Audio Engineering Society

Convention Paper 10690

Presented at the AES 156th Convention
2024 June 15–17, Madrid, Spain

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Filterbank rule for spherical microphone array encoding and EQs for above aliasing, transducers, and binaural playback

Franz Zotter

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

Correspondence should be addressed to Franz Zotter (zotter@iem.at)

ABSTRACT

Compact rigid-sphere microphone arrays are used to record 3D-sound scenes from one recording perspective. They are often used as the main microphone array in field recordings using Ambisonics. For time-invariant Ambisonic encoding, the array signals are typically first decomposed into spherical harmonic coefficient signals by a matrix, which are then equalized by radial filters, which need to be regularized by, e.g., a filter bank. This paper shows: (i) a simple way to determine common transducer responses for equalization based on comparison with an omnidirectional reference, (ii) a heuristic rule to define the filterbank frequencies depending on the Ambisonic order, diffuse-field-normalized weightings `basic` or `max-rE`, and available SNR, (iii) the appropriate handling of frequencies above the array's spatial aliasing limit, and finally while the proposed solution supports loudspeaker playback: (iv) binaural playback requires a different equalization.

1 Introduction

Scattering by rigid spheres was described by [1, Ch.XVII] and numerical evaluation with early computers can be found in [2], or much later [3]. Based on this theory, compact rigid-sphere microphone arrays as they are applied nowadays in spatial audio recording were described by [4, 5, 6, 7, 8, 9], and in textbooks [10, 11, 12, 13]. We will moreover find interest in the diffuse-field sensitivity curve of an ideal sound pressure sensor on a rigid sphere as described in [14], below.

Spherical microphone arrays such as the em32 or em64 Eigenmike or the Zylia ZM-1 can be used for beamforming or for what should be our main concern: to encode their transducer signals into Ambisonics, for spatial audio recordings.

Most recently, and probably most successfully for compact arrays, time-variant encoding processes depend on the recorded signals, they extract time-variant parameters, dynamically extracts and re-encodes directional plus residual signals, and there are several papers, e.g., [15, 16, 17], and the COMPASS and HO-DirAC plugins can be used.

To find a good baseline alternative for loudspeaker playback, our paper focus on a time-invariant encoding process and its remaining potentials for optimization. An ideal conversion of array signals into Ambisonic signals would deliver signals that would be captured by ideal spherical-harmonic far-field directivities. This ideal has often been posed as least-square error problem by some researchers, cf. [18, 19], which is favored by those papers proposing the solution and has some advantages, such as transducer-response equalization, but it also risks designing acoustic inaccuracies in the measurement, e.g. reflection and diffraction on mounting gear, into the encoding filters obtained.

More traditionally, based on the analytic Rayleigh scattering model of the rigid sphere [1, Ch.XVII], signals can be first decomposed into spherical harmonic coefficient signals by a static matrix, then treated by the so-called radial filters, e.g. [6, 10, 11, 12, 13], in order to equalize for phase and magnitude differences between different spherical harmonic orders. These theoretical radial filters are unstable by their n-fold pole at 0 Hz, and their practical realization was described in terms of recursive (IIR) discrete-time filters in second-order sections, obtained via bilinear transform and Tikhonov regularization [6], corrected impulse invariance and regularizing filter-banks that permit compensating amplitude losses [20], or non-recursive (FIR) implementations regularized by soft-knee thresholding [21]. More recent approaches also suggest to use a filter bank and individual, diffuse-field compensated order weights in every band [22, 23]. In [24], a flat diffuse-field response was found to be a perceptually suitable equalization for playback on loudspeakers.

The scattering-model-based encoding has the advantage that all equations are described analytically, however, there is no simple model to set the filterbank frequencies for regularization given the available SNR [22]. Secondly, so far, simple means of equalizing the common transducer responses to a reference microphone are desirable. Thirdly, the continuous scattering theory does not directly yield a way yet to treat the frequency region above the spatial aliasing limit, which can be expected above 3 kHz, 5 kHz, or 8 kHz, depending on the spherical array (ZM-1, em32, em64). And finally, it appears that binaural rendering on headphones requires a different equalization strategy: Recordings seem to sound fine when decoded on with less sophisticated Tikhonov or soft-knee regularization without diffuse-field equalization.

Secs. 2-4 cover the basics of the filterbank encoding method [13], Sec. 5 proposes a simple rule to get its cutoff frequencies, Sec. 6 proposes diffuse-field equalization for frequencies above the array's spatial aliasing limit, Sec. 7 proposes a new, single-measurement procedure to equalize the common transducer responses, and Sec. 8 discusses the spectral discrepancies obtained between loudspeaker and headphone playback.

