How small can a baffled ambisonic microphone array be?

How Small Can a Baffled Ambisonic Microphone Array Be?

Jens Ahrens

Chalmers University of Technology, 412 96 Gothenburg, Sweden, E-Mail: jens.ahrens@chalmers.se

Introduction

A variety of microphone arrays have been proposed for capturing ambisonic signals. Arrays that provide moderate or high orders typically use a baffle onto which the microphones are mounted. This includes classical spherical microphone arrays (SMAs) [1], equatorial microphone arrays (EMAs) [2], and arrays with non-spherical baffles (XMAs) [3]. The physical mechanisms are identical for the different arrays. They therefore all exhibit the same limitations:

1) At high frequencies, spatial aliasing and order truncation reduce the accuracy of the spherical harmonic (SH) coefficients of the captured sound field that the array provides. The frequency above which this occurs can be estimated with the $N = kR$-rule where N is the maximum order that the array provides, $k = 2\pi f_\mathrm{c}/c$ is the wavenumber that is considered, R is the radius of the baffle, and c is the speed of sound [4]. Solving the expression for the critical frequency f_c yields

$$f_\mathrm{c} = \frac{N}{2\pi R} c\,. \tag{1}$$

This demonstrates that we want the radius R of the baffle to be as small as possible so that f_c is as high as possible for a given N. N is determined by the number of microphones and their placement.

2) At low frequencies, massive amplifications of the microphone array signals are required at non-low orders for being able to extract the SH coefficients. This is problematic in practise as sensor mismatch and sensor self noise can impair the result. Also, we found for the application of binaural rendering that measurement inaccuracies in the employed head-related transfer functions (HRTFs) can impair the result. This is why it is common to limit the dynamic range of these so-called radial filters. A range of 40 dB as depicted in Fig. 1 is typical. The consequence of this regularization of the radial filter gain is that the given order will be attenuated in the output of the microphone array in the frequency range in which the regularization is effective. Accordingly, the given order will be attenuated in the binaural signals if binaural rendering of the SH representation of the sound field is performed.

Note that we will exclusively consider binaural rendering of the SH coefficients in this paper as well as linear encoding (i.e. linear methods for obtaining an SH representation of the captured sound field). We comment on loudspeaker array-based reproduction in the Conclusions. Formulating the requirements for parametric encoding like in [6] requires a different approach. Previous works like [7] touched upon the effect of the radial filter regularization on binaurally rendered signals, but we are not aware of a comprehensive analysis. In this paper, we formulate requirements on the combination of the maximum order N of the array and the radius R of the baffle for avoiding impairment of the binaural signals.

[1] `https://github.com/AppliedAcousticsChalmers/ambisonic-encoding`

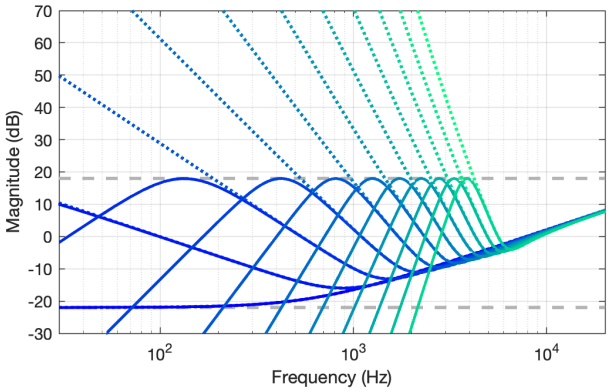

Figure 1: Magnitude transfer functions of the radial filters for orders $n = 0$ (blue) to $n = 10$ (green). The dotted lines denote the theoretic filters, and the solid lines denote the filters when regularized such that they exhibit a maximum dynamic range of 40 dB (from -22 dB to 18 dB, as indicated by the gray dashed lines) achieved through Tikhonov regularization [5, Eq. (32)] as implemented in[1].

