Spherical microphone array equalization for Ambisonics

Franz Zotter[1], Matthias Frank[1], Christian Haar[2]
[1] *Institut für Elektronische Musik und Akustik*
[1,2] *Kunstuniversität Graz, Austria, Email: {zotter, frank}@iem.at*

Introduction

Microphone arrays on rigid spheres elegantly facilitate forming of higher-order spherical harmonic directivity patterns. Hereby, they also enable surround recording with height in higher-order Ambisonics [1]. While the necessary distortionless (either on-axis or diffuse-field equalized) and white-noise-gain constraints (WNG) for spherical-harmonic beamforming seem logical, we lack dependable knowledge about how to equalize spherical arrays for higher-order Ambisonic surround playback.

Processing based on the typical analytical superdirective model of spherical beamforming [2] would heavily amplify noise and mismatch in higher-order directional signals at low frequencies. To avoid enormous noise boosts and strongly mislocalized signal components at low frequencies, higher-order signals are gradually discarded towards low frequencies [1, 2, 3].

Fig. 1 shows spherical harmonic pickup patterns with gradually reduced expansion order, $b = 4 \ldots 0$. While the omni-directional component remains unity, diffuse and on-axis amplitudes suffer amplitude loss when reducing the spherical harmonic expansion order b. Song proposes free-field equalization [2] or measurements [4]; Baumgartner and Lösler discuss diffuse-field equalization [5, 6].

As this is most frequently done to obtain constant loudness in amplitude panning: *Shouldn't it be useful to assume that diffuse-field amplitude determines perceived loudness in large-scale Ambisonic surround playback?* Contrarily, bass over-emphasis dominated correspondingly processed playback of Eigenmike$^{\mathrm{TM}}$ recordings. Therefore, dependable facts are going to be explored here for clarification.

Listening Experiment

To clarify whether amplitude of the diffuse field, free field, omnidirectional component, or any quantity determines perceived frequency response characteristics, we undertook a listening experiment at IEM's anechoic chamber and IEM's CUBE.

To defeat any uncuntrolled influence of an acoustic recording situation, noise, microphone mismatch, etc., and to increase reproducibility, we based the search for a perceptual equalization curve on loudspeaker playback of Ambisonically amplitude-panned band-limited noise. Panning was done in 2D as well as 3D (using AllRAP [7]).

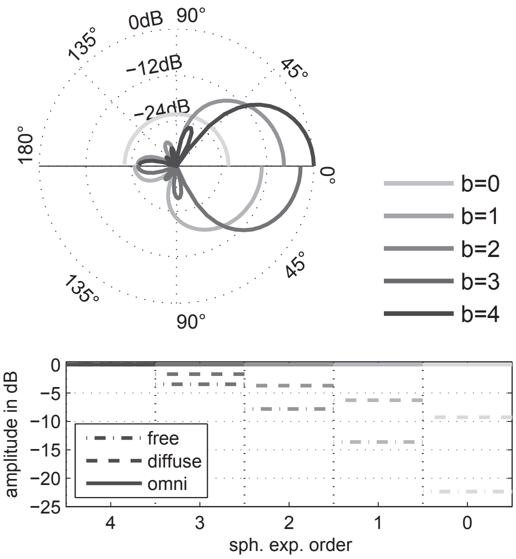

Figure 1: Spherical pickup pattern with gradually reduced spherical harmonics expansion order and their amplitudes: on-axis (free), diffuse-field, omnidirectional. *Which amplitude represents loudness of the pattern in surround playback?*

Band-limited noise w./wo. order reduction

Listeners had to equalize the perceived loudness of a reduced-order sound to a 4^{th}-order reference sound, both encoded in the look direction $\varphi_{\mathrm{s}} = 0°$ on the horizontal plane, and both being band-limited noise signals of the same frequency band. The set of Ambisonic pair comparisons is described in Table 1. Except for N = 0, which employed the same frequency range as N = 1, ranges were taken from [6] to fit the processing for a rigid-sphere array of $r = 4.2\,\mathrm{cm}$. Band limitation was achieved by 4^{th} order Butterworth band-pass filtering.

The continuous noise sounds of each comparison task was periodically switched from one to the other of the comparison pair every 800 ms. Listeners could influence the loudness of the reduced-order sound by moving a slider until they perceived a continuous sound without loudness modulation and then press the proceed button.

