transfer function (BRTF) based on a 2,702-node Lebedev grid SMA and thus contains no considerable spatial aliasing. The dashed curve is based on an 86-node grid and is affected by spatial aliasing that manifests in this representation as

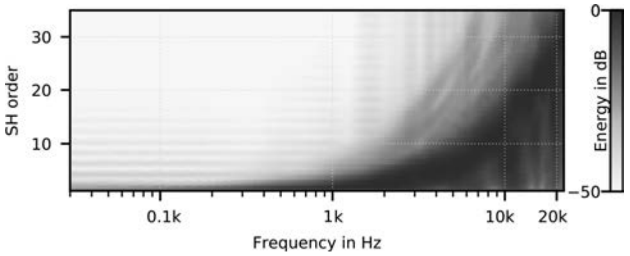


Fig. 1. Normalized logarithmic energy distribution of the head-related transfer functions (HRTFs) of the employed Neumann KU100 dummy head over frequency and SH order n. The color encodes the energy ranging from −50 dB normalized to the maximum values for each frequency bin.

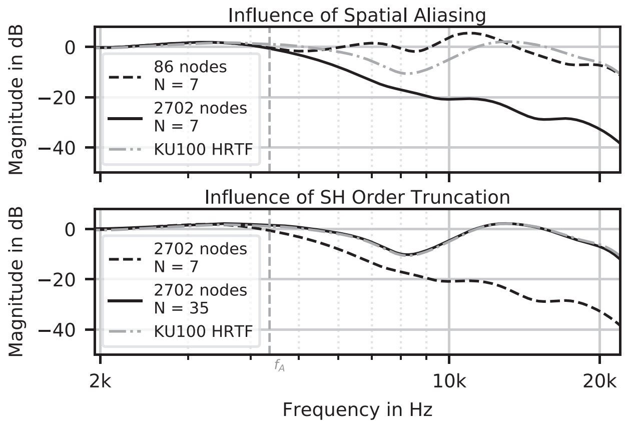


Fig. 2. Binaural signals obtained from the KU100 head-related impulse response (HRIR) set for a simulated plane wave impinging on simulated arrays of varying numbers of sampling nodes from the direction ($\phi = 0°,\ \theta = 90°$) with a maximum permitted radial filter gain of 40 dB. All curves are 1/3-octave-smoothed.

an increase of the magnitude at frequencies above 4 kHz. Both signals were computed for the same SH order $N = 7$ to ensure identical truncation effects.

The left-ear measured KU100 HRIR is depicted in Fig. 2 for reference (grey dash-dotted curve). It can be seen that the SMA rendering up to $N = 35$ based on the 2,702 node grid (bottom) exactly matches the measured HRIR. The top figure shows that the high pass of spatial aliasing and the low-pass of order truncation cancel out each other. However, significant deviations from the reference persist.

1.3 Mitigation Approaches

A number of approaches to mitigate the impairment due to spatial undersampling in binaural rendering of SMA data have been presented in the past years. This section outlines the basic concepts of a selection of approaches. The same approaches are covered in the listening experiment that we conducted.

1.3.1 Bandwidth Extension Algorithm for Microphone Arrays

The Bandwidth Extension Algorithm for Microphone Arrays (BEMA) [13, 3] synthesizes the sound field SH coefficients of the higher time-frequency bands. It thus addresses the spatial ambiguities as well as the influence on the time-frequency transfer function. For this, spatial and spectral properties of the reliably obtainable frequency bands are acquired. The spatial energy distribution is extracted from the SH coefficients of frequency bands below the spatial aliasing frequency f_A as given by Eq. (5). The total energy of the higher frequencies is derived from an additional omnidirectional center microphone, ideally located in the center of the array. This approach is based on the observation that most relevant sound fields exhibit a smooth energy distribution in adjacent frequency bands.

The synthesis of the BEMA SH coefficients can be mathematically expressed as

$$S_{nm,\,\mathrm{BEMA}} = \underbrace{\frac{1}{d_n(\frac{\omega}{c}r)}\,I_{nm}}_{\text{spatial information}} \cdot \underbrace{C_0(\omega)}_{\text{spectral information}} , \tag{7}$$

where C_0 is the energy normalized frequency domain signal of the center microphone and I_n^m the normalized so-called spatiotemporal image

$$I_{nm} = \frac{1}{W}\sum_{\mu=1}^{W} d_n\left(\frac{\omega_a-\mu}{c}r\right) S_{nm}(\omega_a-\mu) \tag{8}$$

with the averaging width W and the cut-off frequency ω_a, above which the BEMA synthesis is effective. The choice of W and ω_a defines the frequency bands denoted as source bands that are included in the calculation of I_{nm}.

1.3.2 Magnitude Least-Squares

Magnitude Least-Squares (MagLS) [14] is a method for reducing the impact of SH order truncation. This method premodifies the HRTF set in such a way that the energy in higher SH modes is reduced without notably decreasing the perceptual quality. If such higher modes are then removed due to truncation, the error becomes less significant. This modification is an advancement of the time alignment approach [15]. According to the duplex theory [16], interaural time differences become perceptually less important at high frequencies than ILDs. However, most of the energy in higher modes is caused by rapid phase changes towards higher frequencies. Thus, removing the linear phase at higher frequencies will decrease the energy in higher modes without significantly modifying the ILDs. The MagLS algorithm not only removes the linear phase slope but also minimizes the distance to the magnitudes of a reference HRTF set at higher frequencies. This is achieved by solving the least-squares problem in an iterative procedure according to

$$\min_{H_{nm}(\omega)} \sum_{q=1}^{Q} [\,|Y_n^m(\Omega_q)\,\mathbf{H}_{nm}(\omega)| - H(\Omega_q,\omega)]^2 . \tag{9}$$

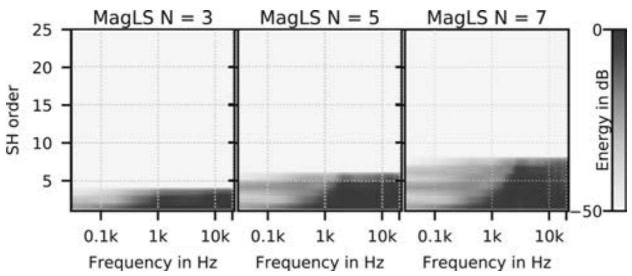


Fig. 3. Normalized logarithmic energy distribution of the HRTFs of the employed Neumann KU100 dummy head over frequency and SH order after MagLS preprocessing for the target orders $N = (3, 5, 7)$. The color encodes the energy ranging from −50 dB normalized to the maximum values for each frequency bin. It can be seen that MagLS modifies the information at low orders to account for the information that was removed from the higher orders.

The energy reduction in higher modes is depicted in Fig. 3. In contrast to the untreated HRTF set in Fig. 1, the energy in higher modes completely vanishes.