2 Rigid-Sphere Scattering

In the frequency domain, the sound pressure observed at an angular offset ϕ from the arrival direction of a plane wave can be analytically formulated as, cf. [25],

$$p(ka,\cos\phi) = \sum_{n=0}^{\infty} \frac{(2n+1)\,\mathrm{i}^{n+1}\,P_n(\cos\phi)}{(ka)^2\,h_n'^{(2)}(ka)}, \quad (1)$$

where the radius of the sphere is a, the wave number $k=\frac{\omega}{c}$, the speed of sound c, and $\omega = 2\pi f$ denotes the dependency on the frequency, $P_n(\cos\phi)$ are the Legendre polynomials, and $h_n'^{(2)}$ is the derived spherical Hankel functions of the second kind and order n.

Based on the spherical harmonic addition theorem $(2n+1)P_n(\boldsymbol{\theta}^\top\boldsymbol{\theta}_\mathrm{s}) = \sum_{m=-n}^{n} 4\pi\, Y_n^m(\boldsymbol{\theta})Y_n^m(\boldsymbol{\theta}_\mathrm{s})$, scattering can also be described in spherical harmonics

$$p(ka,\boldsymbol{\theta}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \frac{4\pi\,\mathrm{i}^{n+1}\,Y_n^m(\boldsymbol{\theta})\,Y_n^m(\boldsymbol{\theta}_\mathrm{s})}{(ka)^2\,h_n'^{(2)}(ka)}, \quad (2)$$

$$p_\chi(ka,\boldsymbol{\theta}) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \frac{\mathrm{i}^{n+1}\,Y_n^m(\boldsymbol{\theta})\,\chi_{nm}}{(ka)^2\,h_n'^{(2)}(ka)}, \quad (3)$$

where Y_n^m are the spherical harmonics (SH) of the order n and degree m at the direction vector $\boldsymbol{\theta} = [\cos\varphi\sin\vartheta,\ \sin\varphi\sin\vartheta,\ \cos\vartheta]^\intercal$, the plane wave direction is $\boldsymbol{\theta}_\mathrm{s} = [\cos\varphi_\mathrm{s}\sin\vartheta_\mathrm{s},\ \sin\varphi_\mathrm{s}\sin\vartheta_\mathrm{s},\ \cos\vartheta_\mathrm{s}]^\intercal$, where φ, ϑ denote the azimuth and zenith, respectively, and $\boldsymbol{\theta}^\intercal\boldsymbol{\theta}_\mathrm{s} = \cos\phi$. In the second line, spherical harmonics $Y_n^m(\boldsymbol{\theta}_\mathrm{s})$ and 4π were replaced by frequency-domain Ambisonic signals $\chi_{nm}(\omega)$, i.e. coefficients of a spherical-harmonic plane-wave decomposition.

3 SH-Encoding and Radial Filters

Spherical arrays receive M sound pressure signals p_l picked up their M transducers. We will work in the frequency domain here, for convenience, but the expansion into spherical harmonics by the coefficients ψ_{nm} works equivalently in both time-domain and frequency domain. The signals p_l received should correspond to a linear combination of the spherical harmonics with unknown coefficients ψ_{nm}, with the spherical harmonics sampled at the transducer locations, specified by $\boldsymbol{\theta}_l$. This equivalence $p_l = \sum_{n=0}^{\mathrm{N}}\sum_{m=-n}^{n}\psi_{nm}Y_n^m(\boldsymbol{\theta}_l)$ is formulated using matrix vector notation containing the ordered, sampled harmonics $\boldsymbol{Y} = [Y_n^m(\boldsymbol{\theta}_l)]$, coefficients $\boldsymbol{\psi} = [\psi_{nm}]$, and pressures $\boldsymbol{p} = [p_l]$,

$$\boldsymbol{p} = \boldsymbol{Y}\boldsymbol{\psi}, \quad (4)$$

and left inversion of $\boldsymbol{Y}$ yields the desired coefficients

$$\boldsymbol{\psi} = \boldsymbol{Y}^\dagger\boldsymbol{p}, \quad (5)$$

so that they minimize $\|\boldsymbol{p}-\boldsymbol{Y}\boldsymbol{\psi}\|^2$. Inversion typically exists at a suitably low order of decomposition N, which can get at most $\mathrm{N}\le\sqrt{M}-1$, and requires a regular microphone layout to permit high decomposition orders, cf. [7, Eq. (3.34)], or for $t\ge 2\mathrm{N}$ designs [10] and N3D it simplifies to $\boldsymbol{Y}^\dagger=\frac{4\pi}{M}\boldsymbol{Y}^\mathsf{T}$. Due to the signal model Eq. (3), high-order coefficients are strongly attenuated at low frequencies. Ambisonic signals χ_{nm} are filtered versions of ψ_{nm} using the so-called radial filters $\hat\rho_n(\omega)$

$$\chi_{nm}(\omega)=\overbrace{\mathrm{i}^{-n-1}\,(ka)^2\,h_n'^{(2)}(ka)}^{\hat\rho_n(\omega)}\;\psi_{nm}(\omega) \tag{6}$$

that enhance high orders and hereby restore a focused directional mapping. However, these filters are unstable with an n-fold pole at $ka=0$, see Fig. 1. They can only be implemented safely when allowing the directional resolution to successively increase with frequency.