Table 1: Comparison pairs: bands and Ambisonic orders N.

f	A	Ref
68. . . 478 Hz	N = 0	N = 4
68. . . 478 Hz	N = 1	N = 4
478. . . 1405 Hz	N = 2	N = 4
1405. . . 2727 Hz	N = 3	N = 4

Table 2: Environmental/algorithmic conditions.

c_1	2D, 10 lspks, anechoic, center seat, ideal setup
c_2	2D, 12 lspks, CUBE, center seat, delay comp.
c_3	3D, 24 lspks, CUBE, center seat, delay comp.
c_4	3D, 24 lspks, CUBE, center seat
c_5	3D, 24 lspks, CUBE, off-center seat

Experimental Setup

The 10-channel loudspeaker arrangement in IEM's anechoic chamber uses ten 8020 Genelec loudspeakers at ear height, equally spaced at $r = 1.5\,\mathrm{m}$ with $\varphi = 0°(\text{front}), 36°, \ldots$ from the subject's perspective. IEM's CUBE is a permanent installation of 24 externally amplified coaxial Tannoy System 1200 with the loudspeaker angles documented, e.g., in [8]. We measured $RT = 675\,\mathrm{ms}$.

In total, we are interested in the influence of environments and conditions of Table 2: IEM's anechoic chamber (c_1) using 10 loudspeakers, IEM's CUBE, in which the loudspeakers do not strictly have equal delays to the center listening spot, with delay compensation in 2D (c_2: 12 horizontal loudspeakers), with and without delay compensation in 3D (c_3, c_4: 24 loudspeaker hemisphere) at the center, and 2.5 m, i.e. half-radius, left-off-center (c_5).

Ambisonic Panning/Decoding

Noise signals were presented through Ambisonic panning to $\varphi_s = 0$. The 2D max-rE-weighted [9] sampling Ambisonic panning of $c_{1\ldots2}$ used the weights $g_0, \ldots, g_{L-1}$

$$g_l = 1 + 2 \sum_{n=1}^{N} \cos(n\,\tfrac{\pi}{2(N+1)})\,\cos(n\,\varphi_l), \qquad (1)$$

to distribute the signal on $L = 10$ loudspeakers, $\varphi_l = \frac{2\pi}{L} l$, in c_1, and $L = 12$, with φ_l corresponding to the azimuth of the first 12 IEM CUBE loudspeakers, in c_2. For 3D Ambisonic panning in $c_{3\ldots5}$, max-rE-weighted AllRAP, cf. [7, 10], was employed.

Listeners and notes on the experiment

For testing the condition c_1, twelve listeners, average age of 26, took part and needed 13 minutes on average to finish 24 comparison tasks, of which we only used results of nine listeners and $4(\text{rep.}) \times 4(N) = 16$ responses, here. Excluded conditions refer to alternative decoding of the $N = 0, 1$ cases on fewer loudspeakers. The excluded 3 listeners gave repeated responses whose standard deviation reached 3 dB, three times as much as for the others.

In experiments concerning the conditions $c_{2\ldots5}$, ten listeners, average age 31, took part and required 35 minutes on average to finish $4(\text{rep.}) \times 4(\text{c.}) \times 4(N) = 64$ comparison tasks, whereof the last 16 were done after relocating to a 2.5 m left-shifted position c_5. Listeners gave their repeated answers with 0.6 dB average standard deviation.

All participants were experienced spatial audio listeners. To every listener, all comparison pairs for one listening position ($\{c_1\}$, $\{c_2, c_3, c_4\}$, $\{c_5\}$) were presented four times, each time in individual random order.

For the $N = 0$ and $N = 1$ comparison tasks and the central listening positions, listeners reported to perceive timbre differences, so that high- (or low-) frequency parts of the presented frequency bands were heard to be switched on and off. At the off-center listening position, a different amount of sound from the side was reported to cause slight difficulties in equalizing the loudness levels.

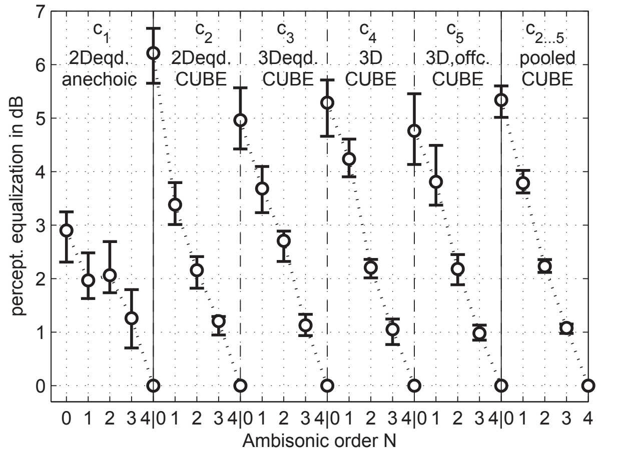

Figure 2: Equalization perceived in medians and confidence intervals making up for level differences in comparison pairs Tab. 1. Conditions Tab. 2 are 2D/3D, anechoic/IEM CUBE, with equal, equalized, unequalized delays, or off-center.