1.3.3 Spectral Equalization

To compensate for the modification of the time-frequency transfer function of the binaural signals, global equalization filters have been proposed. These filters are directly applied to the binaural signals and thus equalize every direction equally.

The so-called Spherical Head Filters (SHFs) [17] have been developed to compensate for the low-pass effect of SH order truncation. The authors determine the systematic magnitude deviation of order-truncated HRTFs based on a spherical head model and propose a global compensation filter without taking the effect of spatial aliasing into account. Applying these filters, which are depicted in Fig. 4, to all directions equally results in improved frequency responses for frontal directions but can make the deviations for lateral and especially contralateral sound incidents even larger.

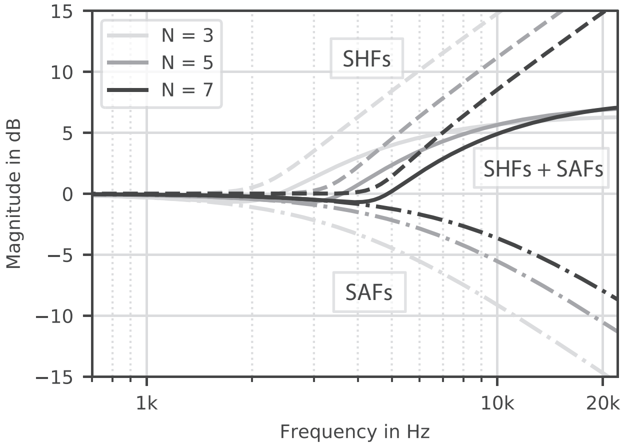


Fig. 4. Spherical Head Filters (dashed line), Spatial Aliasing Filter (dot-dashed line), and the combination of both (solid line) for orders $N = (3, 5, 7)$. Note that the Spherical Head Filters are designed with respect to the current SH rendering order N, the Spatial Aliasing Filters with respect to the maximum order N_{grid} that the sampling scheme permits. We assume $N = N_{\text{grid}}$ here.

An equalization approach to compensate for the high-shelf boost effect of spatial aliasing was proposed in [3]. The authors computed the deviation of dummy head measured room transfer functions to corresponding array renderings. For the array renderings, HRTFs with limited modal resolution were used to design the filters under negligible truncation errors [3, pp. 83], [18]. It was found that for diffuse sound fields, the average logarithmic deviations between dummy head transfer functions and array renderings follows a +6 dB/octave slope above f_A. Thus, aliasing compensation filters can be deployed generically using first order low-pass filters with the cut-off at f_A.

Informal listening showed that the low-pass effect of the truncation error is more noticeable than the high-shelf boost of spatial aliasing and solely applying the low-pass filter to compensate for aliasing yields no considerable perceptual benefit. We therefore combined the SHFs and the +6 dB/octave low-pass spatial aliasing filters (SAF), which results in a global undersampling equalization filter (SHF+SAF). Thus, we exclusively consider the SHFs and SHF+SAFs in the remainder.

1.3.4 Tapering

A method denoted as Spherical Harmonics Tapering to suppress the side lobes induced by order truncation was presented in [12]. Truncating the series of SH coefficients at a given order corresponds to applying a rectangular window over the order n, which results in considerable side lobes. The authors discussed different window functions and proposed a cosine-shaped fade-out towards higher orders as the most effective one. As any order-truncated signal, the resulting binaural signals need to be equalized by the previously discussed SHFs, whereby the Tapering requires slightly modified cut-off frequencies below f_A. In the remainder, we will solely discuss the combination of Tapering and SHF and denote this Tapering+SHF.

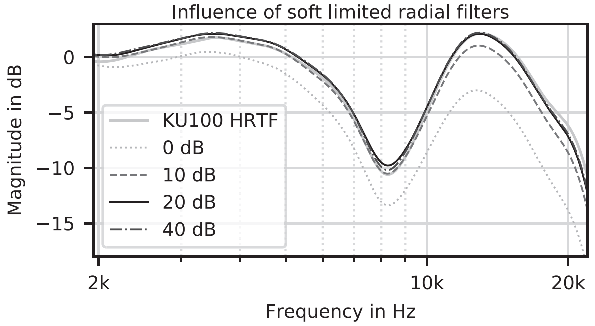


Fig. 5. Left ear magnitude responses of the frontal KU100 HRTF and ARIR binaural renderings up to 35th-order with different radial filter soft limits. The ARIR renderings are based on a simulated broadband plane wave impinging on an SMA with a 2,702-point Lebedev grid from ($\phi = 0°$, $\theta = 90°$). Our experiment employed the 0-dB limit.

2 EMPLOYED DATA

Many investigations are based on array room impulse responses (ARIRs) [19–21, 15, 4] as these allow for more flexibility regarding the design of the microphone array as well as more controlled conditions. Real-time implementations of the binaural rendering pipeline were presented e.g. in [1, 22, 23]. A noteworthy difference between ARIR-based rendering and the rendering of streamed (live-captured or recorded) signals is the fact that the signals from ARIR-based rendering are free from additive noise from the microphones and other stages in the signal chain, which can be strongly amplified by the radial filters d_n in (4). We employ a soft-limiting approach [3, pp. 90–118] that restricts the radial filter magnitudes to 0 dB. This was also done in the experiments in [3, 4] and may be considered to be on the conservative side so that the signal-to-noise ratio in the binaural signals is high even in the case that microphone self-noise and the like are apparent [24]. Fig. 5 illustrates the influence of the soft limiting. It shows the left ear BRTFs resulting from a broadband plane wave impact from ($\phi = 0°$, $\theta = 90°$) on a simulated 2,702-node Lebedev SMA. The BRTFs were calculated up to 35th-order using the different radial filter gain limits of 0, 10, 20, and 40 dB.

We used the `sound_field_analysis-py` Python toolbox [25] and the impulse response data set from [26] to prepare the stimuli. `sound_field_analysis-py` computes the radial filters $d_n(\omega)$ via sampling of the complex analytic frequency-domain representations resulting in impulse responses of length 2048 without time aliasing.

The impulse response data set contains BRIRs measured with a Neumann KU100 dummy head and ARIRs captured on various Lebedev grids under identical conditions. The ARIR measurements were performed with the VariSphear device [27], which is a fully automated robotic measurement system that sequentially captures directional impulse responses on a spherical grid for emulating a sphere microphone array. To obtain impulse responses of a rigid sphere array, an Earthworks M30 microphone was flush-mounted into a wooden spherical scattering body (see [26, Fig. 12]). All measurements were performed in four different rooms at the WDR broadcast studios in Cologne, Germany. We employed the data sets of the rooms Small Broadcast Studio (SBS) and Large Broadcast Studio (LBS) with approximate reverberation times of 1 s and 1.8 s, respectively.

