Zamir Ben-Hur, Fabian Brinkmann, Jonathan Sheaffer, Stefan Weinzierl, Boaz Rafaely

Spectral equalization in binaural signals represented by order-truncated spherical harmonics

Spectral equalization in binaural signals represented by order-truncated spherical harmonics

Zamir Ben-Hur,[1,a)] Fabian Brinkmann,[2] Jonathan Sheaffer,[1] Stefan Weinzierl,[2] and Boaz Rafaely[1]

[1] Department of Electrical and Computer Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

[2] Audio Communication Group, Technical University of Berlin, Einsteinufer 17c, D-10587 Berlin, Germany

The synthesis of binaural signals from spherical microphone array recordings has been recently pro-posed. The limited spatial resolution of the reproduced signal due to order-limited reproduction has been previously investigated perceptually, showing spatial perception ramifications, such as poor source localization and limited externalization. Furthermore, this spatial order limitation also has a detrimental effect on the frequency content of the signal and its perceived timbre, due to the rapid roll-off at high frequencies. In this paper, the underlying causes of this spectral roll-off are described mathematically and investigated numerically. A digital filter that equalizes the frequency spectrum of a low spatial order signal is introduced and evaluated. A comprehensive listening test was conducted to study the influence of the filter on the perception of the reproduced sound. Results indicate that the suggested filter is beneficial for restoring the timbral composition of order-truncated binaural signals, while conserving, and even improving, some spatial properties of the signal.

I. INTRODUCTION

Binaural reproduction plays an important role in virtual acoustics,[1] hearing science,[2] telecommunications,[3] and entertainment technology.[4] One key element in binaural technology is binaural signals, which can be obtained from in-ear recordings or synthesized from spatial sound field measurements.[5] Recently, it has been proposed to render such binaural signals in the spherical harmonics (SH) domain[6] by adding the products of the spherical Fourier coefficients of the plane-wave (PW) density function (which encodes the directional information of the sound field) with the SH representation of the free-field head related transfer functions (HRTFs). This approach is commonly employed in algorithms operating in the SH domain, which require manipulation of the sound field or the corresponding HRTFs in post-processing.[7,8]

When reproducing a measured sound field, the PW density function can be obtained from a spherical microphone array recording.[9] In practice, such microphone arrays have a limited number of sensors, which directly affects the usable spatial bandwidth of the recovered PW decomposition.[10] This, in turn, places a constraint on the maximum SH order of the employed catalog HRTF; such catalog HRTFs are typically available at considerably higher spatial resolutions than measured sound fields. As will be further detailed, truncating the SH order of an HRTF results in a frequency roll-off at high frequencies, an artifact which has been experimentally identified by Avni et al.[11]

This roll-off manifests a relationship between the spatial and frequency bandwidths of the rendered signals. In other words, synthesizing a binaural signal with a low SH order results in a severely constrained frequency bandwidth. The effect is particularly troublesome when conducting psycho-acoustic studies of spatial perception, as it does not allow the experimenter to systematically control the studied spatial resolution as a single independent variable; for example, see Avni et al.[11] In addition, algorithms operating in the SH domain, such as generalized spherical beamformers[7,12] and aliasing cancellation techniques,[8] often require the reduction of the SH order of the processed signals, but without affecting their spectral bandwidth.

In this paper, we address the issue of producing binaural signals that are order-limited in the SH domain, while simultaneously preserving the spectral energy of these signals across a prescribed bandwidth. The underlying causes of spectral roll-off are identified and mathematically described using a spherical head model, followed by a numerical analysis of measured HRTFs (Sec. III). We then introduce a digital filter that equalizes the frequency spectrum of a low SH order binaural signal to that of a high SH order. A listening experiment, described in Sec. V, studies the perceptual performance of the proposed equalization filter, in terms of timbre, sound localization, source extension, and externalization. Results indicate that the proposed filter is beneficial for restoring the timbral composition of order-truncated binaural signals, while conserving the spatial perception of the signals. The psychoacoustics underlying these results is discussed in detail in Sec. VI. In summary, the contributions of this paper are as follows:

(1) The spectral roll-off effect, which so far has been predominantly treated experimentally, is mathematically explained and numerically analyzed.

a)Electronic mail: zami@post.bgu.ac.il

(2) An equalization filter for timbre restoration is formulated. This filter has been proposed by the authors and presented (in less detail) in recent conference publications.[13,14]

(3) A comprehensive listening experiment to study the performance of the equalization filter is conducted, providing a systematic treatment of the perceptual aspects relating to the filter design.

II. REPRESENTATION OF BINAURAL SIGNALS IN THE SH DOMAIN

The approach developed by Rafaely and Avni[6] for rendering binaural signals from the sound pressure measured by a spherical microphone array is outlined in this section. Consider a function $F(\Omega)$, where $\Omega \equiv (\theta, \phi) \in S^2$ is the spatial angle, represented by the elevation angle $\theta \in (0, \pi)$, which is measured downwards from the Cartesian z axis, and the azimuth angle $\phi \in [0, 2\pi)$, which is measured counterclockwise from the Cartesian x axis in the xy-plane. Assuming $F(\Omega)$ is square integrable over S^2, it can be represented in the SH domain by[15]

$$F(\Omega) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} F_{nm} Y_n^m(\Omega),$$ (1)

where F_{nm} is the spherical Fourier transform (SFT) of $F(\Omega)$, given by

$$F_{nm} = \int_{\Omega \in S^2} F(\Omega) \left[Y_n^m(\Omega)\right]^* d\Omega,$$ (2)

where $(\cdot)^*$ denotes the complex conjugate, $\int_{\Omega \in S^2}(\cdot)d\Omega \equiv \int_0^{2\pi} \int_0^{\pi}(\cdot) \sin\theta d\theta d\phi$, and $Y_n^m(\Omega)$ are the SH basis functions, defined by[16]

$$Y_n^m(\Omega) \equiv \sqrt{\frac{(2n+1)}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos\theta) e^{im\phi},$$ (3)

where $P_n^m(\cdot)$ is the associated Legendre function of order n and degree m.

The pressure encountered at a listener's ears can be represented using integration over the sphere as[17]

$$p^{l,r}(k) = \int_{\Omega \in S^2} a(k, \Omega) h^{l,r}(k, \Omega) d\Omega,$$ (4)

where $k = 2\pi f/c$ is the wave number, f is the frequency, and c is the speed of sound. $a(k, \Omega)$ is the PW density function, which represents a sound field composed of a continuum of PWs, and $h^{l,r}(k, \Omega)$ is the HRTF, which is an acoustic transfer function from the sound source to the listener's eardrum.[18] The superscripts l,r denote the left and right ears, respectively. Alternatively, the pressure at the ear can be represented in the SH domain by[6]

$$p^{l,r}(k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \tilde{a}_{nm}^*(k) h_{nm}^{l,r}(k),$$ (5)

where

$$\tilde{a}_{nm}(k) = (-1)^m \left[a_{n(-m)}(k)\right]^*$$ (6)

is the SFT of $a^*(k, \Omega)$. Recall that $(\cdot)^*$ denotes the complex conjugate, and $h_{nm}^{l,r}(k)$ are the SFTs of $h^{l,r}(k, \Omega)$, which can be readily computed by applying Eq. (2), the SFT, to the HRTFs, and $\tilde{a}_{nm}(k)$ can be calculated by capturing the sound field using a spherical microphone array and applying PW decomposition in the SH domain. For theoretical and experimental studies see, for example, Rafaely[9] and Park and Rafaely,[19] respectively.