Table 3: Estimators of perceptually correct $c_{2\ldots5}$ equalization.

estimator	N=0	N=1	N=2	N=3
omni (2D/3D)	0 dB	0 dB	0 dB	0 dB
CLL	4.7 dB	3.7 dB	2.3 dB	1.3 dB
perceived	**5.3 dB**	**3.8 dB**	**2.3 dB**	**1.1 dB**
diffuse (2D)	6.4 dB	4.2 dB	2.4 dB	1.0 dB
diffuse (3D)	10.1 dB	6.1 dB	3.6 dB	1.6 dB
free (2D)	15.2 dB	8.6 dB	4.7 dB	2.0 dB
free (3D)	23.4 dB	13.5 dB	7.7 dB	3.4 dB

Results

The overview of the resulting perceptual equalization levels are depicted in Fig. 2. The level differences of the different-order bands are obvious, and there is a pronounced difference for the orders $N = 0, 1$ between anechoic c_1 and IEM CUBE $c_{2\ldots5}$. What is more, multivariate ANOVA for $c_{2\ldots5}$ revealed there being no significant influence of repetition, delay compensation, and listening position ($0.13 \le p \le 0.44$). By contrast, the 3D $c_{3\ldots5}$ and 2D c_2 playback conditions as well as the subjects have a significant influence ($p \le 0.03$).

The IEM CUBE conditions still resemble quite well, bearing in mind that 1 dB is often named as JND value for levels. Therefore Fig. 2 shows the pooled statistics thereof in its right-most column $c_{2\ldots5}$. Median perceived equalization levels amount to $\boldsymbol{\epsilon}_4$=[5.3, 3.8, 2.3, 1.1] dB.

Interestingly, the anechoic center-seat condition c_1 seems to require much less equalization. We expect this result to apply to binaural rendering using anechoic HRIRs.

Models

Analytic equalization

From all analytical equalizations (diffuse-field, free-field, and omni-directional for max-rE 2D/3D, cf. Fig. 1), an analytic 2D diffuse equalizer comes closest to the experimental medians. For $\max N = 4$, it is calculated by

$$\epsilon_{\mathrm{N},4} = \sqrt{\frac{1+2\sum_{n=1}^{\mathrm{N}}\cos^2\left(n\,\frac{\pi}{2(\mathrm{N}+1)}\right)}{1+2\sum_{n=1}^{4}\cos^2\left(n\,\frac{\pi}{10}\right)}}. \tag{2}$$

In the anechoic, centered 2D condition c_1, the diffuse-field equalization matches quite well above 1 kHz ($\mathrm{N} \geq 2$), which could be explained by a predominantly stochastic signal interference at both ears. Below 1 kHz, equalization levels between diffuse-field and omni-directional fit well.

For the echoic IEM CUBE conditions, the 2D diffuse-field equalization slightly over-estimates the experiments but matches in higher bands. By contrast, 3D equalizers or such for free-field highly over-estimate the required equalization by several dBs, see Tab. 3.

Measurement-based equalization (CLLs)

The composite loudness levels (CLLs) [11, 12] allow technical comparisons of the loudness recorded by a dummy head. CLLs combine levels extracted at both ears to one measure. The ratio of the 4^{th}-order to the reduced-order CLL, $\mathrm{CLL}_{\mathrm{Ref}}/\mathrm{CLL}_{\mathrm{A}}$, estimates a third-octave equalizer.

For the anechoic condition c_1, we could use 10 HRIRs out of the Neumann KU100 measurements by ARI[1]. Moreover, we took BRIRs with Bruel&Kjær HATS 4128C of the IEM CUBE loudspeakers for the conditions $c_{2\ldots5}$.

Before the CLL estimation, BRIRs/HRIRs of the loudspeaker setup were superimposed using the corresponding gains for panning of the orders $\mathrm{N} = 0, \cdots, 4$.

The third-octave $\mathrm{CLL}_{\mathrm{Ref}}(f)/\mathrm{CLL}_{\mathrm{A}}(f)$-ratios falling into the different frequency ranges of Tab. 1 allow to analyze median and IQR of the CLL equalization estimator, see Fig. 3. Its estimation of all responses of the IEM CUBE conditions and their statistical spread is quite good. Estimation for the anechoic condition is comparatively poor.