Binaural rendering of the ARIRs was performed according to Eq. (4) for a pure horizontal grid of head orientations with 1° resolution using the Neumann KU100 HRIR set, which were available on a 2,702-sampling-point Lebedev grid [28]. We denote these data "ARIR renderings" in the remainder. Likewise, the BRIRs of the same dummy head were available for the same head orientations so that a direct comparison of both auralizations was possible.

All mitigation algorithms were implemented with `sound_field_analysis-py`. Solely the MagLS HRIRs were preprocessed with MATLAB code provided by the authors of [14]. Every ARIR parameter (room, order, and sampling grid) set was processed with each of the algorithms MagLS, Tapering+SHF, SHF, and SHF+SAF. An untreated (Raw) ARIR rendering was also produced.

Previous experiments showed that SH representations of an order of less than 8 exhibit audible undersampling artifacts, i.e., a clear perceptual difference to the reference dummy head data was apparent [4]. As the present work investigates undersampled sound fields, we chose to focus on SH orders below 8 for the instrumental and perceptual evaluations as we cannot expect a considerable effect of the mitigation approaches for orders higher than that.

3 INSTRUMENTAL EVALUATION

In this section, the mitigation algorithms are evaluated and compared with a focus on their influence on the time-frequency spectrum. Fig. 6 depicts the logarithmic differences of left-ear BRTFs measured with a dummy head to BRTFs based on ARIR renderings of room SBS using the anechoic HRTF of that same dummy head. The left-hand plots are based on 50 sampling point grids rendered with order 3, the right-hand plots are based on 86 sampling point grids rendered with order 7. The vertical dashed lines indicate the spatial aliasing frequency f_A. The horizontal lines indicate the head orientation for which the rendered sound source is located contralateral to the depicted ear.

It can be seen that significant differences between dummy head BRTF and ARIR signals arise above f_A and especially for the contralateral direction for the Raw rendering, which does not employ any mitigation method.

The BEMA processed ARIR renderings exhibit considerably larger deviations. Even the authors of BEMA reported that the method introduces audible artifacts when applied to nonanechoic sound fields. As shown in [13], BEMA only works well for a single plane wave impact, whereas a low number of three phase-shifted plane waves impinging from different directions already leads to considerable comb filtering artifacts. Also, the averaging of the SH coefficients in the source band leads to a low-pass effect on the binaural signals.

Comparing SHF+SAF and SHF to the Raw condition shows that both equalizations reduce the spectral differences significantly. SHF+SAF exhibits slightly lower deviations than SHF, whereby both approaches still exhibit considerable deviations around the contralateral direction.

The ARIR renderings with SH Tapering exhibit similar spectral differences like the equalization approaches SHF+SAF and SHF. Recall that Tapering incorporates a modified SHF filter. Interestingly, although the modified SHFs were employed for the Tapering algorithm, the spectral differences are more similar to SHF+SAF.

Similarly to SHF, SHF+SAF, and Tapering+SHF, the MagLS processed ARIR renderings show significantly lower spectral differences than the Raw rendering. In the case of a 3rd-order rendering, MagLS clearly yields the result closest to the reference BRIR. For the more sophisticated SMA ($N = 7$), on the other hand, MagLS does not outperform the other approaches.

In summary, the instrumental evaluation shows that SHF, SHF+SAF, Tapering+SHF, and MagLS all reduce deviations of the time-frequency spectrum to a similar extent, whereas BEMA increases them. All methods cause deviations particularly for sources that are contralateral.

4 PERCEPTUAL EVALUATION

We conducted a listening experiment in order to examine to what extent the above introduced mitigation approaches provide perceptual improvements for the binaural rendering of undersampled SMA data. The subjects' task was to compare head-tracked auralizations of SMA data that were preprocessed with one of the mitigation methods to head-tracked auralizations of corresponding dummy head measurements.

4.1 Methods

4.1.1 Stimuli

The stimuli were generated for the SH orders 3, 5, and 7 as described in Sec. 2, for a pure horizontal grid with 1° resolution allowing for direct comparison of dummy head and ARIR auralizations. Informal pilot tests revealed that there are rather small audible differences of the mitigation methods for acoustically dry environments, we chose to use the data of the rooms SBS and LBS with exhibt reverberation times of 1 s or more (cf. Sec. 2). We used the ARIRs measured on the 50-sampling-point Lebedev grid for the ARIR renderings of SH order 3 and 5 and the 86-sampling-point Lebedev grid for order 7.

Previous studies showed a significant dependency of the perceived difference on the position of the auralized sound source [4, 15]. We therefore generated all ARIR renderings for two nominal head orientations: such that the virtual sound source appeared straight ahead ($\phi = 0°,\ \theta = 90°$), as well as such that it appeared lateral ($\phi = 90°,\ \theta = 90°$). Anechoic drum recordings were used as the test signal in particular because drums have a wide spectrum and strong transients, which makes them a critical test signal. Previous studies showed that certain aspects are only revealed with critical signals [3, 4]. To support transparency, static

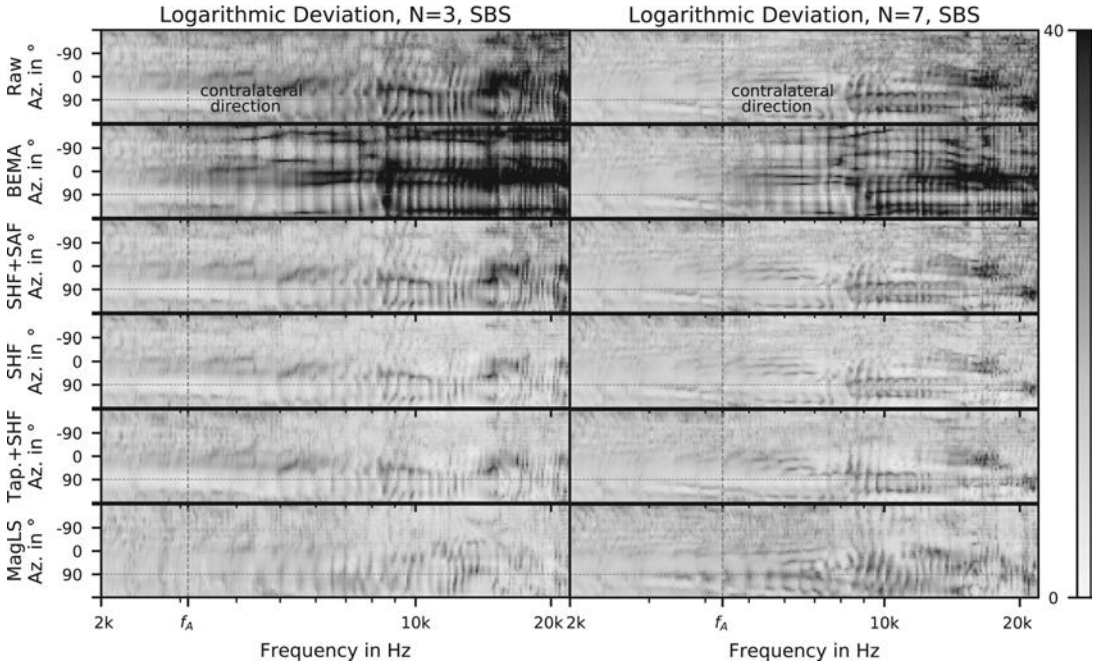


Fig. 6. Logarithmic deviations for the left ear of SBS ARIR renderings from the corresponding dummy head BRTFs with respect to azimuth angle of head orientation (vertical axes) and frequency (horizontal axes). The ARIR renderings were processed with each of the discussed algorithms. The shade of grey encodes the magnitude of the deviation ranging from 0–40 dB.