In practice both $h_{nm}^{l,r}(k)$ and $\tilde{a}_{nm}(k)$ may be computed from samples of $h^{l,r}(k, \Omega)$ and $a^*(k, \Omega)$, and so the integral in Eq. (2) can be approximated by a summation. This approximation is realized in practice using the matrix representation of the discrete SFT. For further reading regarding the computation of the discrete SFT from samples the reader is referred to Rafaely.[16] Furthermore, in this work it is assumed that $h_{nm}^{l,r}(k)$ and $\tilde{a}_{nm}(k)$ are computed with negligible spatial aliasing errors.

In the practical case, $h_{nm}^{l,r}(k)$ or $\tilde{a}_{nm}(k)$ will only be available up to an order N, so the computation of the pressure at the ear can only be performed for the order limited function, and, therefore, Eq. (5) can be written with a finite summation over n,

$$p^{l,r}(k) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \tilde{a}_{nm}^*(k) h_{nm}^{l,r}(k).$$ (7)

III. SPECTRAL ROLL-OFF AND EQUALIZATION OF ORDER LIMITED SIGNALS

As was identified by Avni *et al.*,[11] the truncation of the series not only restricts the spatial resolution of the binaural signal, but also affects its frequency content. Figure 1(c) shows an example for a binaural signal composed of a single unit amplitude PW incident at $\Omega = (90°, 45°)$ and truncated at different SH orders, computed using Eq. (7) with measured HRTFs from the Cologne HRTF database for the Neumann KU100 dummy head.[20] Note the rapid roll-off at high frequencies, which is characterized by a lower cutoff frequency as the series truncation order is decreased. Figure 1(a) shows the PW density function of a unit amplitude single PW, $a(k, \Omega)$, on the horizontal plane, at $\theta = 90°$. Figure 1(b) shows the left ear HRTF, $h^l(k, \Omega)$, at the incident direction $\Omega = (90°, 45°)$. Both of the above are used to reproduce the binaural signal in Fig. 1(c). In Fig. 1(a) the effect of the truncation order can be identified as the widening of the main lobe, which could affect the perceived source width of the reproduced signal. Figure 1(b) implies that the roll-off in the reproduced signal is mainly caused by the order-truncated HRTFs.

To identify the causes for this roll-off effect, the HRTF can be approximated by modeling the head as a rigid sphere.[21] Consider a single PW arriving from an arbitrary direction, Ω_0, with unit amplitude, at the surface of a rigid sphere of radius r_0. The pressure on some point Ω on the surface of the sphere is given by[22]
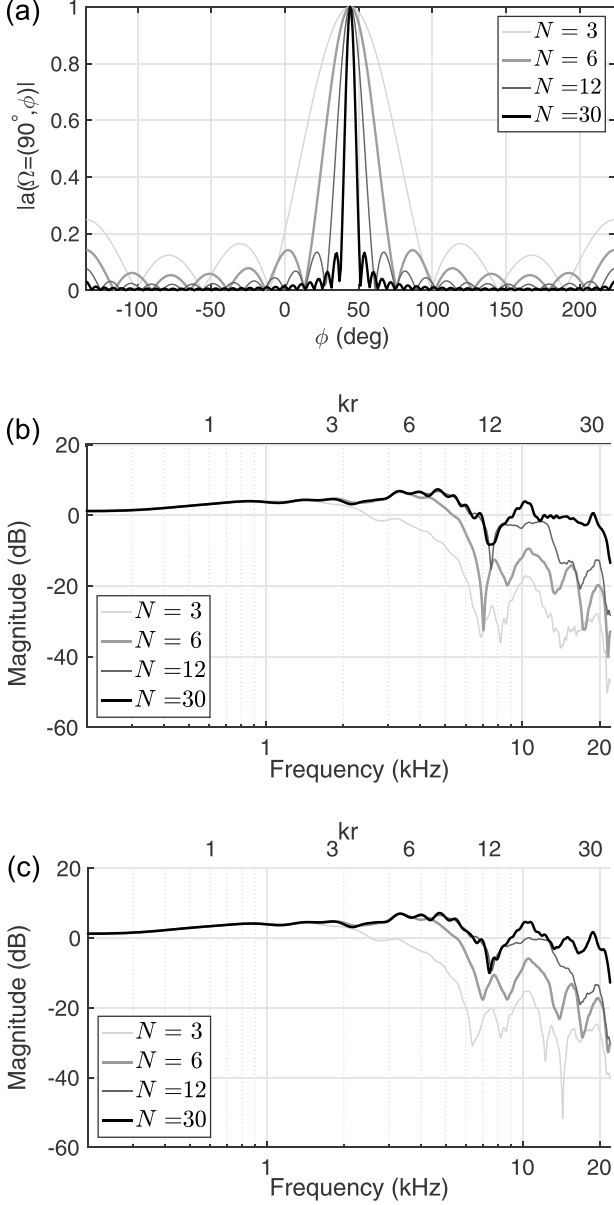
FIG. 1. Magnitude of (a) the PW density function of a sound field composed of a single unit-amplitude PW with incident direction $\Omega_0 = (90°, 45°)$, (b) left ear HRTF at the direction of the incoming PW, $h^l[k, \Omega = (90°, 45°)]$, and (c) the produced left ear pressure, $p^l(k)$, composed of $a(k, \Omega)$ from (a) and the $h(k, \Omega)$ from (b). All are truncated at different SH orders. In (b) and (c) kr was calculated assuming $r = 8.75$ cm.

$$p(kr_0, \Omega, \Omega_0) = \sum_{n=0}^{N} \sum_{m=-n}^{n} b_n(kr_0) \left[Y_n^m(\Omega_0) \right]^* Y_n^m(\Omega), \quad (8)$$

where $b_n(kr_0)$ is a radial function, defined for a rigid sphere as[9]

$$b_n(kr_0) = 4\pi i^n \left[j_n(kr_0) - \frac{j_n'(kr_0)}{h_n'(kr_0)} h_n(kr_0) \right], \quad (9)$$

where $j_n(\cdot)$ is the spherical Bessel function, $h_n(\cdot)$ is the outgoing spherical Hankel function and $j_n'(\cdot)$ and $h_n'(\cdot)$ represent their first derivatives. Assuming a reverberant setting, the

sound field can be considered as diffuse, i.e., as having equal power from all directions; this provides an analytic means to quantify the frequency related effects by constructing an incident wave with a finite number of SH, and averaging over all incident directions. Assuming that a receiver, representing the sound pressure at the ear, is placed at some point Ω on the surface of a rigid sphere, the average response magnitude over all incident wave directions, Ω_0, can be written as[23]

$$\bar{p}(kr_0, \Omega) = \sqrt{\frac{1}{4\pi} \int_{\Omega_0 \in S^2} |p(kr_0, \Omega, \Omega_0)|^2 d\Omega_0}. \quad (10)$$

Substituting Eq. (8) and applying the orthogonality property of SH and the addition theorem,[22] Eq. (10) can be written as function of the SH order, N,[14]

$$\bar{p}(kr_0)|_N = \frac{1}{4\pi} \sqrt{\sum_{n=0}^{N} (2n+1)|b_n(kr_0)|^2}. \quad (11)$$

Figure 2 shows the average pressure for different SH orders, computed for a rigid sphere with radius $r_0 = 8.75$ cm. It has been shown[9] that for the frequency range of $kr > N$, the order-truncation error becomes increasingly significant. This can explain the rapid decrease in the magnitude of the sound pressure in Fig. 2, and the deviation from the true pressure for frequencies in the range that satisfies $kr > N$. This behavior is a result of the nature of the radial functions, $b_n(kr)$. As the frequency increases, the spatial variation of the sound pressure around the sphere becomes more complex, and high SH orders are required for its accurate representation.