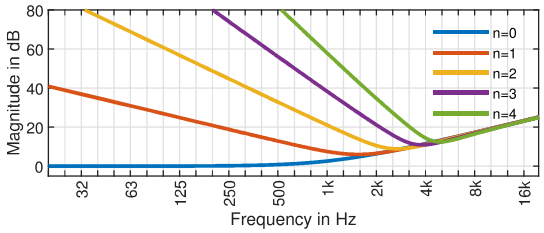


Fig. 1: Analytic radial filters for an array radius $a=0.05$ from Eq. (6) for $0\le n\le 4$.

4 Filterbank-Regularized Radial Filters

In [22], a bank of linear-phase bandpass filters was proposed that cuts out each frequency band b, in which the microphone array data can be resolved for all orders $b\le n\le\mathrm{N}$. With diffuse-field-equalized order weights $w_{n,b}$ and band-pass filters $H_b(\omega)$, these practical radial filters become, cf. [13],

$$\rho_n(\omega)=\left[\sum_{b=n}^{\mathrm{N}} w_{n,b}\,H_b(\omega)\right]\mathrm{i}^{-n-1}\,(\mathrm{ka})^2\,h_n'^{(2)}(\mathrm{ka})\,e^{\mathrm{i ka}}. \tag{7}$$

The phasor $e^{\mathrm{i ka}}$ removes the linear phase of $h_n'^{(2)}$, and the band-specific order weight $w_{n,b}$ is

$$w_{n,b}=\begin{cases}\hat w_n\sqrt{\frac{\sum_{n=0}^{\mathrm{N}}(2n+1)(\hat w_n)^2}{\sum_{n=0}^{b}(2n+1)(\hat w_n)^2}}, & \text{for } n\le b\\ 0, & \text{otherwise.}\end{cases} \tag{8}$$

This equalization was perceptually tested to perform best on loudspeakers in [24], and here it is used for across-band diffuse-field equalization of two alternative order weights $\hat w_n$,

$$\hat w_n=\begin{cases}P_n\big(\cos\frac{137.9^\circ}{b+1.51}\big), & \text{for max } -\boldsymbol{r}_\mathsf{E},\\ 1, & \text{for } \texttt{basic}.\end{cases} \tag{9}$$

The filter bank $H_b(\omega)$ is devised from zero-phase filters $\hat H_b(ka)$ that are equalized for perfect reconstruction and shifted by a suitable causal time offset τ

$$H_b(\omega)=\frac{\hat H_b(ka)}{\sum_{b=0}^{\mathrm{N}}\hat H_b(ka)}\,e^{-\mathrm{i}\omega\tau}, \tag{10}$$

and to work for different array radii a, the filters $H_b(\omega)$ are written as band passes depending on the Helmholtz number ka instead of ω. The filters are band-pass for $1\le b\le\mathrm{N}-1$

$$\hat H_b(ka)=\frac{\left(\frac{ka}{ka_b}\right)^{b+1}}{1+\left(\frac{ka}{ka_b}\right)^{b+1}}\,\frac{1}{1+\left(\frac{ka}{ka_{b+1}}\right)^{b+2}}, \tag{11}$$

and we get two exceptions for $b=0$ to be purely low-pass and for $b=\mathrm{N}$ to be purely high-pass

$$\hat H_0(ka)=\frac{1}{1+\left(\frac{ka}{ka_0}\right)^2},\quad \hat H_\mathrm{N}(ka)=\frac{\left(\frac{ka}{ka_\mathrm{N}}\right)^{\mathrm{N}+1}}{1+\left(\frac{ka}{ka_\mathrm{N}}\right)^{\mathrm{N}+1}}.$$

5 Proposed Filterbank Frequency Rule

By the shape $(\frac{ka}{\sqrt{M}-1})^{-n}$ of the curves in Fig. 1 and a given SNR, we assume a limiting filterbank to exhibit logarithmically decreasing frequencies $ka_b\approx c\,e^{\frac{1}{b}}$. However, cutoffs need to be designed that the entire system of order-weighted radial filters $\rho_n(\omega)$ do not introduce excessive diffuse noise boosts above $10^{\frac{SNR}{10}}$. This overall noise boost is described relative to the zeroth-order filter $-\mathrm{i}(ka)^2\,h_n'^{(2)}(ka)=-(1+\mathrm{i}ka)$ as low-noise reference:

$$|G(\omega)|^2=\frac{\sum_{n=0}^{\mathrm{N}}(2n+1)|\rho_n(\omega)|^2}{|1+\mathrm{i}ka|^2}\le 10^{\frac{SNR}{10}}. \tag{12}$$

Optimal cutoffs frequencies maximize the boost $\frac{1}{\Omega}\int|G(\omega)|^2\,\mathrm{d}\omega$ by their minimization $ka_b\to\min$, subject to the available $SNR\ge\max_\omega 20\lg|G(\omega)|$ as limitation at any ω and staying ascending $ka_{b-1}<ka_b$.