stimuli for both tested sound source positions are publicly available.[1]

4.1.2 Setup

The experiment was conducted in a quiet acoustically damped audio laboratory at Chalmers University of Technology. The SoundScape Renderer (SSR) [29, 30] in binaural room synthesis (BRS) mode was used for dynamic auralization. It convolves arbitrary input test signals with a pair of BRIRs corresponding to the instantaneous head orientation of the listener, which was tracked along the azimuth with a Polhemus Patriot tracker. A change of head orientation as well as switching between stimuli results in a cross-fade with cosine ramps over the course of one processing block. All stimuli were time aligned so that no artifacts occurred during the fade.

The binaural renderings were presented to the participants using AKG K702 headphones with a Lake People G109 headphone amplifier at a playback level of about 66 dBA. The output signals of the SSR were routed to an Antelope Audio Orion 32 DA converter at 48 kHz sampling frequency and a buffer length of 512 samples. Equalization according to [26] was applied to compensate for the headphone transfer function. All involved software components were running on the same iMac Pro 1.1 computer.

4.1.3 Paradigm

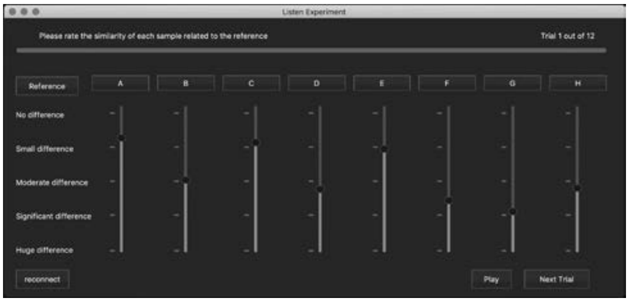
We used a test design based on the Multiple Stimulus with Hidden Reference and Anchor (MUSHRA) method as proposed by the International Telecommunication Union (ITU) [31]. The subjects' task was to rate the overall perceived difference between the ARIRs renderings and the dummy head reference. We used a non-head-tracked diotic 0° dummy head reference BRIR of the room under test as anchor, which was low-pass filtered with a cutoff at 3 kHz.

Table 1. The stimuli employed in the listening experiment. All algorithms were presented in each of the trials. Each such set was rendered for 3 SH orders, 2 source positions, and 2 rooms. This results in 12 trials with 8 stimuli each.

Algorithm	SH order (grid)	Position	Room
BEMA MagLS SHF SHF + SAF Tapering + SHF Raw Reference Anchor	3 (50) 5 (50) 7 (86)	$\phi = 0°, \theta = 90°$ $\phi = 90°, \theta = 90°$	LBS SBS

Each trial required 8 ratings to be performed by the subject (BEMA, MagLS, SHF, Tapering, SHF+SAF, Raw, hidden reference, anchor) against the dummy head reference. The experiment consisted of 12 trials: 3 SH orders $(3, 5, 7)$ $\times$ 2 nominal source positions $(0°,\ 90°)$ $\times$ 2 rooms (LBS, SBS), as summarized in Tab. 1. The subjects were provided with a graphical user interface (GUI) with continuous sliders named as “No difference” (100), “Small difference” (75), “Moderate difference” (50), “Significant difference” (25), and “Huge difference” (0) as depicted in Fig. 7.

[1] https://doi.org/10.5281/zenodo.3759343



Listen Experiment

Please rate the similarity of each sample related to the reference

Trial 1 out of 12

Fig. 7. PyQt GUI used in the listening experiment.

4.1.4 Procedure

Twenty participants, 4 of them female, between the ages of 22 and 50 took part in the experiment. Most of them were M.Sc. students or staff at the Division of Applied Acoustics of Chalmers University of Technology. Sixteen participants reported that they had previously participated in a listening experiment. The subjects were sitting in front of a computer screen with a keyboard and a mouse. It was possible to listen to each stimulus as often and long as desired. The participants were allowed and strongly encouraged to move their heads during the presentation of the stimuli. At the beginning of each experiment, the subjects had to rate four training stimuli that covered a representative set of perceptual differences of the presented stimuli in the subsequent test. These training stimuli consisted of a BEMA and MagLS rendering of SBS data at 3rd order for the lateral sound source position, as well as the corresponding anchor and reference. The experiment took on average about 40 minutes per participant.

4.2 Results

All anchor and reference ratings were post-screened before applying statistical analysis according to the recommendation of the ITU [31]. All anchor ratings were below 30 and most reference ratings above 80. Only two reference ratings (50, 49) and two anchor ratings (40, 38) were conspicuous, which constitutes a low portion of in total 96 ratings per participant. We performed statistical analyses including and excluding the respective subjects' data, which led to identical results. We report only the results over the complete data set here.

Our subjects compared the different algorithms for a given combination of SH order, room, and source position in each trial. We therefore highlight that a direct comparison of the ratings for different orders, rooms, and source positions has to be performed with reservation. The anchors and references were conceptually the same across all trials and the stimulus and condition order were randomized per participant. A certain amount of consistency in the subjects'