The Ear Signals

The binaural signals $B(\omega)$ can be computed from the microphone array output via [8, 9]

$$B(\omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} \breve{S}_{n,m}(\omega)\; \mathring{H}_{n,m}(\omega)\,. \tag{2}$$

Eq. (2) needs to be formulated for each ear separately. The sound field coefficients $\breve{S}_{n,m}(\omega)$ are the same in either formulation. $\mathring{H}_{n,m}(\omega)$ are the SH coefficients of the HRTFs of the corresponding ear for with (2) is formulated. Natural sound fields tend to have an order of $N = \infty$. Microphone arrays are only capable of extracting $\breve{S}_{n,m}(\omega)$ up to a finite N.

Eq. (2) is difficult to investigate based on measured data as $\breve{S}_{n,m}(\omega)$ generally diverges at low frequencies and $\mathring{H}_{n,m}(\omega)$ tends to zero. We therefore use simple models for both the sound field and the HRTFs so that we can simulate $\breve{S}_{n,m}(\omega)$ and $\mathring{H}_{n,m}(\omega)$ with arbitrary accuracy based on analytical expressions. Specifically, we assume that the sound field that is being rendered is a spherical

wave that originates from a distance r_s straight ahead in the horizontal plane and that the HRTFs are those of a rigid spherical head with radius of $R_\mathrm{head} = 0.075\,\mathrm{m}$ and whose ear canal entrances are located at $\pm 90°$.

Inserting the corresponding analytical expressions into (2) yields

$$B_\mathrm{sph.}(\omega) = \sum_{n=0}^{N} \sum_{m=-n}^{n} (-\mathrm{i}) \frac{\omega}{c}\, h_n(\omega r_\mathrm{s}/c)\; Y_{n,m}(\beta_\mathrm{s}, \alpha_\mathrm{s}) \\ b_n(\omega R_\mathrm{head}/c)\; Y_{n,m}(\pi/2, \alpha_\mathrm{ears})\,. \quad (3)$$

for the resulting ear signals $B_\mathrm{sph.}(\omega)$ of the spherical head. $b_n(\omega R_\mathrm{head}/c)$ are the radial terms for a rigid sphere with radius R_head [1]. Note that (3) simply represents the sound pressure on the surface of a rigid sphere with radius R_head at the two ear positions due to our spherical wave. This demonstrates that (2) does not produce any systematic error. An error in the binaural signal $B(\omega)$ occurs only if the sound field coefficients $\breve{S}_{n,m}(\omega)$ are not extracted accurately by the given microphone array.

Evaluation

The Effect of the Order Truncation

We investigate first the consequences of the unavoidable order truncation on the binaural signals. This is required for being able to identify the effect of the attenuation of $\breve{S}_{n,m}(\omega)$ due to the radial filter regularization. Order truncation means that the coefficients $\breve{S}_{n,m}(\omega)$ are only available for $n \leq N$. We will analyse this aspect assuming that ideal radial filters are in place so that no unwanted attenuation whatsoever occurs.

Fig. 2 depicts the ground truth (infinite order) ear signal (red line) for a sound source at a distance of 3 m and the listener with the rigid spherical head facing the source. The blue and green lines represent the ear signals for the cases where only the SH coefficients up to the indicated order N were used to evaluate (3). The deviation from the red line represents the truncation error. Interestingly, it is evident from Fig. 2 that different orders n contribute to the ear signal primarily in a very confined frequency range. For example, the ear signal for $N = 2$ does not deviate from the ground truth below approx. 1.2 kHz, which means that all orders $n > 2$ do not contribute to the ears signals below 1.2 kHz and so forth. Fig. 3 depicts the same data for the listener facing to the side.

The observation of high orders not contributing to the binaural signals at low frequencies can be deduced from the $N = kR$ rule that is represented by (1): A given maximum order N yields a critical frequency f_c. Consequently, a maximum order of $N+1$ yields an accordingly higher f_c'. This means that the order $n = N + 1$ contributes to the binaural signals primarily in the frequency range $f_\mathrm{c} < f < f_\mathrm{c}'$.