The CLL tend to under-estimate the required equalization.

[1] http://sofacoustics.org/data/database/ari(artificial)

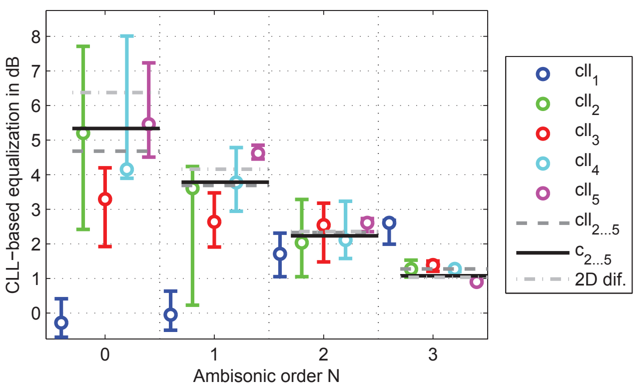

Figure 3: Median and IQR of all binaural third-octave CLLs within the bands specified in Tab. 1, compared to the median experimental results.

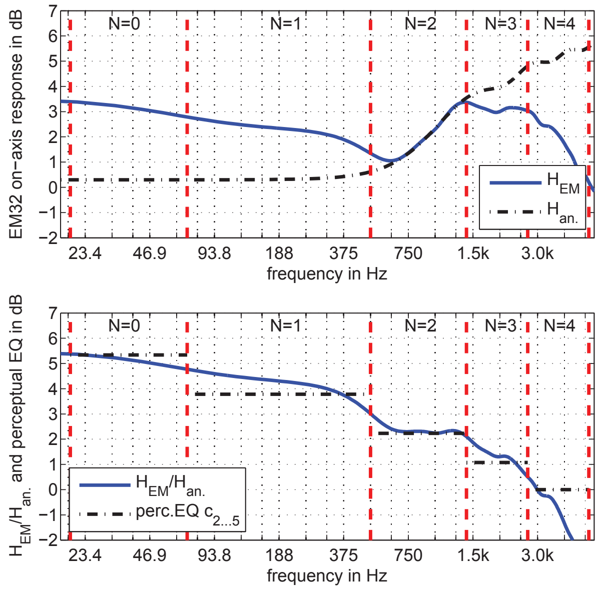

Figure 4: Measured Eigenmike EM32 on-axis responses compared to analytical (top) and their difference (+2.3 dB offset) compared to perceived equalization (bottom).

Discussion of low-frequency results

In the delay-compensated 2D CUBE condition c_2 for $\mathrm{N} = 0$, listeners {1,2,3,6,8,10} used a 5.3 dB average equalization, while listeners {4,5,7,9} used 7.3 dB, what we found a remarkably large difference. Reviewing the CLLs of the comparison pair reveals an ambiguous choice dividing the listener groups: In c_2, the relative $\mathrm{CLL}_{\mathrm{Ref}}/\mathrm{CLL}_{\mathrm{A}}$ yields 1.7 dB in the 80 Hz octave and 6.1 dB in the 200 Hz octave. Listeners {4,5,7,9} apparently focused on equalizing the 200 Hz octave, while accepting more fluctuation in the 80 Hz octave than the listeners {1,2,3,6,8,10}. Obviously, an improved low-frequency equalization would require sub-division of the $\mathrm{N} = 0$ band.

Discussion of Eigenmike equalization

Despite the results, omni-directional equalization sounded most natural for the playback of Eigenmike EM32 (ser.nr. 27) recordings at IEM CUBE and the IEM mAmbA[2] with our own algorithms [6]. As reasonable explanation, pressure pickup and signal conditioning might already, as a side benefit, equalize for a perceptually flat response.

To gain insight, we reviewed results from EM32 directivity measurements[3] taken in 2013 in order to extract the average on-axis response $H_{\mathrm{EM}}(f)$, as shown in Fig. 4 (top). Analytically, the on-axis response $H_{\mathrm{an.}}(f)$ of the pressure sensed on a rigid sphere increases by up to +6dB at high frequencies (dash-dotted curve). After dividing by this increase, $H_{\mathrm{EM}}/H_{\mathrm{an.}}$, the bottom diagram in Fig. 4 shows that the EM32 already provides perceptual equalization for surround playback below the spatial aliasing frequency.