To correct this low-pass effect on the binaural signal due to order truncation, the authors previously suggested the Spherical Head Filter (SHF) method for timbre correction.[13] The filter, applied directly to the binaural signal, equalizes the magnitude of a low-order binaural signal to a desired, normally higher-order, signal. This method is based on the rigid sphere approximation for the human head and on the average response of a diffuse field, as calculated in Eq. (11). Consider the response of two order limited signals with different SH orders. A correcting equalization filter from the lower to the higher order will attempt to restore the distorted timbre in the lower order signal. Accordingly, one can
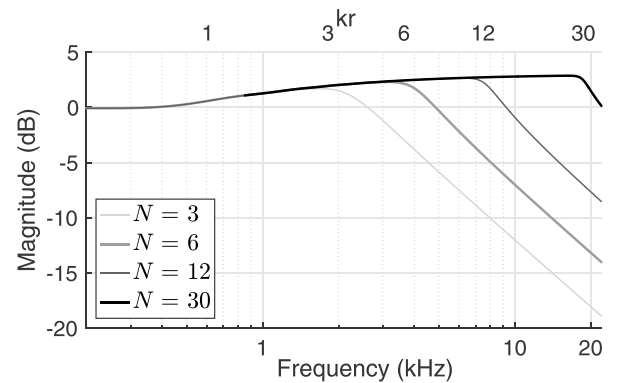

FIG. 2. The averaged magnitude of the pressure at some point on a rigid sphere due to a diffuse sound field for different SH orders.

describe the transfer function of a timbre correction filter, which equalizes the frequency response of the low order signal (corresponding to order N) to that of the high-order one (with order N_{high}), as follows:

$$G_{\text{SHF}}(kr)|_{N \to N_{\text{high}}} = \frac{\bar{p}(kr)|_{N_{\text{high}}}}{\bar{p}(kr)|_N}. \qquad (12)$$

Figure 3 shows examples for the SHF for different orders N equalizing to order $N_{\text{high}} = 30$, calculated for a rigid sphere of radius $r = 8.75$ cm. As a second approach toward equalization, an HRTF based filter (HRF) can be obtained in an analogous way to the derivation of Eq. (12) by averaging the HRTF across all incident angles[14]

$$\bar{h}^{l,r}(k)|_N = \sqrt{\frac{1}{4\pi} \sum_{n=0}^{N} \sum_{m=-n}^{n} |h_{nm}^{l,r}(k)|^2}, \qquad (13)$$

followed by the division of high-order and order-truncated coefficients

$$G_{\text{HRF}}(k)|_{N \to N_{\text{high}}} = \frac{\bar{h}(k)|_{N_{\text{high}}}}{\bar{h}(k)|_N}. \qquad (14)$$

Equation (14) requires averaging across ears, i.e., $\bar{h}(k) = [\bar{h}^l(k) + \bar{h}^r(k)]/2$, to avoid the introduction of interaural changes. While this has the advantage of considering individual HRTF characteristics, the resulting zeroth order filter exhibits narrow peaks of high gain that might cause audible ringing artifacts,[24] which fade away with increasing truncation order. They result from peaks in $\bar{h}(k)|_{N_{\text{high}}}$ that are not canceled by the rather smooth low order function $\bar{h}(k)|_0$. SHF and HRF filters were analyzed in terms of energy differences in 39 auditory filter bands

$$\Delta G(H_1, H_2, f_c) = 10 \log \frac{\int C(f, f_c)|H_1(f)|^2 df}{\int C(f, f_c)|H_2(f)|^2 df},$$

with $50 \text{ Hz} \leq f, f_c \leq 20\,000 \text{ Hz}, \qquad (15)$
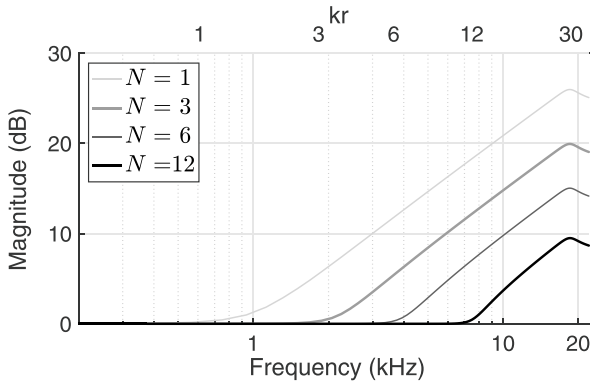
where C is a Gammatone filter with center frequency f_c in Hz, as implemented in the Auditory Toolbox,[25] and $H_1 = G_{\text{HRF}}$ and $H_2 = G_{\text{SHF}}$ are the equalization filters, as specified in Eqs. (12) and (14). The integral was evaluated between 50 Hz and 20 kHz and f_c was restricted accordingly. Please note that the calculation of Eq. (15) is similar to the calculation of the internal cochlea spectrum suggested by Salomons.[26] The HRF was calculated using Neumann KU100 HRTFs from the Colonge database.[20] Results shown in Fig. 4 suggest that differences between SHF and HRF filters are small for truncation orders of 3 or higher. In these cases, differences are well below 2 dB for frequencies below 9 kHz, and smaller than 4 dB above 9 kHz. Because this finding was also supported in an exploratory listening test,[14] where no significant differences were found between filter types, only SHF filters were considered in this study. An additional advantage of this choice is the fact that the SHF can be generated to arbitrarily high-orders, whereas catalog HRTFs may be of relatively low resolution and, therefore, may not be suitable for designing HRFs of high-orders.

The sphere radius was manually adjusted for a good fit to the KU100. However, it could also be estimated from the head related impulse responses (HRIRs) by using the time of arrival (TOA) model from Ziegelwanger and Majdak,[27] which calculates the TOAs using a rigid sphere model, and adjusts the parameters by minimizing the difference between modeled TOAs and the TOAs estimated directly from the HRIRs. While this is similar to the work of Algazi et al.,[28] who suggested using modeled and estimated interaural time differences (ITDs), the TOA model already comprises the TOA estimation and outlier removal. For the KU100, a radius of 9.1 cm was estimated by the TOA model, and a comparison between SHF and HRF filters for this radius yielded almost identical results as those displayed in Fig. 4 (not shown here). This implies a robustness of the SHF model to small changes in r_0 on one hand, and the validity of manually choosing and automatically estimating r_0 on the other. To assess the generalizability to other HRTF catalogs, SHF and HRF filters were calculated using HRIRs numerically obtained from a head only mesh from the FABIAN head and torso simulator[29] and mesh2HRTF.[30] SHF filters were calculated for a radius of 8.74 cm, as estimated by the TOA model. Differences between the two filter types are

FIG. 3. Magnitude response of the SHF equalizing order N to $N_{\text{high}} = 30$.
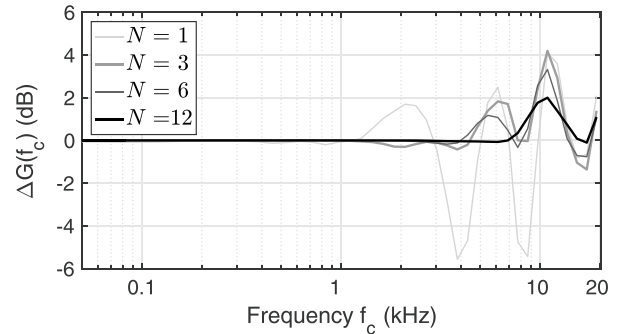