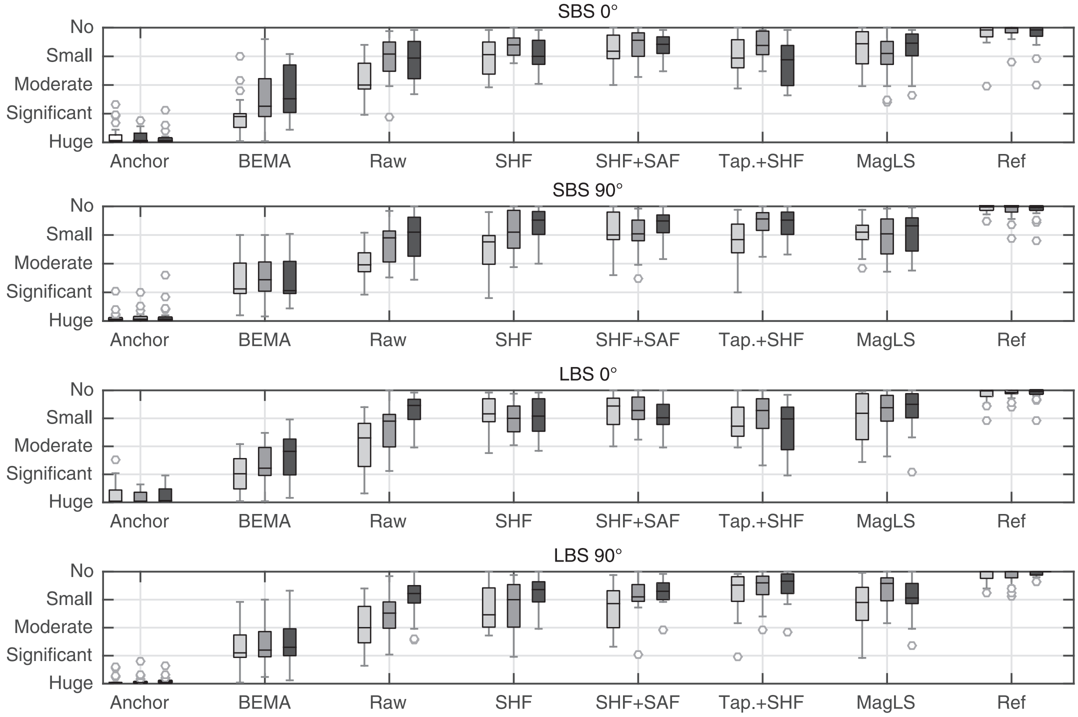


Fig. 8. Boxplots illustrating the ratings of the perceptual difference between the stimulus and the dummy head reference for each room and virtual source position separately. Each figure depicts the boxplots for each algorithm at the SH order 3 (light grey), 5 (dark grey), and 7 (black), respectively. Each box indicates the 25th and 75th percentiles, the median value (black line), the outliers (grey circles), and the minimum/maximum ratings not identified as outliers (black whiskers).

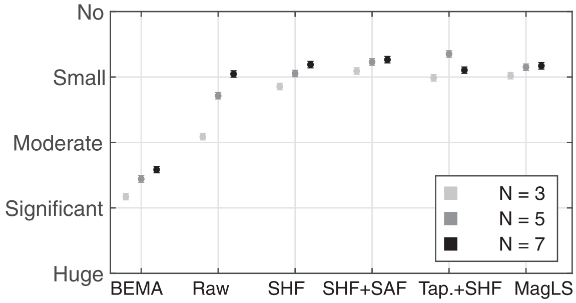


Fig. 9. Mean difference ratings pooled over source position and room with respect to algorithm (abscissa) and SH order (colors). The 95% within-subject confidence intervals were calculated according to [33, 34].

responses may therefore be assumed. In the following, we present a statistical analysis that includes the comparison between orders and positions as it is commonly performed with MUSHRA data. This facilitates discussing the results in relation to the literature as we will do in Sec. 4.3.

An overview of the results is presented as boxplots in Fig. 8, illustrating the ratings for the rooms SBS and LBS and source positions at 0° and 90° separately. The boxplots confirm that subjects rated the hidden anchor and the reference consistently. Furthermore, perceptual differences between Raw and dummy head renderings tended to become smaller with increasing SH order. All algorithms with the exception of BEMA led to a smaller perceptual difference to the reference than the Raw renderings.

For statistical analysis of the results, repeated measures ANOVAs were performed. A Lilliefors test for normality was applied to test the requirements for the ANOVA. It failed to reject the null hypothesis in 14 of 72 conditions at a significance level of $p = 0.05$. However, parametric tests such as the ANOVA are generally robust to violations of normality assumption [32]. For further analysis, Greenhouse–Geisser-corrected p values are considered, with the associated ϵ-values for correction of the degrees of freedom of the F-distribution being reported.

A four-way repeated measures ANOVA with the within-subject factors algorithm (BEMA, MagLS, Tapering+SHF, SHF, SHF+SAF, and Raw), order $(3, 5, 7)$, room (SBS, LBS), and nominal source position (0°, 90°) was performed. In addition, a number of nested repeated measures ANOVAs were performed. For each of the six algorithms, one three-way ANOVA with the factors order $(3, 5, 7)$, room (SBS, LBS), and source position (0°, 90°), as well as a four-way ANOVA with the subset of MagLS, Tapering+SHF, SHF, and SHF+SAF for the factor algorithm and the factors order $(3, 5, 7)$, room (SBS, LBS), and source position (0°, 90°) were applied. The results of the ANOVA incorporating all algorithms are presented in Tab. 2.

The results of the experiment are depicted in aggregate form in Fig. 9. The mean values with respect to algorithm and SH order are depicted separately. Each value was calculated by averaging the ratings of all participants, source positions and rooms. Furthermore, 95% within-subject confidence intervals as proposed by [33, 34], based on the main effect of the algorithm, are shown. The plots confirm the observations taken from the boxplots and additionally show that the ratings do not scale linearly with the rendering order. It is noteworthy that on average, the Tapering renderings were rated with a larger perceptual difference when rendered at SH order 7 than with order 5.

Overall, the ratings of the algorithms SHF, SHF+SAF, Tapering+SHF, and MagLS are located in a similar range. We therefore preliminary conclude that all algorithms achieve a similar magnitude of improvement compared to Raw renderings.

The following analysis refers to main effects and first order interactions only. It was found for the four-way ANOVA involving all algorithms that the algorithm and order main effects as well as the first order interaction effects algorithm×order, algorithm×position, and order×position were significant. These effects will be examined successively in the following paragraphs.

The main effects of the algorithm ($F(5, 95) = 194.9$, $p < .001$, $\eta_p^2 = .911$, $\epsilon = .684$) as well as of the order ($F(2, 38) = 40.75$, $p < .001$, $\eta_p^2 = .682$, $\epsilon = .765$) support the trends identified in the boxplots in Fig. 8 and the mean plots in Fig. 9. All algorithms significantly affect the perceived similarity and for all algorithms other than Tapering+SHF, higher SH orders yield more perceived similarity. Furthermore, the interaction effect algorithm×order ($F(10, 190) = 8.06$, $p < .001$, $\eta_p^2 = .298$, $\epsilon = .612$) suggests that both factors do not just exclusively influence the perceived differences, but the algorithms may lead to different levels of improvements with respect to the SH order.