[2] http://iem.kug.ac.at/darmstadt2014/international-summer-course-for-new-music-darmstadt-2014/mamba.html

[3] http://iaem.at/kurse/sommer-13/ahlu/2013_protokoll_eigenmike_bucheggerhackkeller.pdf

The reason for the particular shape of $H_{\mathrm{EM}}/H_{\mathrm{an.}}$ is unclear, as it deviates from an expected raw-transducer frequency response. Nevertheless, an independently measured frequency response of another EM32 (by Matthias Kronlachner) seemed to confirm the shape.

Conclusions

We could present and discuss the necessity and amount of equalization required for surround playback of direct sound recorded by compact spherical microphone arrays.

Array signal processing is forced to limit the resolution (Ambisonic order) towards low frequencies, which, technically, causes an attenuation of the diffuse-field sensitivity. We wanted to know whether this causes perceivable bass attenuation. To isolate the effect, our listening experiments compared the loudness of Ambisonic 4^{th}-order panning with the loudness of reduced-order panning within typical frequency bands, for different loudspeaker setups.

Under anechoic conditions and for a centered listener, e.g. binaural rendering, our results indicate that diffuse-field equalization is suitable for frequencies above 1 kHz, but not below, where levels between no (omni-directional) and diffuse-field equalization were preferred.

Under studio and sound reinforcement conditions (IEM CUBE), the 2D diffuse-field equalization appears to match best, also for 3D. Interestingly, the particular choice of rendering (2D/3D, w/wo. delay compensation) was only of little influence.

Moreover, the perceptual equalization can be explained by relative composite loudness levels. Comments of the listeners and CLLs indicate that low frequencies might require a more frequency-selective equalization. All the more, low frequencies should be equalized for the particular playback room.

Acknowledgments

We thank Stefan Lösler for his recent work on spherical array filter design, the students of the Acoustic Holography and Holophony lab course 2013 measuring the characteristics of the Eigenmike EM32 on a grid of 480 directions, and Matthias Kronlachner for an independent on-axis measurement.

This work was supported by the project ASD, which is funded by Austrian ministries BMVIT, BMWFJ, the Styrian Business Promotion Agency (SFG), and the departments 3 and 14 of the Styrian Government. The Austrian Research Promotion Agency (FFG) conducted the funding under the Competence Centers for Excellent Technologies (COMET, K-Project), a program of the above-mentioned institutions.

References

[1] S. Moreau, “Etude et realisation d'outils avances d'encodage spatial pour la technique de spatialisation sonore higher order Ambisonics : microphone 3D et controle de distance,” PhD Thesis, Academie de Nantes, Le Mans, France, 2006.

[2] W. Song, W. Ellermeier, and J. Hald, “Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise,” *JASA*, 123, 2008.

[3] B. Bernschütz, C. Pörschmann, S. Spors, and S. Weinzierl, “Soft-limiting der modalen Amplitudenverstärkung bei sphärischen Mikrofonarrays im Plane-Wave-Decomposition-Verfahren,” *Fortschritte der Akustik, DAGA*, 2011.

[4] W. Song, W. Ellermeier, and J. Hald, “Psychoacoustic evaluation of multichannel reproduced sounds using binaural synthesis and spherical beamforming,” *JASA*, 130, 2011.

[5] R. Baumgartner, H. Pomberger, and M. Frank, “Practical implementation of radial filters for Ambisonic recordings,” *ISCA*, 2011.

[6] S. Lösler and F. Zotter, “Comprehensive radial filter design for practical higher-order Ambisonic recording,” *Fortschritte der Akustik, DAGA*, 2015.

[7] F. Zotter and M. Frank, “All-round Ambisonic panning and decoding,” *J. Audio Eng. Soc*, vol. 60, no. 10, 2012.

[8] F. Zotter, H. Pomberger, and M. Noisternig, “Energy-preserving Ambisonic decoding,” *Acta Acustica u. A.*, vol. 98, no. 1, 2012.

[9] J. Daniel, J.-B. Rault, and J.-D. Polack, “Ambisonics encoding of other audio formats for multiple listening conditions,” in *Audio Engineering Society Convention 105*, Sept. 1998.

[10] A. J. Heller and E. M. Benjamin, “The Ambisonic Decoder Toolbox: extensions for partial-coverage loudspeaker arrays,” *Linux Audio Conference*, 2014.

[11] K. Ono, V. Pulkki, and M. Karjalainen, “Binaural modeling of multiple sound source perception: Methodology and coloration experiments,” in *Audio Engineering Society Convention 111*, Nov. 2001.

[12] K. Ono, V. Pulkki, and M. Karjalainen, “Binaural modeling of multiple sound source perception: Coloration of wideband sound,” in *Audio Engineering Society Convention 112*, Apr. 2002.