FIG. 4. Differences between SHF and HRF (KU100, $r_0 = 8.75$ cm) filters $\Delta G(G_{\text{HRF}}, G_{\text{SHF}}, f_c)$, computed using Eq. (15).
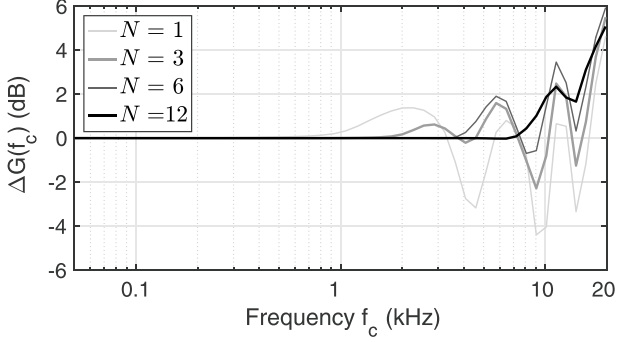
FIG. 5. Differences between SHF and HRF (FABIAN, $r_0 = 8.74$ cm) filters $\Delta G(G_{\text{HRF}}, G_{\text{SHF}}, f_c)$, computed using Eq. (15).

shown in Fig. 5. They are similar to the corresponding differences for the KU100, thus implying a good generalizability.

Since the suggested approach makes no assumptions concerning the signal phase, filters were designed with a linear phase using interpolation of the desired responses in the spectral domain. Then, by applying the inverse Fourier transform, the temporal filter coefficients were calculated and represented by 60 tap linear phase finite impulse response filters.

IV. PHYSICAL EVALUATION

In this section, an investigation of the effect of the equalization filters on the resulting binaural signals is presented. In a first step, these were assessed under the assumption of a diffuse field HRTF (i.e., having equal power from all directions, denoted here as the common transfer function, CTF) by computing Eqs. (7) and (10) for a sound field composed of uncorrelated PWs arriving from all directions. For this purpose, a data set consisting of 2702 HRTFs of a Neumann KU-100 manikin, taken from the Colonge database,[20] was used. The data set was acquired with a Lebedev sampling scheme,[31] which can produce HRTFs up to an order of 44.[10] The SH coefficients were computed using the discrete SFT.[16] Differences in auditory filters from the high-order CTF to order truncated CTFs without equalization, i.e., $\Delta G(\text{CTF}_N, \text{CTF}_{N_{\text{high}}}, f_c)$, and with equalization, i.e., $\Delta G(\text{CTF}_{N_{\text{SHF}}}, \text{CTF}_{N_{\text{high}}}, f_c)$, are shown in Fig. 6. Here, subscripts N and SHF denote the truncation order and the application of the equalization filter that was applied by convolution, i.e., element wise multiplication of the CTF and SHF frequency-domain spectra. Without equalization, a clear low-pass behavior with differences up to 28 dB can be observed, whereas for the equalized CTFs the differences are smaller than 6.5 dB and concentrated around 0 dB. Notably, differences below 5 kHz—a range that is essential for speech transmission—are smaller than 1 dB for truncation orders of 3 and higher.

Binaural signals for 770 single PWs with incidence angles distributed nearly-uniform across the sphere, using the Lebedev sampling scheme of order 23,[31] were calculated according to Eq. (7). These were then compared to the high-order reference $N_{\text{high}} = 30$ to assess the effect of the equalization filter on single PW sound fields, such as the direct sound and early reflections. In this case, the
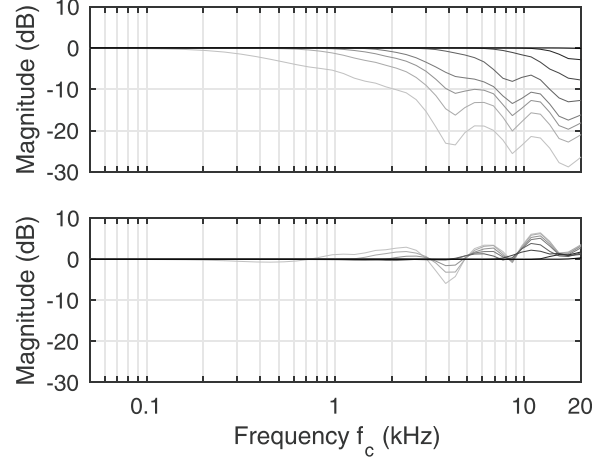

FIG. 6. Differences between high-order and order-truncated CTFs for truncation orders 0, 1, 2, 3, 6, 12, 20, and 29 without equalization (top) and with equalization (bottom). Darker colors denotes higher truncation orders.

assumption of diffuseness, on which the filter design was based, is clearly violated. Differences from the high-order reference to order-truncated left ear HRTFs without equalization, i.e., $\Delta G(p^l_N(f), p^l_{N_{\text{high}}}(f), f_c)$, and with equalization, i.e., $\Delta G(p^l_{N_{\text{SHF}}}(f), p^l_{N_{\text{high}}}(f), f_c)$, were calculated, resulting in 39 (number of auditory filters) $\times$ 770 (number of PWs) values for each truncation order. Here, $p^l_N(f)$ is the pressure at the left ear calculated according to Eq. (7) up to order N, and $p^l_{N_{\text{SHF}}}(f)$ is the signal after SHF equalization. Frequency dependent absolute differences averaged according to the quadrature weights α across the 770 PWs ($\sum_{l=1}^{770} \alpha_l |\Delta G_l|$) are depicted in Fig. 7. Unsurprisingly, differences (i) are smaller for higher truncation orders, (ii) increase with frequency f_c, and (iii) are smaller if equalization is applied. However, while maximum errors of approximately 23 dB monotonically decrease with increasing truncation order if equalization is not applied, they are close to 5 dB for most truncation orders if equalization is applied. This suggests
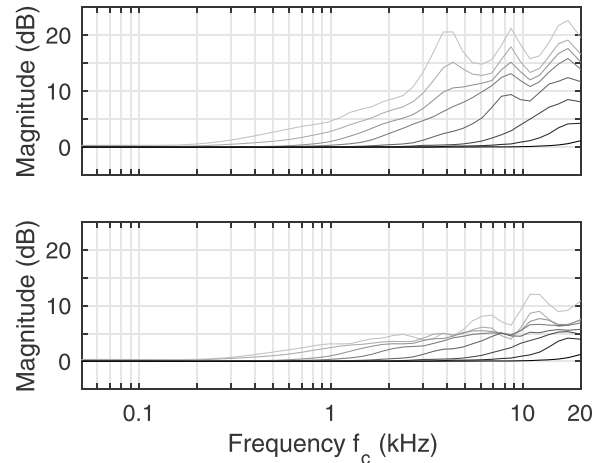

FIG. 7. Absolute differences between high-order and order-truncated left ear binaural signals for truncation orders 0, 1, 2, 3, 6, 12, 20, and 29 without equalization (top) and with equalization (bottom). Darker color means higher truncation order. Results were averaged across 800 PWs with different incident angles.

that the equalization is highly beneficial, especially for low truncation orders.