The same matter can be illustrated from a slightly different angle through Fig. 4, which depicts the energy of the sound coefficients $\breve{S}_{n,m}(\omega)$ and of the HRTF coefficients $\mathring{H}_{n,m}(\omega)$ for each order n (i.e. summed up over all m for a given n). The HRTFs exhibit very low energy at high

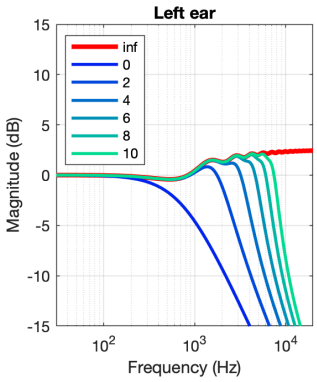

Figure 2: Ear signals for different maximum orders N for a listener facing the sound source assuming ideal radial filters

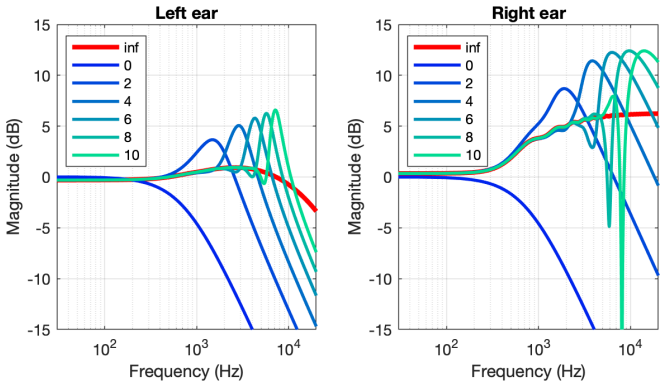

Figure 3: Ear signals for different maximum orders N for sound incidence from the right assuming ideal radial filters

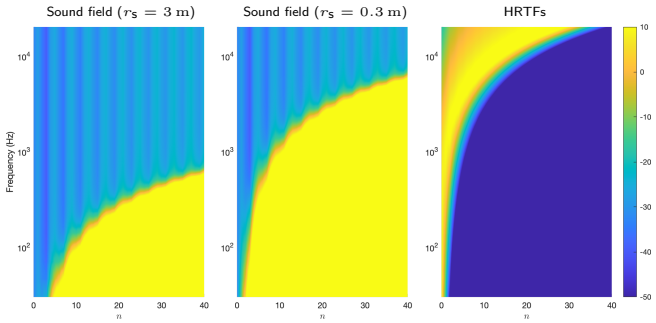

Figure 4: Energy of the HRTFs (right) and of the sound field (left and middle) per order n

orders and low frequencies so that, according to (2), the sound field coefficients experience a strong attenuation int he computation of the ear signals.

It is evident from Fig. 2 and 3 that the order truncation causes an attenuation of high frequencies. For lateral sound incidence, a considerable amplification occurs at frequencies just below the roll off. Practical implementations of microphone arrays use finite sets of discrete microphones, which causes spatial aliasing that compensates partly for the attenuation at high frequencies [7]. Further improvements in terms of the spectral balance of the binaural signals are achieved by employing MagLS equalized HRTFs [10] so that the binaural signals obtained from microphone arrays of a moderately high order of, say, $N = 5$ can provide high quality.

It is great news that each order n contributes to the binaural signals only in a confined frequency range because the regularization of the radial filters attenuates the extracted sound field coefficients $\check{S}_{n,m}(\omega)$ at high orders at low frequencies. Fig. 2 and 3 demonstrated that the high-order information is indeed dispensable at low frequencies.

The smaller the radius R of the microphone array baffle is the higher is the frequency up to which a given order n is attenuated by the regularization. This suggests that for each desired maximum order N, there is a minimum radius R that the baffle needs to have so that the radial filter regularization does not affect the binaural signals. The remainder of this paper is dedicated to establishing this minimum radius R as a function of the desired maximum order N.