By inspecting the results for various orders and SNRs, $\mathrm{N}=2,\ldots,7$, $SNR=6,\ldots,30\,\mathrm{dB}$, and both weightings, it was possible to come up with a generic rule to tune the cutoff frequencies ka_b

$$ka_b = \sqrt{2}\cdot\alpha_b\cdot\sigma^{\beta_b}, \tag{13}$$

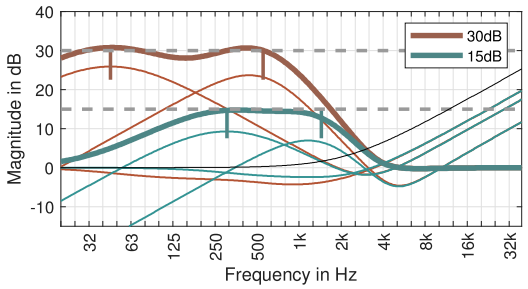
with specific values depending on the allowable noise boost SNR in dB, the maximum order N of decomposition, the band index b, and order weights, for instance

$$\sigma = 10^{\frac{-(SNR+3)}{20}}, \quad \alpha_b = \begin{cases} 1+\frac{b}{2}, & \text{for max}-\boldsymbol{r}_\mathrm{E}, \\ 1+\frac{b}{\sqrt{2}}, & \text{for } \texttt{basic}, \end{cases} \tag{14}$$

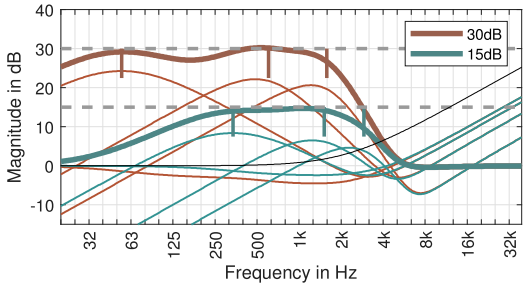
$$\beta_b = \begin{cases} \frac{1+\frac{1}{(\mathrm{N}+1)^2}}{b+1}, & \text{for max}-\boldsymbol{r}_\mathrm{E}, \\ \frac{1+\frac{0.5}{(\mathrm{N}+1)^2}}{b+1}, & \text{for } \texttt{basic}. \end{cases} \tag{15}$$

These values provide a reasonably accurate control $-0.3\,\mathrm{dB} \le \max_\omega 20\lg|G(\omega)| - SNR \le 0.65\,\mathrm{dB}$ for any practical purpose with $2 \le \mathrm{N} \le 7$.

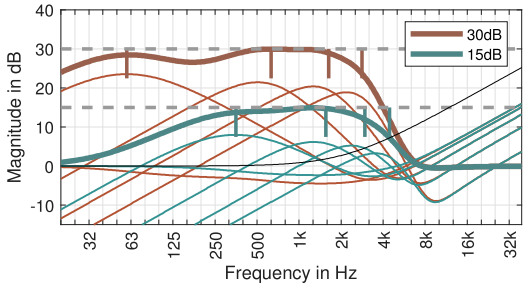
Fig. 2 shows the results for $|G(\omega)|^2$ for different settings of SNR, the filters $\rho_n(\omega)$, and zeroth-order reference, for max-$\boldsymbol{r}_\mathrm{E}$ weights and the radius $a=0.05$ m.



(a) N=2



(b) N=3



(c) N=4

Fig. 2: Noise boosts $|G(\omega)|^2$ due to cutoffs ka_b for $SNR=\{15,30\}$ dB (bold) for max-$\boldsymbol{r}_\mathrm{E}$, radial filters $\rho_n(\omega)$ for $a=0.05$ m (thin,color), cutoffs $f_b=\frac{c(ka_b)}{2\pi a}$ (vertical), and zero-order reference (thin,black).