An example of incident angle dependent absolute differences averaged across the 39 auditory filters ($\frac{1}{39}\sum_{f_c}|\Delta G|$) is given in Fig. 8 for $N = 4$. In general, differences across source positions are smaller when equalization is applied, but the distribution of differences also changes. Without equalization, differences are smaller for contralateral source positions ($180 \leq \phi \leq 360$), and smallest for sources close to the left ($\theta = 90, \phi = 90$) and right ($\theta = 90, \phi = 270$) ear. The equalization inverts this pattern, and differences become larger for contralateral sources, whereas the largest differences shift to sources close the left and right ear. This suggests that the binaural signals for ipsilateral sources are more similar to the corresponding CTF—which was used for computing the equalization filter—than the signals for contralateral sources. As a side effect of equalization, the largest differences shift away from the median plane ($\phi = \{0, 180\}$), which might be practically advantageous, because listeners often move their head toward sources in natural listening conditions. Similar observations were made for different truncation orders, whereby the overall differences decreased with increasing truncation order. However, the large differences at the ipsilateral ear disappeared for high truncation orders if equalization was applied.

V. PERCEPTUAL EVALUATION

The results from Sec. IV imply that the equalization filter can improve the spectral roll-off of an order limited binaural signal, and shift the regions of largest errors to the side of the listener. To study the influence of the filter on the spatial perception of sound, a comprehensive listening test was conducted. It aimed to investigate the perceptual aspects relating to the filter design using the Spatial Audio Quality Inventory (SAQI),[32] which is an instrument used to evaluate the perceptual qualities of virtual acoustic environments.

A. Methodology

The binaural signals were rendered using the method described in Sec. II, again using Neumann KU-100 HRTFs.[20] Depending on the experimental condition, the HRTFs were truncated in the SH domain to the desired order. The equalization filters were computed as in Eq. (12), with radius $r = 8.75$ cm.

The PW density function, $a_{nm}(k)$, was simulated using the multi-channel room acoustic simulator (McRoomSIM),[33] for a rectangular room of dimensions $15.5 \times 9.8 \times 7.5$ m, as can be seen in Fig. 9, with a mean absorption coefficient $\alpha = 0.221$, and a mean $T_{20} = 1.086$ s. The spherical microphone array and an omni-directional source were placed as described in the figure at a height of 1.7 m. The critical distance is $r_{cd} = 1.742$ m, and the source is situated at a distance of $1.5 r_{cd}$ at an angle of $30°$ from the array, relative to the HRTF coordinate system. The signals were generated for head orientations between $-80°$ and $80°$ with a $1°$ resolution. This was achieved by rotating $h_{nm}^{l,r}(k)$ by the respective Wigner-D functions,[34] thus keeping the scene geometry constant and independent of the head movements.

Five perceptual qualities and six truncation orders were selected for perceptual testing based on informal listening. The zeroth order, evoking a perception similar to a mono recording, was used as a low anchor. Orders 1, 2, and 3 were chosen because they resemble cases of existing microphone arrays (e.g., Ambisoncs B-Format,[35] EigenMike[36]), and orders 6 and 12 were selected to cover the range where differences between the order truncated and high-order reference signals tend to become perceptually irrelevant, regardless of the equalization.

The perceptual qualities *overall difference* and *externalization* were directly taken from the SAQI catalogue, while the integrative qualities *coloration*, *source position*, and *source extension* were condensed from multiple SAQI items in order to limit the duration of the listening test. Coloration was chosen to assess the effect of the equalization filter on the timbre of the resulting signals, while source position and extension were expected to be related to the spatial
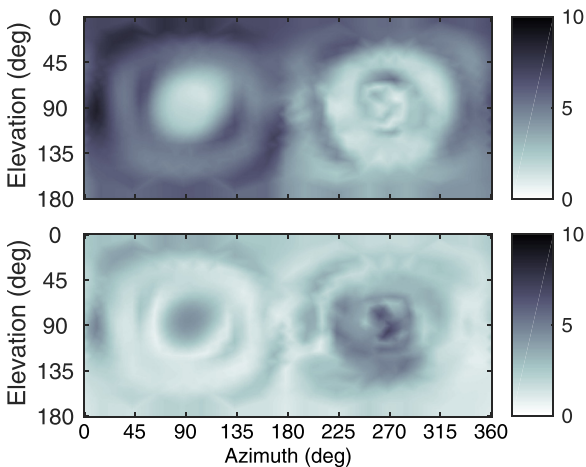

FIG. 8. (Color online) Absolute differences between order truncated ($N = 4$) and high-order ($N_{high} = 30$) left ear binaural signals calculated according to Eq. (7) without equalization (top) and with equalization (bottom). Color represents difference in dB. Results were averaged across 39 auditory filters.
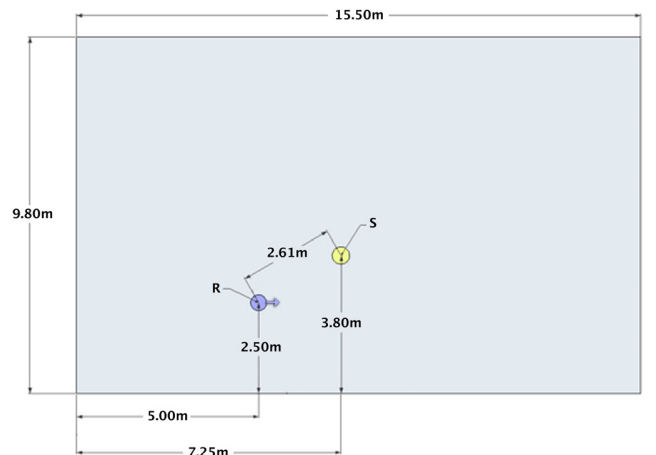

FIG. 9. (Color online) The room layout used in the McRoomSIM to simulate the sound field. R and S are the spherical microphone array and the source positions, respectively.

smoothing of the order truncated PW density function $a(k, \Omega)$, as seen in Fig. 1(a). In addition, two audio contents were used: pulsed pink noise (1 s of identical noise burst, 0.75 s pause, 20 ms fade in/out) was considered to be suitable for detecting overall differences, as well as differences in coloration and source position. An anechoic sample of German male speech (5 s) was included as a typical real life content for microphone array applications and to reveal differences in externalization and source extension.

To assess the effects of truncation order and equalization, manipulated stimuli were directly compared against a high-order reference ($N_{high} = 30$) using the Whisper[37] toolbox for MATLAB, enabling instantaneous switching between the stimuli. The reference order was chosen in accordance with the assumption of an upper wave-number limit of $kr = N$, when a typical HRTF of a human head with radius $r \approx 8\,$cm leads to an order limit of $N \approx 30$ at $f \leq 20\,$kHz. Moreover, the spatial detail achieved with a truncation order of 30 appears to be sufficiently high considering that (i) Ahrens et al.[38] reported that HRTFs of truncation order 15 were already perceptually identical to the original, and (ii) the HRTF after SH processing shown in Fig. 1(c) can be assumed to be perceptually identical to the original HRTF in Fig. 1(b)—the only by-products of the SH processing are a slight smoothing and an additional, but possibly imperceptible, notch at 15 kHz.[24] The presentation order of the stimuli was randomized across subjects, while being grouped in blocks with the same content. Dynamic auralization, including a real-time exchange of binaural room impulse responses (BRIRs) according to the current head position of the listener, was done using fWonder[39] and a high precision Polhemus Patriot head tracker. Sennheiser HD800 headphones were used for audio playback, and were compensated using *PEQ regularization*.[40] The listening test took approximately 1 h, including a training phase to familiarize the subjects with the user interface and stimuli. Subjects were explicitly instructed to move their head to different positions while listening and to take their time before giving an answer. To increase the reliability of the ratings, subjects could switch between the order truncated and reference stimuli as often and as fast or slow as they wanted, and compared six conditions in a multi slider rating interface at a time. For orientation, the sliders had numeric labels from -7 to 7 next to them, where a rating of 0 indicates that the test and reference stimuli were perceptually indistinguishable, and a rating of 7 (-7) indicates perceptual superiority of the test (reference) stimulus. Twenty subjects (3 female, median age 29) participated in the test, of which 14 had participated in listening tests before, and there were no reported known hearing impairments.