The Effect of the Radial Filter Regularization

Fig. 5, 6, and 7 depict ear signals computed via (3) for an order of $N = 7$ and sound incidence from the right. The red line represents the ground truth (infinite order) ear signal, the blue line represents the Nth-order ear signal with ideal radial filters (i.e. with no regularization of the radial filters), and the green line represents the Nth-order ear signals obtained with radial filters that are regularized to a maximum dynamic range of 40 dB as depicted in Fig. 1. The vertical gray lines in Fig. 5-7 denote f_c according to the $N = kR$ limit given by (1).

Any deviation of the green lines from the blue ones is a consequence of the radial filter regularization. The deviations tend to be largest for close sound sources and lateral sound incidence. We therefore show results only for a source distance of 0.3 m and sound incidence from 90° to the right. The deviations for far sources and the listener facing the sound source are only marginally smaller.

Only small deviations occur for $R = 0.08\,\mathrm{m}$ at frequencies below 100 Hz (green vs. blue lines in Fig. 5). These are very likely to be uncritical. They can actually be compensated for by using eMagLS binaural rendering [11]. Reducing R to 0.06 m produces also deviations at higher frequencies, which are likely to also be uncritical (Fig. 6). Reducing R further to 0.04 m as depicted in Fig. 7 produces deviations of several dB in the mid-frequency range, which are likely to be audible and therefore have to be considered an impairment. In how far these deviations can be compensated for by eMagLS rendering is not clear at this stage as eMagLS (and MagLS for that matter) can extend the duration of the energy support of the corresponding impulse responses slightly, and it is difficult to anticipate if a possible reduction of the deviations of magnitude spectra of the binaural signal dominates over the impaired time response. An equalization according to [3, App.] is also conceivable but has the same potential downsides. The smallest baffle radius R that supports the considered order of $N = 7$ is therefore approx. 0.06 m.

Note that the deviations in Fig. 7, for example, can be avoided by increasing the dynamic range of the radial filters from 40 dB to 60 dB, which is feasible but is certainly