6 Proposed Equalization above Aliasing

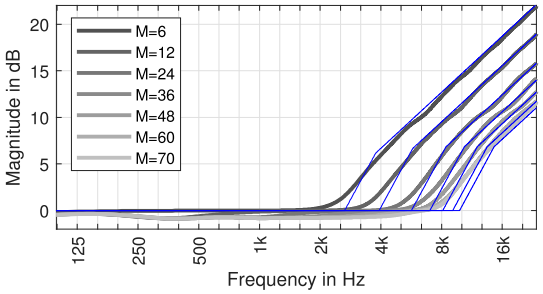
The radial filters in Figs. 1,2 exhibit first-order high-pass slopes for frequencies above $ka \ge \mathrm{N}$, where spatial aliasing starts, as e.g. shown in [26, Sec.3.8.2].

The Ambisonic signal spectra due to very-high-order spherical harmonic sound fields received by the array, e.g. with orders $\mathrm{N}_\mathrm{sim}=25$, can be modeled as

$$\boldsymbol{\chi}_\mathrm{diff}(\omega) = \mathrm{diag}\{[\rho_n(\omega)]\}\boldsymbol{Y}_\mathrm{N}^\dagger \boldsymbol{Y}_{\mathrm{N}_\mathrm{sim}}\,\mathrm{diag}\{[\tfrac{\mathrm{i}^{n+1}}{(ka)^2 h_n'^{(2)}(ka)}]\}, \tag{16}$$

to consider spatially aliased components in their physical strength. With sufficiently high simulation order $\mathrm{N}_\mathrm{sim} \ge \max ka$ and assuming ideal transducers,

Fig. 3: Diffuse-field sensitivities (bold) and models (thin) of filterbank-regularized encoders for t-designs of $M=[6,12,24,36,48,60,70]$ microphones, max-$\boldsymbol{r}_\mathrm{E}$ weights, and orders $N=1\ldots7$.



the Frobenius norm of the matrix $\boldsymbol{\chi}_{\text{diff}}$ expresses the diffuse-field sensitivity of the encoded array signals

$$R_{\text{diff}}(\omega) = \|\boldsymbol{\chi}_{\text{diff}}(\omega)\|_{\text{Fro}}. \tag{17}$$

Fig. 3 displays the resulting curves for the $SNR = 30$ dB and max-$\boldsymbol{r}_{\text{E}}$ setting and various t-designs as microphone layouts, assuming an array radius of $a = 0.05$ m. The blue curves are rough models consisting of a constant part, a 12 dB/oct., and a 6 dB/oct. high-pass slope,

$$\hat{R}_{\text{diff}}(ka) = \max\{\min\{R_1(ka), R_2(ka)\}, 1\}, \tag{18}$$

$$R_1(ka) = \left[\frac{ka}{1.1(\sqrt{M}-1)}\right]^2,$$

$$R_2(ka) = \frac{ka}{0.75(\sqrt{M}-0.15)}.$$

We neglect the influence of the SNR setting, the order weights, and the decomposition order N, here. The inverse of Eq. (18) provides a specification for low-pass filters to equalize the diffuse-field sensitivity above the spatial aliasing frequency $ka \geq N$.

7 Proposed Transducer Equalization

Despite a rather accurate matching of array transducers by the manufacturer can be assumed, there is still an overall frequency response common to all transducers, which should be equalized. We propose a measurement procedure delivering equalization of the common response with regard to a reference microphone. The procedure uses the array, a broadband loudspeaker, and an omnidirectional microphone, with no need for an anechoic chamber or truncating room impulse responses.

We propose the sum of the squared array transducer magnitude spectra as a notional diffuse field: the M matched transducers receive the same signals as one array transducer would receive if a perfect copy of the rotated surrounding sound field could be sequentially reproduced in M rotations. The sum of squares models incoherent superposition and perfect rotational copies model isotropy of an ideal diffuse field; be it from direct original fields, or even more so for already diffuse ones.

This notion can only be utilized for equalization when taking into account that diffuse-field response of an ideal sound pressure transducer on the rigid sphere is not constant. Peterson [14] computed this response, see crosses in Fig. 4. It is formally expressed as

$$H_{\text{diff}}(\omega) = \sum_{n=0}^{\infty} \frac{2n+1}{(ka)^4\,|h_{n'}^{\prime(2)}(ka)|^2}, \tag{19}$$

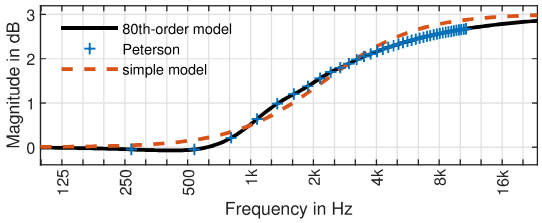


Fig. 4: Diffuse-field sensitivity of a rigid sphere (radius assumed was $a = 0.05$ m) modeled with N = 80 order diffraction formula Eq. (19), and Eq. (20)

cf. [25], see black curve in Fig. 4. Because the series converges poorly at high frequencies, a simplified practical version of reasonable accuracy (± 0.2 dB) is shown in red,

$$\hat{H}_{\text{diff}}(\omega) = \frac{1+\sqrt{2}\,\mathrm{i}\,\frac{ka}{\sqrt{2\pi}}}{1+\mathrm{i}\,\frac{ka}{\sqrt{2\pi}}}. \tag{20}$$