In total, 24 conditions were statistically tested in a three factorial, fully repeated measures MANOVA (multivariate analysis of variance) design, with the factors *truncation order* (0, 1, 2, 3, 6, 12), *equalization* (true, false), and *content* (noise, speech), as well as 5 dependent variables (overall difference, coloration, source location, source extension, externalization).

B. Results

Listening test results for all conditions and subjects are shown in Fig. 10 by means of raw and median ratings, and interquartile ranges. The ratings show relative differences, i.e., specify if an order truncated stimulus was more or less externalized in comparison with the high-order reference. As expected, a clear effect of the truncation order and equalization on (nearly) all dependent measures can be seen. Results showed consistently better ratings for conditions with equalization, and higher truncation orders. The noise content
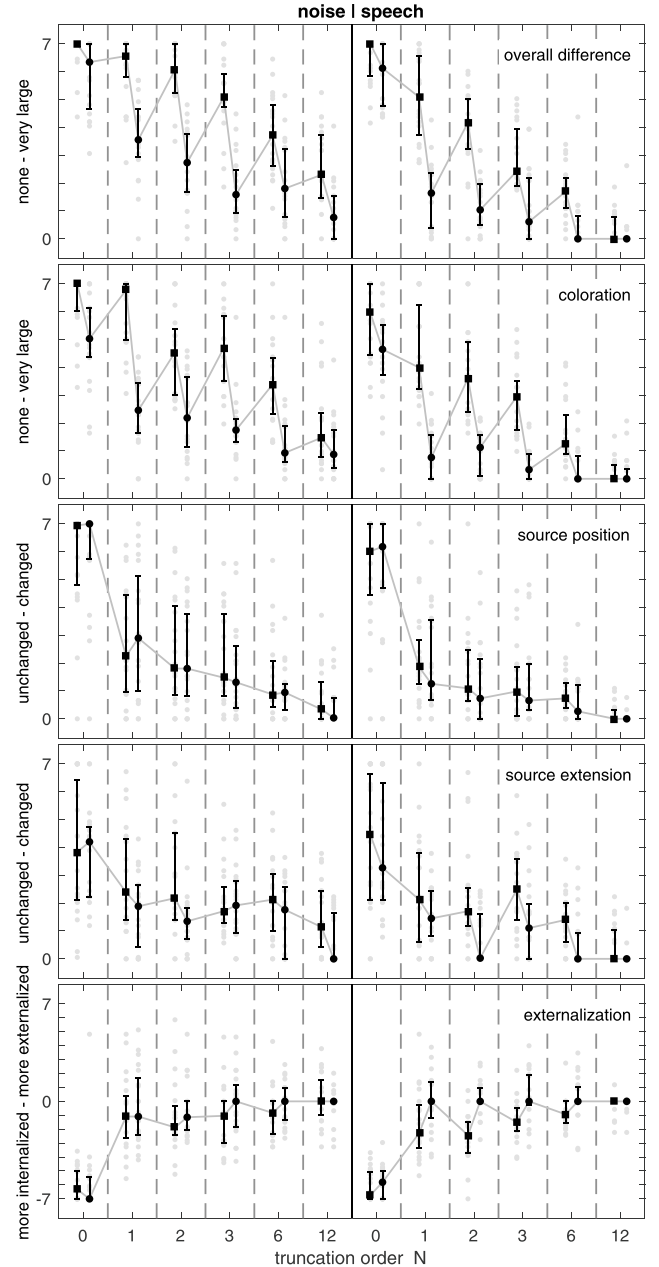
FIG. 10. Ratings for all subjects and test conditions. Perceptual qualities and scale labels are given in the text. Ratings for the noise and speech content are on the left and right, respectively. Individual ratings are given by light gray dots. Median ratings are given by squares (conditions without SHF equalization) and dots (conditions with SHF equalization). Interquartile ranges are given by solid lines. Connecting lines between conditions are given to improve readability.

FIG. 11. Standardized MANOVA model residuals for of all perceptual qualities (gray bars) and best fit normal distribution for comparison (black lines).

article.[41] For an introduction to multivariate statistical analysis, including MANOVA, see, for example, Raykov and Marcoulides.[42] A Lilliefors test showed that the MANOVA requirement of normally distributed model residuals was only met for the *overall difference*, but a visual inspection implied that the violations on the other variables were negligible (cf. Fig. 11). For a detailed analysis of the effects we conducted follow up univariate analyses of variance (ANOVA) with their type I error inflation being protected by the significant MANOVA result, and pairwise comparisons with Bonferroni correction.

The main effects, i.e., differences due to truncation order, equalization, and content, were significant for the vast majority of the perceptual qualities and factors (cf. Table I for an overview). They always resulted in better ratings for conditions with equalization and higher truncation orders. Pairwise comparisons for the truncation order show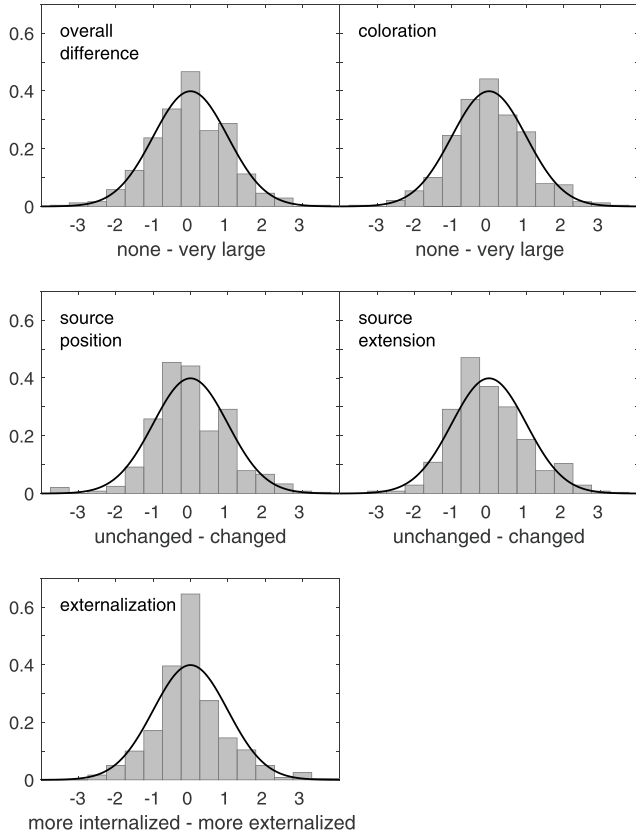ed that (i) differences between $N = 0$ and all remaining factor levels were significant, (ii) differences between all factor level combinations were significant for *overall difference* and *coloration*, (iii) differences between successive factor levels (i.e., $N = 3$ and $N = 6$) were not significant for *source position* and *source extension*, and (iv) only the difference between $N = 2$ and $N = 12$ was significant for *externalization*.