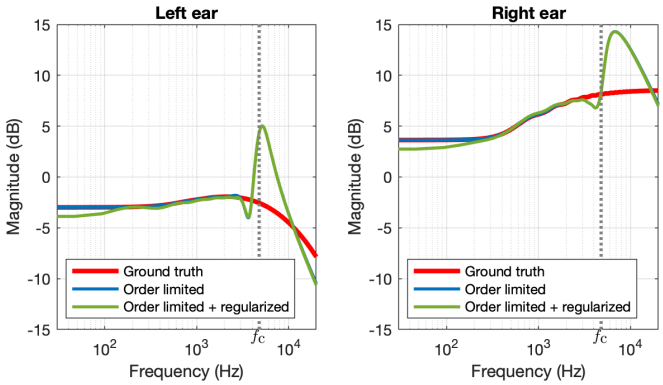

Figure 5: Binaural signals, $R = 0.08\,\mathrm{m}$, $N = 7$

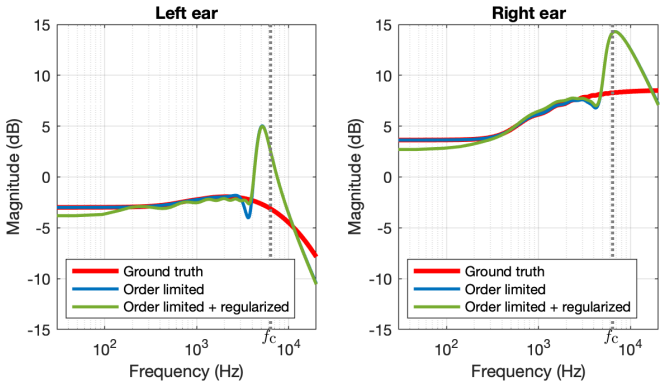

Figure 6: Binaural signals, $R = 0.06\,\mathrm{m}$, $N = 7$

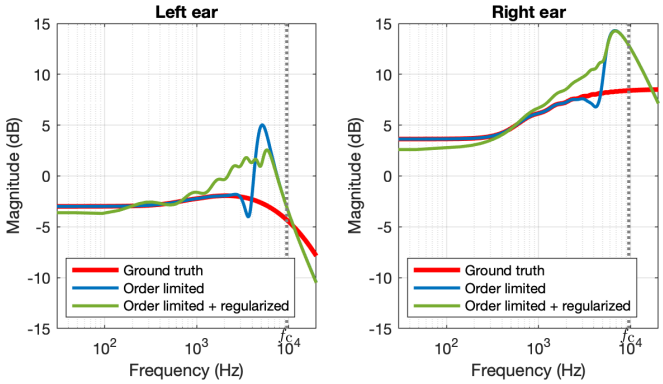

Figure 7: Binaural signals, $R = 0.04\,\mathrm{m}$, $N = 7$

on the more extreme side. It can entail a variety problems in practise due to microphone mismatch and self noise or inaccuracies in the HRTFs at low frequencies. Note that other authors prefer a lower dynamic range of approx. 30 dB [12].

We conducted the same exercise for orders $N = 3\,..\,12$ to determine the minimum radius R that is required for binaural rendering. The results are summarized in Tab. 1, where also the corresponding critical frequency f_c according to (1) is provided. If the employed baffle has the stated radius R, then the binaural output signals of the processing pipeline will be accurate up to f_c. Note that this statement is only valid for the rigid sphere HRTFs that we employed here. Limiting the order of a human HRTF set to, say, $N = 3$ may impair monaural cues even if MagLS HRTFs are used.

N	3	4	5	6	7	8	9	10	11	12
$R_{30\text{dB}}$	0.05 m*	0.06 m	0.06 m	0.07 m	0.07 m	0.07 m	0.07 m	0.07 m	0.08 m	0.08 m
$f_{\text{c},30\text{dB}}$	3.3 kHz	3.6 kHz	4.5 kHz	4.7 kHz	5.4 kHz	6.2 kHz	7.0 kHz	7.8 kHz	7.5 kHz	8.2 kHz
$R_{40\text{dB}}$	0.03 m	0.04 m	0.05 m	0.06 m	0.06 m	0.06 m	0.07 m	0.07 m	0.07 m	0.07 m
$f_{\text{c},40\text{dB}}$	5.5 kHz	5.5 kHz	5.5 kHz	5.5 kHz	6.4 kHz	7.3 kHz	7.0 kHz	7.8 kHz	8.6 kHz	9.4 kHz
$R_{60\text{dB}}$	0.015 m	0.02 m	0.03 m	0.035 m	0.04 m	0.04 m	0.045 m	0.05 m	0.05 m	0.05 m
$f_{\text{c},60\text{dB}}$	10 kHz	10 kHz	9.1 kHz	9.4 kHz	9.6 kHz	10 kHz	10 kHz	10 kHz	12 kHz	13 kHz

Table 1: Minimum baffle radius R and resulting critical frequency f_c according to (1) for different desired maximum array orders N assuming a dynamic range of the radial filters of 30 dB (top), 40 dB (middle), and 60 dB (bottom) as well as rigid-sphere HRTFs. The asterisk denotes a parameter set with larger deviations than the others.

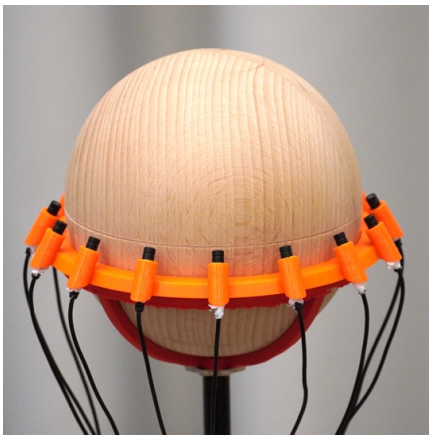

Figure 8: 15-channel 7th-order EMA with $R = 6$ cm. The design is available at[2].