The curves Eq. (20) or Eq. (19) approach 3 dB at $ka = \sqrt{2\pi}$ and define the spectral difference that the superimposed magnitude-square transducer spectra $|p_l(\omega)|^2$ should have after equalization to the spectrum $|p_{\text{ref}}(\omega)|^2$ of an omnidirectional reference microphone. Accordingly, the common transducer frequency response is found by dividing with both the omnidirectional reference microphone spectrum and diffuse-field response $H_{\text{diff}}(\omega)$ of the ideal transducer on the sphere:

$$H_{\text{mic}}(\omega) = \frac{\frac{1}{M}\sqrt{\sum_{l=1}^{M}|p_l(\omega)|^2}}{|H_{\text{diff}}(\omega)\,p_{\text{ref}}(\omega)|}. \tag{21}$$

Reasonably smooth response curves are obtained with nth-octave smoothed measured room responses $|p_l(\omega)|^2$ in numerator and denominator $|p_{\text{ref}}(\omega)|^2$.

7.1 Measurement and EQ for em32 and ZM-1

We took impulse response measurements with an Earthworks 1/8" reference microphone, a Genelec 8020 loudspeaker at 1.5 m distance at IEM CUBE (RT$\approx$ 0.5 s), using a swept-sine sweep signal, and repeated the measurement with em32 Eigenmike replacing the measurement microphone and then the Zylia ZM-1. After deconvolution with the measurement signal, the impulse responses of the array transducers and the reference microphone were truncated to 0.5 s. Their $p_l(\omega)$, $p_{\text{ref}}(\omega)$ spectra were 6^{th}-octave-smoothed, then combined with H_{diff} for $a = 0.042$ m (em32) or $a = 0.049$

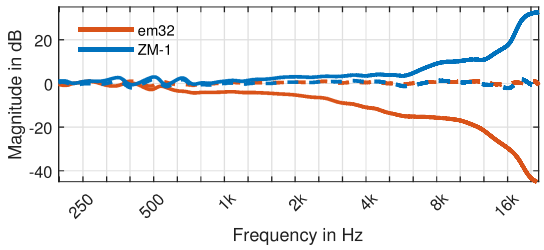


Fig. 5: Common magnitude responses of the Zylia ZM-1 and em32 Eigenmike array transducers (solid), and equalized responses (dashed), using Tab.1.

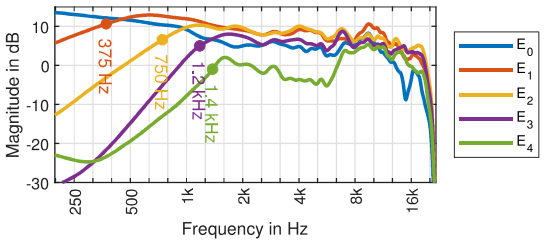


Fig. 6: Ambisonic HRTF order sensitivities of 4th-order MagLS.

(ZM-1) as in Eq. (21). Fig. 5 shows the resulting array magnitude responses H_{mic} for the ZM-1 (MEMS) and em32 (electret condenser), and the responses after equalization by 2nd order peak filters of Tab. 1.

8 Equalizing for Binaural Playback

The equalization target of a flat diffuse-field response as obtained by a flat power-spectral density of all the encoded spherical harmonic signals, above, primarily works for loudspeakers in rooms. Unfortunately, this does not deliver a spectrally flat playback on headphones. Unlike decoding to loudspeakers in a room, the HRTFs of a binaural decoder are not equally transparent to all spherical harmonic orders as shown in Fig. 6 for $N=4$ and the order energies of the MagLS binaural decoding filters $B_{nm}^{(l)}(\omega)$ for the both ears $l=1,2$,

$$E_n(\omega) = \sum_{m=-n}^{m} \sum_{l=1}^{2} |B_{nm}^{(l)}(\omega)|^2; \qquad (22)$$

suitable cut-on frequencies for $n=1,2,3,4$ could be 375 Hz, 750 Hz, 1.2 kHz, 1.4 kHz, but aren't necessarily supported by a `basic` spherical microphone array encoder. The spectral deficit can be severe and observed in detail by combining Eq. (16) with the $2\times(\mathrm{N}+1)^2$ binaural decoder $\boldsymbol{B}(\omega)$, yielding the diffuse-field sensitivity of the microphone-array, encoder, and binaural-decoder system

$$R_{\text{bin}}(\omega) = \frac{\|\boldsymbol{B}(\omega)\boldsymbol{\chi}_{\text{diff}}(\omega)\|_{\text{Fro}}}{\|\boldsymbol{B}(\omega)\|_{\text{Fro}}\|\boldsymbol{\chi}_{\text{diff}}(\omega)\|_{\text{Fro}}}, \qquad (23)$$