An overview of the obtained significant interaction effects is given in Table I along with the types of interaction, where an ordinal interaction means that the main effect remains unambiguously interpretable. The interaction effects involving the truncation order imply very much better ratings when equalization is applied and $N \geq 1$. Notably, median ratings for *externalization* were close to zero (no difference) for all conditions with equalization and truncation orders $N \geq 1$, whereas conditions without equalization and $N \leq 6$ were perceived to be more internalized (cf. Fig. 10). Disordinal but negligible interactions for the truncation order were observed in the case of $N = 0$ (*externalization*) and $N = 12$ (*source extension*). In these extreme cases, the effect of the equalization appears to be small, and the conditions without equalization yield slightly better ratings. The obtained interaction effects involving content and equalization were caused by (i) a stronger effect of the equalization on the ratings for the noise content (only pertaining to the *overall difference*), and (ii) a stronger effect of the equalization on the ratings for the speech content (only pertaining to the *externalization*).

produced higher differences for all perceptual qualities except *externalization*, where differences were larger for the speech content. Furthermore, more distinct response patterns can be observed for *overall difference* and *coloration* with pink noise, while speech yielded a clearer pattern for *source extension* and *externalization*. In the case of *source extension*, only absolute ratings were considered for statistical evaluation, because subjects did not agree on whether the source extension of the order truncated stimuli was smaller or larger compared to the reference.

The MANOVA test showed that all the major effects described above, as well as their first order interactions, were highly significant ($p \leq 0.01$, Pillai's trace). Please note that a detailed display of MANOVA results was omitted for brevity, but can be found in the digital addendum to this

TABLE I. Overview of significant main effects and first order interactions from follow up ANOVA (** $p \leq 0.01$; * $p \leq 0.05$).

	Main effects			Interactions		
	truncation order	equalization	content	order vs equalization	order vs content	content vs equalization
Difference	**	**	**	** ordinal	** ordinal	** ordinal
Coloration	**	**	**	** ordinal	* ordinal	
Position	**		**			
Extension	**	**	**	* hybrid[a]	** disordinal	
Externalization	**	**		** disordinal		** hybrid[b]

[a]Ordinal interaction of equalization; disordinal interaction of order.
[b]Ordinal interaction of equalization; disordinal interaction of content.

VI. DISCUSSION

Insights into the SH based reproduction of binaural signals, as outlined in Secs. II and III, showed two main effects of order truncation: reduced spatial resolution of the sound field represented by the PW density function $a(k, \Omega)$, and spectral roll-off in the resulting HRTFs. In Sec. IV, an analysis of the suggested equalization filter proved that the spectral roll-off can be accounted for if the underlying assumption of a diffuse sound field holds. This assumption is, however, clearly violated for the direct sound and early reflections of BRIRs, in which case the equalization decreases the overall error and shifts the region of largest errors to contralateral source positions. This shift might be perceptually beneficial as, for example, the ipsilateral ear is more important for source localization in sagittal planes.[43] Because the spectral shape plays an important role in the localization process,[44] it might be concluded that coloration at the contralateral ear is perceptually less relevant.

The perceptual evaluation reported in Sec. V confirmed the physical error analysis and revealed a positive effect of the equalization, not only with respect to reduced *coloration*, but also with significantly better conservation of *source extension* and *externalization*. In the case of externalization, this might be caused by the unnatural perception evoked by the spectral roll-off for conditions without equalization. This assumption can be supported with empirical evidence from Boyd *et al.*,[45] who discovered a decrease in externalization when applying a tenth order Butterworth low-pass filter with a cutoff frequency of 6.5 kHz to binaural signals recorded in a dry listening room ($T = 0.3$ s). Notably, order truncated stimuli were perceived to be fully externalized when equalization was applied. This is in contradiction to the results reported by Hartmann and Wittenberg,[46] who found the exact spectral shape (which is affected by the order truncation regardless of the equalization) to be required for externalization. In our case, the change in spectral shape might be counter balanced by the combination of motion cues provided by the dynamic binaural synthesis, and spatial cues inherent to the early and late reflections.[47] The effect on the *source extension* might also be caused by the unnatural, low-pass like characteristic of the conditions without equalization. Interestingly, no significant effect of the equalization on *source position* was observed. This might be due to the fact that the equalization does not change ITDs and interaural level differences whereas the effect on the derivative of the spectral shape with respect to frequency, which is essential for localization in sagittal planes,[44] can be assumed to be perceptually negligible, at least for the tested source position.

The effect of the truncation order was largest for *overall difference* and *coloration* which is indicated by the steep slope of the ratings, and can clearly be assigned to the spectral roll-off in the latter case. In the case of the noise content, coloration remains audible for the highest truncation order $N = 12$ even if equalization is applied. Since the physical evaluation suggested a nearly perfect equalization of the diffuse components for this condition, this can be attributed to remaining coloration in the direct sound and early reflections. This remaining coloration appears to be perceptually irrelevant for truncation orders $N \geq 3$ in the case of the speech content, where median ratings are close to zero, if equalization was applied. Considering the similarity between ratings for *overall difference* and *coloration*, and the smaller effect of the truncation order on the remaining qualities, the coloration appeared to be the dominating factor for the overall difference. Smaller effects of the truncation order on the perceived *source position*, *source extension*, and *externalization* were found if disregarding the low anchor condition of truncation order $N = 0$. Apparently, the relevant binaural cues are already coded in a perceptually convincing manner with low order SH representations. This is in accordance with the results from Romigh *et al.*,[48] who found out that accurate source localization is possible even if using HRTFs truncated to order 4.

VII. CONCLUSION

In the current study, we suggested a SH based approach for the spectral equalization of order-truncated binaural signals. The equalization filter was calculated as the ratio of high-order and truncated-order diffuse field HRTFs, which were modeled by a rigid sphere. In line with earlier studies, we showed that the simplifying assumption of using rigid sphere, rather than dummy head, HRTFs for the calculation of the equalization filter only slightly affects the performance of the equalization filter. Our perceptual evaluation showed significant improvements in the quality of a binaural simulation when equalization was applied. For male speech content, the coloration was irrelevant for a truncation order of 3 or higher, and small even for truncation orders of 1 and 2. Moreover, virtual sources were perceived to be fully externalized, and remaining differences in the source position and extension were small for truncation order 3, and 0 (on average) for truncation order 6 or higher.

Future work could evaluate the use of spherical or elliptical head and torso models for the calculation of equalization filters, which would make the approach more applicable to HRTFs measured on human subjects. Moreover, the proposed SHF filter might also be used for HRIR individualization by considering the geometry of the listener head during the design stage.

ACKNOWLEDGMENTS

[1]D. R. Begault and L. J. Trejo, *3-d Sound for Virtual Reality and Multimedia* (Academic Press Professional, Cambridge, MA, 1994), pp. 1–30.

[2]M. L. Hawley, R. Y. Litovsky, and J. F. Culling, "The benefit of binaural hearing in a cocktail party: Effect of location and type of interferer," J. Acoust. Soc. Am. **115**(2), 833–843 (2004).

[3]S. L. Gay and J. Benesty, *Acoustic Signal Processing for Telecommunication*, Vol. 551 (Springer Science & Business Media, New York, 2012), pp. 283–300.

[4]J. Blauert and J. Braasch, "Modeling binaural processing: What next?," in *Proceedings of Meetings on Acoustics*, Vol. 18, Acoustical Society of America (2014), p. 015005.

[5]H. Møller, "Fundamentals of binaural technology," Appl. Acoust. **36**(3), 171–218 (1992).

[6]B. Rafaely and A. Avni, "Interaural cross correlation in a sound field represented by spherical harmonics," J. Acoust. Soc. Am. **127**(2), 823–828 (2010).

[7]M. Jeffet and B. Rafaely, "Study of a generalized spherical array beamformer with adjustable binaural reproduction," in *IEEE 4th Joint Workshop on Hands-free Speech Communication and Microphone Arrays (HSCMA)* (2014), pp. 77–81.