To validate the observation that the algorithms SHF, SHF+SAF, Tapering+SHF, and MagLS achieved similar improvements, a four-way repeated measures ANOVA was performed taking into account only the results for these algorithms. The factor algorithm was not significant ($p = .107$), showing that all algorithms except BEMA achieved similar perceptual improvements. The ANOVAs conducted for each algorithm separately suggested no significant effect for the SH order for the algorithm MagLS only ($p = .202$). This indicates that MagLS performs comparably similar at all orders.

We found no main effect of the factor source position ($p = .49$), but an interaction effect of algorithm×position ($F(5, 95) = 5.563$, $p < .001$, $\eta_p^2 = .227$, $\epsilon = .001$). This suggests that the algorithms perform differently dependent on the presented source position. Moreover, the ANOVA revealed a significant interaction of order×position ($F(2, 38) = 194.9$, $p < .026$, $\eta_p^2 = .187$, $\epsilon = .858$). Thus, the position dependency varies with respect to the order. The results of the ANOVA can be seen in Fig. 10 (left, right), presenting the mean values calculated similarly to Fig. 9 but separated into frontal and lateral nominal source position. The plots indicate that the 7th-order renderings processed with the SHF, SHF+SAF, and Tapering+SHF algorithms were rated with larger perceptual difference for frontal than for lateral sound source positions.

Interestingly, an ANOVA over exclusively the data of any one given algorithm suggests that Tapering+SHF is the only single algorithm for which a significant main effect of the source position ($F(1, 19) = 15.61$, $p < .001$, $\eta_p^2 = .451$,

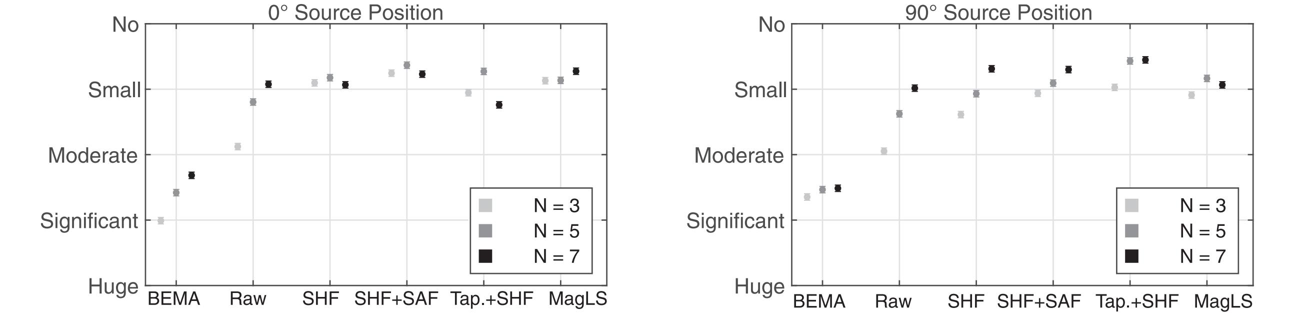


Fig. 10. Mean difference ratings for 0° (left) and 90° (right) sound source position pooled over both rooms with respect to algorithm (abscissa) and SH order (colors). The 95% within-subject confidence intervals were calculated according to [33, 34].

$\epsilon = 1$) may be apparent. To further dissect this observation, we performed multiple *t* tests (with Hochberg correction to correct for multiple hypothesis testing), comparing 5th and 7th-order renderings processed with SHF, SHF+SAF, and Tapering+SHF for frontal and lateral source positions. The tests suggest a significant difference between the ratings for frontal and lateral source position only for 7th-order renderings with Tapering+SHF ($t(39) = 4.879$, $p < .001$, $d_z = .772$). This indicates a rather weak influence of source position and order in the present data set. Concerning the influence of the room, we found neither a main effect nor any interaction effect.

4.3 Discussion

The results of the perceptual evaluation show that all presented algorithms other than the BEMA approach yield perceivable improvements of binaural array renderings. No algorithm was rated significantly better than the others. All analysis of the dependency of the ratings on the rendering order, room, and source position has to be performed with reservation as this requires comparing ratings across different trials. As we argued in Sec. 4.2, a considerable amount of consistency of the ratings may be assumed across trials. We discuss our data in relation to findings from similar studies and analyses in the literature in the following.

Our listening experiment confirms that higher-order renderings were mostly rated closer to the dummy head than lower order renderings. However, all orders we tested were rated significantly different compared to the reference. This matches the findings from [3, 4] where it was found that renderings of an order below 8 exhibit audible differences to dummy head auralizations and that these differences are induced by spatial undersampling. The soft limiting that we applied to the radial filters may have led to audible differences of ARIR and dummy head auralization independent of undersampling. This may have caused a saturation of the perceptual improvement towards higher orders. Similar results were obtained in [3, 4] where similar soft limiting was applied. We assume that less-conservative radial filter limits lead to more similarity to the dummy head in particular at higher frequencies, as indicated by Fig. 5. The cost is a lower signal-to-noise ratio in the binaural signals if additive sensor noise is apparent [24]. Similarly to [3, 4], we observed no room dependency in the ratings.

Table 2. Results of the four-way repeated measures ANOVA with the within-subject factors algorithm (BEMA, MagLS, Tapering+SHF, SHF, SHF+SAF, Raw), order (3, 5, 7), source position (0°, 90°), and room (SBS, LBS).

Effect	*df*	*F*	ϵ_{GG}	η_p^2	p_{GG}
Algorithm	5	194.898	.684	.911	<.001*
Order	2	40.750	.765	.682	<.001*
Position	1	.495	1.000	.025	.490
Room	1	1.617	1.000	.078	.219
Algorithm × Order	10	8.055	.612	.298	<.001*
Algorithm × Position	5	5.565	.653	.227	.001*
Order × Position	2	4.372	.858	.187	.026*
Algorithm × Room	5	1.001	.742	.050	.409
Order × Room	2	.731	.709	.037	.446
Position × Room	1	.181	1.000	.009	.676
Algorithm × Order × Position	10	4.479	.644	.191	<.001*
Algorithm × Order × Room	10	3.218	.549	.145	.008*
Algorithm × Position × Room	5	1.445	.713	.071	.233
Order × Position × Room	2	.478	.913	.025	.607
Algorithm × Order × Position × Room	10	.909	.622	.046	.494

ϵ_{GG}: Greenhouse–Geisser epsilons
η_p^2: Partial eta squared
p_{GG}: Greenhouse–Geisser corrected *p* values
Statistical significance at 5% level are indicated by asterisks

Previous studies observed a dependency of the ratings on the sound source position. In [4], the participants compared dummy head auralizations and raw ARIR renderings in terms of spaciousness and timbre separately. The timbre of ARIR renderings of SH orders of 8 and higher was perceived noticeably closer to the dummy head auralization for lateral sound sources than for frontal sources, which at first glance seems surprising considering the deviations of truncated SH representations at the contralateral ear for lateral source positions (cf. Fig. 6). The authors concluded that spectral differences of frontal sound sources can be perceived more reliably than the spectral differences of lateral sources and attributed it to the higher spatial resolution of the human auditory system in the front [35].