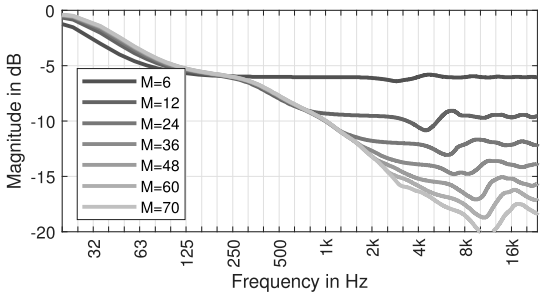
relative to the diffuse field sensitivity of the encoded microphone signals $\|\boldsymbol{\chi}_{\text{diff}}(\omega)\|_{\text{Fro}}$, see Eqs. (16),(19), times those of the mag-LS decomposed HRIRs $\|\boldsymbol{B}(\omega)\|_{\text{Fro}}$. Fig. 7 shows the resulting binaural diffuse-field sensitivities in the configurations used for Fig. 2. For a spherical array with $M=6$ and $N=1$, the respective frequency response in Fig. 7 has the characteristics of a high-shelf filter with -6 dB with cut-off at 25 Hz. For a larger number of microphones M and order N of the encoder, we see an additional shelving characteristics attenuating by additional $-4,-6,-8,-10,-12,-14$ dB at about 500 Hz. The resulting emphasis on the frequency range below 500 Hz is quite disturbing.

It appears reasonable to encode a spherical microphone array recording specifically for binaural playback. Weights should be `basic` and cut-on frequen-

Table 1: Parametric EQs for em32 and ZM-1

em32			ZM-1		
lvl dB	f Hz	Q	lvl dB	f Hz	Q
-1	400	0.8	-1.5	350	1
+3	800	0.8	-1.5	1.8k	1
+3	6k	0.8	-1.5	5k	1
-3	10k	1.2	+4	13k	3
+38	21k	6	-3.5	21k	2.5
+8	22k	5	-30	21k	0.17

Fig. 7: Diffuse-field sensitivities of binaurally decoded, equalized rigid-sphere encoder with `basic` weighting and SNR=30 dB.



cies either align with by the orders required to represent the HRTFs at every frequency, or different bands should use MagLS decoders of matching orders (not carried out, here). The most promising time-invariant alternative might be eMagLS2 [27].

While the proposed diffuse-field equalized encoding filters typically deliver loudspeaker playback with flat frequency response, in studio rooms, they only deliver a flat frequency response in headphone playback after employing shelving filters as a quick fix to equalize the pronounced bass boosts displayed in Fig. 7.

9 Conclusion

We presented a neat set of equation to obtain cutoff frequencies for time-invariant, regularizing encoding filters of compact spherical microphone arrays. We outlined a way to define diffuse-field equalization low-pass filters for the encoded signals above the spatial aliasing frequency. And we proposed a simple measurement to find equalization filters for the common transducer response of the array by comparison with a reference microphone, and we could provide useful filter settings for the em32 Eigenmike and Zylia ZM-1 from such a measurement. The equalization in the filterbank suits loudspeaker playback, but does not yield a balanced frequency response in the binaural decoding for playback on headphones, yet. We visualized the coloration that needs to be equalized as a quick fix for headphones and outlined more systematic alternatives.

The approach discussed here provides a very high level of control in terms of noise boost, sidelobe suppression, frequency response control, and it yields a signal-independent, time-invariant encoding of compact spherical arrays to be played back on loudspeakers.

Yet, the accomplishment does not fit the classical ideas of the higher-order Ambisonic production workflow based on directional encoding of single-channel signals: Frequency-specific order weighting at the encoder side contradict the advantage of truncating and tapering the higher orders only at the decoder end, specific to the capacities of the loudspeaker layout, while producing at the highest convenient Ambisonic order.

While this mainly requires unweighted `basic` decoding for the spherical microphone array signals and encoded mono sounds could be decoded on a different, sidelobe-suppressing max-r_E decoder, headphone playback is even more complicated: it requires a different encoding or filtering of the spherical microphone array than the loudspeaker playback.

All of this is inherently solved by the accomplishments of parametric audio processing, e.g. [17], i.e. adaptive, so time-variant and signal-dependent encoding strategies. Despite their advantages, such methods can sometimes cause time-varying artifacts or musical noise, and we hope that the methods presented here support the development of future baseline solutions that keep the need for adaptive processing small.

References

[1] Strutt, J. W., *Theory of Sound*, volume II, Macmillan and Co., London, 1878.