Conclusions

Tab. 1 demonstrates that especially for the 40-dB-dynamic range of the radial filters, there is indeed a benefit in designing a microphone array with higher order N given that the critical frequency f_c above which spatial aliasing and order truncation are apparent increases with the chosen order N if the baffle radius R is selected to be the minimum that is required. Given that the physical mechanisms underlying SMAs, EMAs, and XMAs, are the same, our results hold for all these microphone array types if linear encoding is performed.

SMAs with orders $N > 7$ are difficult to design in practise as the required number Q of microphones becomes excessive ($Q \geq (N+1)^2$) especially given the comparably small required radii R. EMAs require only $Q = 2N + 1$ microphones, which appears feasible even for high orders N as the microphones can be mounted outside of the baffle. An example EMA implementation of an array with optimal radius is depicted in Fig. 8. More designs are available at[2].

We remind the reader that our results hold only if binaural rendering is performed. It is likely that the situation is similar for loudspeaker array-based reproduction. It can be deduced from the $N = kr$-rule that, at a given frequency, lower SH orders n contribute primarily to the sound field close to the center of expansion, and high orders contribute to the sound field farther away. As the distance is directly proportional to the wavelength, low orders can be sufficient to cover the entire listening area at low frequencies.

[2] `https://github.com/AppliedAcousticsChalmers/ambisonic-microphone-arrays`

References

[1] Boaz Rafaely. Analysis and design of spherical microphone arrays. *IEEE Trans. on Speech and Audio Proc.*, 13(1):135–143, 2005.

[2] J. Ahrens, H. Helmholz, D. Alon, and S. V. Amengual Gari. Spherical Harmonic Decomposition of a Sound Field Based on Observations Along the Equator of a Rigid Spherical Scatterer. *J. Acoust. Soc. Am.*, 150(2), 2021.

[3] J. Ahrens, H. Helmholz, D. Alon, and S. V. Amengual Gari. Spherical Harmonic Decomposition of a Sound Field Using Microphones on a Circumferential Contour Around a Non-Spherical Baffle. *ACM/IEEE Trans. on Audio, Speech, and Lang. Proc.*, 30:3110–3119, 2022.

[4] Rodney A. Kennedy, Parastoo Sadeghi, Thushara D. Abhayapala, and Haley M. Jones. Intrinsic limits of dimensionality and richness in random multipath fields. *IEEE Transactions on Sig. Proc.*, 55(6):2542–2556, 2007.

[5] Sébastien Moreau, Jérôme Daniel, and Stéphanie Bertet. 3D sound field recording with higher order ambisonics – objective measurements and validation of a 4th order spherical microphone. In *120th Conv. of the AES*, May 2006.

[6] Archontis Politis, Sakari Tervo, and Ville Pulkki. COMPASS: Coding and Multidirectional Parameterization of Ambisonic Sound Scenes. In *IEEE ICASSP*, 2018.

[7] Benjamin Bernschütz. Microphone arrays and sound field decomposition for dynamic binaural recording. PhD thesis, Technische Universität Berlin, 2016.

[8] Boaz Rafaely and Amir Avni. Interaural cross correlation in a sound field represented by spherical harmonics. *J. Acoust. Soc. Am.*, 127(2):823–828, 2010.

[9] Jens Ahrens. Binaural Audio Rendering in the Spherical Harmonic Domain: A Summary of the Mathematics and Its Pitfalls. Technical note v. 2, Chalmers University of Technology, 2022.

[10] Christian Schörkhuber, Markus Zaunschirm, and Robert Höldrich. Binaural rendering of ambisonic signals via magnitude least squares. In *Fortschritte der Akustik (DAGA)*, March 2018.

[11] Thomas Deppisch, Hannes Helmholz, and Jens Ahrens. End-to-end magnitude least squares binaural rendering of spherical microphone array signals. In *Immersive and 3D Audio: From Architecture to Automotive, I3DA*, 2021.

[12] Franz Zotter. A linear-phase filter-bank approach to process rigid spherical microphone array recordings. In *IcETRAN*, 2018.