In contrast, another study presented in [15] showed that ARIR renderings treated with the time alignment approach (a predecessor of MagLS) were rated lower for lateral than for frontal sources. However, even though the plots of our results in Fig. 10 indicate some amount of source position dependency, the statistical analysis revealed a significant effect of source position only for the algorithm Tapering+SHF. There is therefore no statistical evidence in our data that supports the observations of a general influence of source position on difference ratings that was made in [4, 15].

The statistically significant effect of order and source position for the algorithm Tapering+SHF, e.g., that 5th-order renderings were sometimes rated higher than 7th-order renderings, might have been caused by unfavorable effects of the tapering with higher-order data. The higher the rendering order, the more SH modes are attenuated by the window, as the tapering starts always at $n = 1$ independent of the maximum order. Tapering in its present form might therefore be most beneficial for rendering orders below 7. A modified approach that tapers only the last few orders below the maximum order is conceivable.

As discussed above, these observation can ultimately only be proven based on data from a direct comparison of stimuli of different orders.

5 CONCLUSIONS

A listening experiment comparing different algorithms to mitigate the perceptual impairment of binaural rendering of SMA data due to spatial undersampling was presented. The subjects' task was to compare array renderings enhanced with state-of-the-art algorithms to corresponding auralizations of dummy head impulse response data in terms of overall perceived difference.

We found that the Magnitude Least-Squares HRIR preprocessing approach, the Spherical Head Filters, and the Spherical Harmonics Tapering (including the SHF), as well as a global undersampling equalization filter, all yield a significant improvement of the SMA renderings. We only evaluated the overall perceived difference to the dummy head auralization and can therefore not break down the differences into individual attributes. A follow-up study, evaluating these attributes such as spaciousness or timbre separately may expose individual advantages and disadvantages of the investigated approaches in more detail. It may be assumed that appropriate equalization of the spectrum yields improvements in particular for the timbre, whereas MagLS and Tapering may be more beneficial for improving the localization and thus, the spaciousness of the binaural synthesis.

The Bandwidth Extension Algorithm for Microphone Arrays is the only algorithm aiming at recovering the loss of spatial information due to spatial aliasing that seemed to produce more harm than benefit. This is mainly caused by the low-pass effect of the involved SH coefficient averaging. Magnitude Least-Squares and Tapering have shown to be appropriate algorithms to mitigate the truncation artifacts, but also, a simple equalization of the binaural time-frequency response, i.e., the Spherical Head Filters and the global undersampling equalization filter, yielded perceptually equivalent results. Simple (global) equalization has the disadvantage of shifting coloration impairments, by design, mostly to the contralateral side.

Although most tested algorithms are successful in improving the array auralizations, there are still audible differences to the corresponding dummy head reference. These differences may be related to spatial ambiguities of spatial aliasing. Instrumental analysis as well as informal listening revealed that the modification of the time-frequency response is more affected by SH order truncation than by spatial aliasing. It remains to be clarified whether the overall perceptual influence of truncation is more significant, and whether spatial aliasing artifacts can even be neglected for sufficient auralizations.

Some amount of the saturation of the observed perceived differences may also be attributed to the choice of radial filter limit, which caused a slight attenuation of the signal at higher frequencies.

6 ACKNOWLEDGMENT

We thank all participants of the listening experiment for their voluntary support and Christian Schörkhuber, Markus Zaunschirm, and Franz Zotter of Institute of Electronic Music and Acoustics at the University of Music and Performing Arts in Graz for providing us with their code of MagLS. We also thank all anonymous reviewers for very valuable and stimulating feedback.

7 REFERENCES

[1] F. Zotter and M. Frank, *Ambisonics A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality* (Springer-Verlag, Berlin, Germany, 2019), doi:10.1007/978-3-030-17207-7, https://plugins.iem.at/.

[2] J. Ahrens, *Analytic Methods of Sound Field Synthesis* (Springer, Berlin, Germany, 2012), doi:10.1007/978-3-642-25743-8.

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

[4] J. Ahrens and C. Andersson, "Perceptual Evaluation of Headphone Auralization of Rooms Captured With Spherical Microphone Arrays With Respect to Spaciousness and Timbre," *Journal of the Acoustical Society of America*, vol. 145, no. 4, pp. 2783–2794 (2019 Apr.), doi:10.1121/1.5096164.

[5] B. Rafaely, *Springer Topics in Signal Processing Springer Topics in Signal Processing* (Springer, Berlin, Germany, 2015), doi:10.1007/978-3-642-11130-3.

[6] E. G. Williams, *Fourier Acoustics* (Academic Press, London, United Kingdom, 1999), doi:10.1016/B978-0-12-753960-7.X5000-1.

[7] C. Andersson, *Headphone Auralization of Acoustic Spaces Recorded with Spherical Microphone Arrays*, Master's thesis, Chalmers University of Technology (2017).

[8] S. Lösler and F. Zotter, "Comprehensive Radial Filter Design for Practical Higher-Order Ambisonic Recording," presented at the *41st DAGA*, 1, pp. 452–455 (2015), doi:10.3758/BF03210951.

[9] C. E. Shannon, "Communication in the Presence of Noise," *Proceedings of the IRE*, vol. 37, no. 1, pp. 10–21 (1949), doi:10.1109/JRPROC.1949.232969.

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

[11] Z. Ben-Hur, D. L. Alon, B. Rafaely, and R. Mehra, "Loudness Stability of Binaural Sound With Spherical Harmonic Representation of Sparse Head-Related Transfer Functions," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2019, no. 1 (2019 Dec.), doi:10.1186/s13636-019-0148-x.

[12] C. Hold, H. Gamper, V. Pulkki, N. Raghuvanshi, and I. J. Tashev, "Improving Binaural Ambisonics Decoding by Spherical Harmonics Domain Tapering and Coloration Compensation," presented at the *International Conference on Acoustics, Speech and Signal Processing*, vol. 2, pp. 261–265 (2019), doi:10.1109/ICASSP.2014.6854442.

[13] B. Bernschütz, "Bandwidth Extension for Microphone Arrays," presented at the *133rd Convention of the Audio Engineering Society* (2012 Oct.), convention paper 8751.

[14] C. Schörkhuber, M. Zaunschirm, and R. Holdrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," presented at the *44th DAGA*, 4, pp. 339–342 (2018).

[15] 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 Jun.), doi:10.1121/1.5040489.

[16] L. Rayleigh, "XII. On Our Perception of Sound Direction," *Philosophical Magazine Series 6*, vol. 13, no. 74, pp. 214–232 (1907), doi:10.1080/14786440709463595.