[2] Schwarz, L., "Zur Theorie der Beugung einer ebenen Schallwelle an der Kugel," *Akustische Zeitschrift*, VIII, pp. 91–117, 1943.

[3] Cooper, D. H., "Calculator program for head-related transfer function," *J. Audio Eng. Soc.*, 30(1/2), pp. 34–38, 1982.

[4] Meyer, J. and Elko, G., "A highly scalable sherical microphone array based on an orthonormal decomposition of the soundfield," in *Proc. IEEE ICASSP'02*, Orlando, 2002.

[5] Petersen, S. O., "Localization of sound sources using 3d microphone array," *Master's Thesis, University of South Denmark, Odense*, 2004.

[6] Daniel, J. and Moreau, S., "Further study of sound endcoding with higher order ambisonics," in *prepr. 6017, 116th AES Conv*, Berlin, 2004.

[7] Rafaly, B., "Analysis and design of spherical microphone arrays," *IEEE Transactions on Speech and Audio Processing*, 13(1), pp. 135–143, 2005.

[8] Song, W.-K., Ellermeier, W., and Hald, J., "Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise," *J. Acoust. Soc. Am., vol. 123, no. 2*, 129(2), pp. 910–924, 2008.

[9] Moreau, S., *Étude et réalisation d'outils avancés d'encodage spatial pour la technique de spatialisation sonore Higher Order Ambisonics: microphone 3d et contrôle de distance*, Ph.D. thesis, Université du Maine, 2006.

[10] Teutsch, H., *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*, Springer, 2007.

[11] Rafaely, B., *Fundamentals of Spherical Array Processing*, Springer, 2015.

[12] Jarrett, D. P., Habets, E., and Naylor, P. A., *Theory and Applications of spherical micrphone array processing*, Springer, 2017.

[13] Zotter, F. and Frank, M., *Ambisonics*, SpringerOpen, 2019.

[14] Peterson, A. P., "Sound diffraction for a spherical microphone," *J. Acoust. Soc. Am.*, 78(1), pp. 266–267, 1985.

[15] Politis, A. and Pulkki, V., "Higher-Order Directional Audio Coding," in V. Pulkki, S. Delikaris-Manias, and A. Politis, editors, *Parametric Time-Frequency Domain Spatial Audio*, pp. 141–159, Wiley&Sons, 2018.

[16] Politis, A., Tervo, S., and Pulkki, V., "COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes," in *2018 IEEE ICASSP*, pp. 6802–6806, 2018.

[17] McCormack, L., Politis, A., Gonzalez, R., Lokki, T., and Pulkki, V., "Parametric Ambisonic Encoding of Arbitrary Microphone Arrays," *IEEE/ACM TASLP*, 30, pp. 2062–2075, 2022.

[18] Farina, A., Capra, A., Chiesi, L., and Scopece, L., "A Spherical Microphone Array for Synthesizing Virtual Directive Microphones in Live Broadcasting and in Post Production," in *40th AES Conf*, Tokyo, 2017.

[19] Politis, A. and Gamper, H., "Comparing modeled and measurement-based spherical harmonic encoding filters for spherical microphone arrays," in *WASPAA-17*, NY, 2017.

[20] Baumgartner, R., Pomberger, H., and Frank, M., "Practical implementation of radial filters for ambisonic recordings," in *Proc. 1st ICSA*, Detmold, 2011.

[21] B. Bernschütz, C. P. and Spors, S., "Soft-limiting bei modaler Amplitudenverstärkung bei sphärischen Mikrofonarrays im plane-wave decomposition Verfahren," in *Fortschritte der Akustik - DAGA*, Düsseldorf, 2011.

[22] Zotter, F., "A linear-phase filter-bank approach to process rigid spherical microphone array recordings," in *IcETRAN*, Palić, 2018.

[23] Langrenne, C., Bavu, E., and Garcia, A., "A linear phase IIR filterbank for the radial filters of ambisonic recordings," in *EAA Spatial Audio Signal Processing Symposium*, pp. 127–132, Paris, 2019.

[24] F. Zotter, M. F. and Haar, C., "Spherical microphone array equalization for Ambisonics," in *Fortschritte der Akustik - DAGA*, Nürnberg, 2015.

[25] Jarrett, D. P., "Rigid sphere room impulse response simulation: Algorithm and applications," *J. Acoust. Soc. Am.*, 132, pp. 1462–1472, 2012.

[26] Bernschütz, B., *Microphone Arrays and Sound Field Decomposition for Dynamic Binaural Recording*, Ph.D. thesis, TU Berlin, 2016.

[27] Deppisch, T., Helmholz, H., and Ahrens, J., "End-to-End Magnitude Least Squares Binaural Rendering of Spherical Microphone Array Signals," in *Proc. I3DA*, Bologna, 2021.