[8]D. L. Alon and B. Rafaely, "Spherical microphone array with optimal aliasing cancellation," in *IEEE 27th Convention of Electrical & Electronics Engineers in Israel (IEEEI)* (2012), pp. 1–5.

[9]B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," J. Acoust. Soc. Am. **116**(4), 2149–2157 (2004).

[10]B. Rafaely, "Analysis and design of spherical microphone arrays," IEEE Trans. Speech Audio Process. **13**(1), 135–143 (2005).

[11]A. Avni, J. Ahrens, M. Geier, S. Spors, H. Wierstorf, and B. Rafaely, "Spatial perception of sound fields recorded by spherical microphone arrays with varying spatial resolution," J. Acoust. Soc. Am. **133**(5), 2711–2721 (2013).

[12]N. R. Shabtai and B. Rafaely, "Generalized spherical array beamforming for binaural speech reproduction," IEEE/ACM Trans. Audio, Speech, Lang. Process. **22**(1), 238–247 (2014).

[13]J. Sheaffer, S. Villeval, and B. Rafaely, "Rendering binaural room impulse responses from spherical microphone array recordings using timbre correction," in *EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany (2014), pp. 81–85.

[14]J. Sheaffer and B. Rafaely, "Equalization strategies for binaural room impulse response rendering using spherical arrays," in *IEEE 28th Convention Electrical & Electronics Engineers in Israel (IEEEI)* (2014), pp. 1–5.

[15]J. R. Driscoll and D. M. Healy, "Computing Fourier transforms and convolutions on the 2-sphere," Adv. Appl. Math. **15**(2), 202–250 (1994).

[16]B. Rafaely, "Fundamentals of spherical array processing," in *Springer Topics in Signal Processing*, Vol. 8 (Springer-Verlag, Berlin, 2015), pp. 1–189.

[17]D. Menzies and M. Al-Akaidi, "Nearfield binaural synthesis and ambisonics," J. Acoust. Soc. Am. **121**(3), 1559–1563 (2007).

[18]J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA, 1997), pp. 36–93.

[19]M. Park and B. Rafaely, "Sound-field analysis by plane-wave decomposition using spherical microphone array," J. Acoust. Soc. Am. **118**(5), 3094–3103 (2005).

[20]B. Bernschütz, "A spherical far field hrir/hrtf compilation of the Neumann ku 100," in *Proceedings of the 40th Italian (AIA) Annual Conference on Acoustics and the 39th German Annual Conference on Acoustics (DAGA)* (2013), p. 29.

[21]R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**(5), 3048–3058 (1998).

[22]E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic Press, New York, 1999), pp. 183–197.

[23]S. Villeval, "Spatial soundfield processing for headphone reproduction," M.S. thesis, Ben-Gurion University of the Negev, Israel, 2015.

[24]B. C. J. Moore, S. R. Oldfield, and G. J. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. **85**(2), 820–836 (1989).

[25]M. Slaney, "Auditory toolbox," version 2, Technical report #1998-010, Interval Research Corporation, University of Purdue (1998).

[26]A. M. Salomons, "Coloration and binaural decoloration of sound due to reflections," Doctoral thesis, Technical University Delft, The Netherlands, 1995.

[27]H. Ziegelwanger and P. Majdak, "Modeling the direction-continuous time-of-arrival in head-related transfer functions," J. Acoust. Soc. Am. **135**(3), 1278–1293 (2014).

[28]V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," J. Audio Eng. Soc. **49**(6), 472–479 (2001).

[29]F. Brinkmann, A. Lindau, M. Müller-Trapet, M. Vorländer, and S. Weinzierl, "Cross-validation of measured and modeled head-related transfer functions," in *German Annual Conference on Acoustics—DAGA* (2015), pp. 1118–1121.

[30]H. Ziegelwanger, W. Kreuzer, and P. Majdak, "Mesh2HRTF: An open-source software package for the numerical calculation of head-related transfer functions," in *22nd International Congress on Sound and Vibration*, Florence, Italy (July 2015).

[31]V.-I. Lebedev and A.-L. Skorokhodov, "Quadrature formulas of orders 41, 47 and 53 for the sphere newblock," Russian Acad. Sci. Dokl. Math **45**(3), 587–592 (1992).

[32]A. Lindau, V. Erbes, S. Lepa, H. J. Maempel, F. Brinkman, and S. Weinzierl, "A spatial audio quality inventory (saqi)," Acta Acust. Acust. **100**(5), 984–994 (2014).

[33]A. Wabnitz, N. Epain, C. Jin, and A. van Schaik, "Room acoustics simulation for multichannel microphone arrays," in *Proceedings of the International Symposium on Room Acoustics* (2010), pp. 1–6.

[34]B. Rafaely and M. Kleider, "Spherical microphone array beam steering using Wigner-d weighting," IEEE Signal Process. Lett. **15**, 417–420 (2008).

[35]P. G. Craven and M. A. Gerzon, "Coincident microphone simulation covering three dimensional space and yielding various directional outputs," U.S. patent 4,042,779 (August 16, 1977).

[36]MH Acoustics, 25 Summit Ave, Summit NJ 07901. em32 eigenmike microphone array release notes, February 2009. https://www.mhacoustics.com/products#eigenmike1 (Last viewed May 16, 2017).

[37]S. Ciba, A. Wlodarski, and H. J. Maempel, "Whisper—A new tool for performing listening tests," in *126th Audio Engineering Society Convention*, Convention Paper 7749, Munich, Germany (May, 2009).

[38]J. Ahrens, M. R. P. Thomas, and I. J. Tashev, "HRTF magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data," *IEEE Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Asia-Pacific (2012), pp. 1–5.

[39]A. Lindau, T. Hohn, and S. Weinzierl, "Binaural resynthesis for comparative studies of acoustical environments," in *122th Audio Engineering Society Convention*, Convention Paper 7032, Vienna, Austria (May, 2007).

[40]A. Lindau and F. Brinkmann, "Perceptual evaluation of headphone compensation in binaural synthesis based on non-individual recordings," J. Audio Eng. Soc. **60**(1/2), 54–62 (2012).

[41]See supplementary material at http://dx.doi.org/10.1121/1.4983652 for a detailed display of results for the statistical analyses.

[42]T. Raykov and G. A. Marcoulides, *An Introduction to Applied Multivariate Analysis* (Taylor & Francis Group, Routledge, New York, 2008), Chap. 1, pp. 1–30.

[43]M. Morimoto, "The contribution of two ears to the perception of vertical angle in sagittal planes," J. Acoust. Soc. Am. **109**(4), 1596–1603 (2001).

[44]R. Baumgartner, P. Majdak, and B. Laback, "Modeling sound-source localization in sagittal planes for human listeners," J. Acoust. Soc. Am. **136**(2), 791–802 (2014).

[45]A. W. Boyd, W. M. Whitmer, J. J. Soraghan, and M. A. Akeroyd, "Auditory externalization in hearing-impaired listeners: The effect of pinna cues and number of talkers," J. Acoust. Soc. Am. **131**(3), EL268–EL274 (2012).

[46]W. M. Hartmann and A. Wittenberg, "On the externalization of sound images," J. Acoust. Soc. Am. **99**(6), 3678–3688 (1996).

[47]D. R. Begault, E. M. Wenzel, A. S. Lee, and M. R. Anderson, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," J. Audio Eng. Soc. **49**(10), 904–916 (2001).

[48]G. D. Romigh, D. S. Brungart, R. M. Stern, and B. D. Simpson, "Efficient real spherical harmonic representation of head-related transfer functions," IEEE J. Sel. Topics Signal Process. **9**(5), 921–930 (2015).