[17] Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, and B. Rafaely, "Spectral Equalization in Binaural Signals Represented by Order-Truncated Spherical Harmonics," *The Journal of the Acoustical Society of America*, vol. 141, no. 6, pp. 4087–4096 (2017 Jun.), doi:10.1121/1.4983652.

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

[19] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "SOFiA Sound Field Analysis Toolbox," presented at the *International Conference on Spatial Audio (ICSA)*, pp. 8–16 (2011 Jan.).

[20] J. Sheaffer, S. Villeval, and B. Rafaely, "Rendering Binaural Room Impulse Responses From Spherical Microphone Array Recordings Using Timbre Correction," presented at the *Joint Symposium on Auralization and Ambisonics (EAA)*, pp. 3–5 (2014 Apr.).

[21] A. Neidhardt, *Untersuchungen zur räumlichen Genauigkeit bei der binauralen Auralisation von Kugelarraydaten*, Master's thesis, Technische Universität Graz (2015).

[22] H. Helmholz, C. Andersson, and J. Ahrens, "Real-Time Implementation of Binaural Rendering of High-Order Spherical Microphone Array Signals," presented at the *45th DAGA*, pp. 2–5 (2019), https://github.com/AppliedAcousticsChalmers/ReTiSAR.

[23] L. Mccormack and A. Politis, "SPARTA & COMPASS: Real-Time Implementations of Linear and Parametric Spatial Audio Reproduction and Processing Methods," presented at the *AES Conference on Immersive and Interactive Audio* (2019 Mar.), conference paper 111, doi:10.1121/1.3278605, https://github.com/leomccormack/SPARTA.

[24] H. Helmholz, D. L. Alon, S. V. A. Garí, and J. Ahrens, "Instrumental Evaluation of Sensor Self-Noise in Binaural Rendering of Spherical Microphone Array Signals," presented at the *Forum Acusticum*, pp. 1–8 (2020).

[25] C. Hohnerlein and J. Ahrens, "Spherical Microphone Array Processing in Python With the Sound Field Analysis-Py Toolbox," *Fortschritte der Akustik – DAGA 2017*, pp. 1033–1036 (2017).

[26] P. Stade, B. Bernschütz, and M. Rühl, "A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios," presented at the *27th Tonmeistertagung - VDT International Convention*, pp. 551–567 (2012 Nov.).

[27] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, "Entwurf und Aufbau eines variablen sphärischen Mikrofonarrays für Forschungsanwendungen in Raumakustik und Virtual Audio," presented at the *36th DAGA*, pp. 717–718 (2010).

[28] B. Bernschütz, "A Spherical Far Field HRIR/HRTF Compilation of the Neumann KU 100," presented at the *39th DAGA*, pp. 592–595 (2013).

[29] M. Geier, J. Ahrens, and S. Spors, "The Soundscape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods," presented at the

124th Convention of the Audio Engineering Society (2008 May), convention paper 7330.

[30] M. Geier, J. Ahrens, and S. Spors, "The SoundScape Renderer," http://spatialaudio.net/ssr/ (2019), version 0.4.2, retrieved 2019-07-01.

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

[32] J. Bortz and C. Schuster, *Statistik für Human- und Sozialwissenschaftler*, 7th ed. (Springer-Verlag, Gießen, Germany, 2010), doi:10.1121/1.3278605.

[33] G. R. Loftus, "Using Confidence Intervals in Within-Subject Designs," *Psychonomic Bulletin & Review*, vol. 1, no. 4, pp. 1–15 (1994), doi:10.3758/BF03210951.

[34] J. Jarmasz and J. G. Hollands, "Confidence Intervals in Repeated-Measures Designs: The Number of Observations Principle," *Canadian Journal of Experimental Psychology*, vol. 63, no. 2, pp. 124–138 (2009 Jun.), doi:10.1037/a0014164.

[35] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, Massachusetts, 1997).

THE AUTHORS

Tim Lübeck

Hannes Helmholz

Johannes M. Arend

Christoph Pörschmann

Jens Ahrens

Tim Lübeck received his B.Sc. degree in Electrical Engineering in 2017 and his M.Sc. degree in Communication Engineering in 2019 from TH Köln, Cologne, Germany. He completed his master's thesis in cooperation with the Division of Applied Acoustics at Chalmers University. Since 2019, he has been a Research Fellow and working toward a Ph.D. at TH Köln and TU Berlin in the field of virtual acoustics, binaural technology, auditory perception, and audio signal processing.

Hannes Helmholz received his B.Sc. degree in Applied Computer Science from the University of Applied Sciences, Berlin, Germany, in 2013. and his M.Sc. degree in Audio Communication and Technology from Technische Universität Berlin, Berlin, Germany, in 2018. Since 2018, he has been working toward a Ph.D. in Applied Acoustics at Chalmers University of Technology, Gothenburg, Sweden, in the field of spatial audio, binaural technology, auditory perception, and audio signal processing.

Johannes M. Arend received his B.Eng. degree in media technology from HS Düsseldorf, Düsseldorf, Germany, in 2011 and his M.Sc. degree in media technology from TH Köln, Cologne, Germany, in 2014. Since 2015, he has been a Research Fellow and working toward a Ph.D. at TH Köln and TU Berlin in the field of spatial audio with a focus on binaural technology, auditory perception, and audio signal processing.

Christoph Pörschmann studied Electrical Engineering at the Ruhr-Universität Bochum (Germany) and at Uppsala Universitet (Sweden). In 2001, he obtained his Doctoral Degree (Dr.-Ing.) from the Electrical Engineering and Information Technology Faculty of the Ruhr-Universität Bochum as a result of his research at the Institute of Communication Acoustics. Since 2004, he has been Professor of Acoustics at TH Köln – University of Applied Sciences. His research interests are in the fields of virtual acoustics, spatial hearing, and the related perceptual processes.

Jens Ahrens has been an Associate Professor within the Division of Applied Acoustics at Chalmers University since 2016. He has also been a Visiting Professor at the Applied Psychoacoustics Lab at the University of Huddersfield, Huddersfield, UK, since 2018. Jens received his Diploma (equivalent to a M.Sc.) in Electrical Engineering/Sound Engineering jointly from Graz University of Technology and the University of Music and Dramatic Arts, Graz, Austria, in 2005. He completed his Doctoral Degree (Dr.-Ing.) at the Technische Universität Berlin, Berlin, Germany, in 2010. From 2011 to 2013, Jens was a Postdoctoral Researcher at Microsoft Research in Redmond, Washington, USA, and in the fall and winter terms of 2015/2016, he was a Visiting Scholar at the Center for Computer Research in Music and Acoustics (CCRMA) at Stanford University, Stanford, California, USA. He is an Associate Editor of the IEEE Signal Processing Letters and a Guest Editor of the EURASIP Journal on Audio, Speech, and Music Processing.