LOCALIZATION ACCURACY IN PRESENTING MEASURED SOUND FIELDS VIA HIGHER ORDER AMBISONICS

Samuel Clapp, Jonas Braasch, and Ning Xiang

Graduate Program in Architectural Acoustics
Rensselaer Polytechnic Institute
Greene Building, 110 8th Street
Troy, NY 12180, USA
clapps@rpi.edu

Anne Guthrie

Arup Acoustics
77 Water Street
New York, NY 10005, USA
anne.guthrie@arup.com

ABSTRACT

A spherical microphone array can encode a measured sound field into its spherical harmonic components. Such an array will be subject to limitations on the highest spherical harmonic order it can encode and encoding accuracy at different frequencies. Ambisonics is a system designed to reproduce the spherical harmonic components of a measured or virtual sound field using multiple loudspeakers. In ambisonic systems, the size of the sweet spot is wavelength dependent, and thus decreases in size with an increase in frequency. This paper examines how to reconcile the limitations of the recording and playback stages to arrive at the optimum ambisonic decoding scheme for a given spherical array design. In addition, binaural models are used to evaluate these systems perceptually.

1. INTRODUCTION

Spherical microphone arrays have been studied extensively for beamforming [1, 2], source localization, and other applications. Likewise, much theoretical and practical work has been done on the optimum methods for ambisonic decoding [3, 4, 5]. Both systems use the concept of spherical harmonics: spherical microphone arrays can decompose a sound field into its spherical harmonic components, while ambisonic decoding can reconstruct the spherical harmonic components of a sound field for a listener.

The performance of spherical microphone arrays is frequency-dependent, affected primarily by the array's size and number of sensors. In many applications, such as beamforming and direction-of-arrival (DOA) estimation, a narrow frequency band is used, where the array's performance is optimum. However, restricting oneself to a narrow frequency band is untenable when presenting auditory scenes to listeners. Thus, we require a way to utilize information from outside of the optimum band, where higher spherical harmonic orders might not be accurately decomposed, but lower orders are.

Much of the literature on ambisonics deals with the reproduction of simulated sound fields, where the exact DOA of every sound event is known, and the restrictions on the highest spherical harmonic order come only from the number of channels in the loudspeaker array.

This paper examines a method for determining mixed-order ambisonic decoding schemes determined by the constraints of the two systems, as well as a way to evaluate the perceptual accuracy of these schemes using binaural models.

2. SPHERICAL HARMONICS

The homogeneous wave equation is given in its general form by:

$$\nabla^2 p = \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}. \tag{1}$$

If expressed in spherical coordinates, solutions can be obtained through a separation of variables, yielding sets of functions for the radial, azimuthal, and elevation components [6]. Solutions to the azimuthal component are given by either sine and cosine terms or complex exponentials, while solutions to the elevation component are given by the associated Legendre functions ($P_n^m(x)$) in the cosine of the elevation angle. Combining these two components (and adding a normalization term) yields the (complex-valued) expression for spherical harmonics:

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

(The real-valued expression uses sine and cosine terms for the azimuthal component.)

The spherical harmonics form an orthonormal basis on the sphere:

$$\int_0^{2\pi}\int_0^{\pi} Y_{n'}^{m'}(\theta,\phi)^* Y_n^m(\theta,\phi)\, \sin\theta\, d\theta\, d\phi = \delta_{nn'}\delta_{mm'}. \tag{3}$$

3. SPHERICAL MICROPHONE ARRAY PROCESSING

3.1. Spherical Harmonic Decomposition

When a plane wave impinges upon a rigid sphere, the sphere will radiate a spherical wave whose intensity will vary as a function of the incident wave's angle of incidence and wavelength (in relation to the radius of the sphere). It is shown in [6] and [7] that by solving the wave equation with the appropriate boundary conditions, we arrive at an expression for the pressure at a point on

a rigid sphere of radius a (denoted by its angular position (θ, ϕ)) due to a plane wave incident from (θ_i, ϕ_i) with amplitude P_0 and wavenumber $k = 2\pi f/c$:

$$p(\theta, \phi, ka) = 4\pi P_0 \sum_{n=0}^{\infty} i^n b_n(ka) \sum_{m=-n}^{n} Y_n^m(\theta, \phi) Y_n^m(\theta_i, \phi_i)^*, \tag{4}$$

with

$$b_n(ka) = j_n(ka) - \frac{j_n'(ka)}{h_n^{(1)'}(ka)} h_n^{(1)}(ka), \tag{5}$$

a quantity referred to as the modal amplitude, which is shown for several orders in Fig. **??**. The slope of each curve is 3 dB per octave multiplied by the order.

Now, if we apply a weighting factor W to each point on the sphere:

$$W_{n'}^{m'}(\theta, \phi, ka) = \frac{Y_{n'}^{m'}(\theta, \phi)^*}{4\pi i^{n'} b_{n'}(ka)}, \tag{6}$$

and then integrate over the entire sphere, we can use the orthonormality of the spherical harmonics (Eq. 3) to yield the following result:

$$\int_0^{2\pi} \int_0^{\pi} W_{n'}^{m'}(\theta, \phi, ka)\, p(\theta, \phi, ka)\, \sin\theta \, d\theta \, d\phi = P_0 Y_{n'}^{m'}(\theta_i, \phi_i)^*. \tag{7}$$

Thus, we can determine the spherical harmonic components of the incident plane wave. However, evaluating the integral in Eq. 7 requires a continuous spherical transducer, and in practice, we must sample the pressure at Q discrete points on the sphere (denoted by their angular positions (θ_q, ϕ_q)), leading to the following summation:

$$\sum_{q=1}^{Q} W_{n'}^{m'}(\theta_q, \phi_q, ka)\, p(\theta_q, \phi_q, ka)\, C_n^m(\theta_q, \phi_q) \approx P_0 Y_{n'}^{m'}(\theta_i, \phi_i)^*, \tag{8}$$

where C_n^m are the quadrature coefficients. Using a nearly uniform sampling scheme, we can estimate the spherical harmonic components up to order N such that $Q \geq (N+1)^2$.

3.2. Error Sensitivity

Spherical microphone arrays lend themselves well to beamforming, a set of techniques for multi-channel sensors that allow for spatial filtering [2, 8, 9, 10]. These techniques involve solving for a set of weighting factors that are applied to each channel on the array, allowing the array to "look" in a particular direction at a particular frequency. The robustness of the beamformer is given by a quantity known as the white noise gain (WNG):

$$\mathrm{WNG}(\theta_0, \phi_0, \theta_q, \phi_q, ka) = 10 \log_{10} \left(\frac{|\mathbf{d}^T \mathbf{W}|^2}{\mathbf{W}^H \mathbf{W}} \right), \tag{9}$$

where $\mathbf{d}^T$ is a column vector of the microphone pressure values due to a plane wave impinging on the sphere from some direction (θ_i, ϕ_i), and $\mathbf{W}$ the sensor weights to look in that direction. WNG represents the array's sensitivity to noise and microphone positioning errors, with negative values representing an amplification and positive values representing an attenuation of (spatially uncorrelated) white noise. A spherical array is most sensitive to these types of errors at lower frequencies, where higher order spherical harmonic components are low in level and must be amplified considerably. The lefthand portion of Fig. 1 shows the frequencies at which WNG = -20 dB for a 64-channel array of varying radius. This value can be used as a guideline to determine the lowest frequency at which a certain order of spherical harmonic components can be accurately measured by a given array.

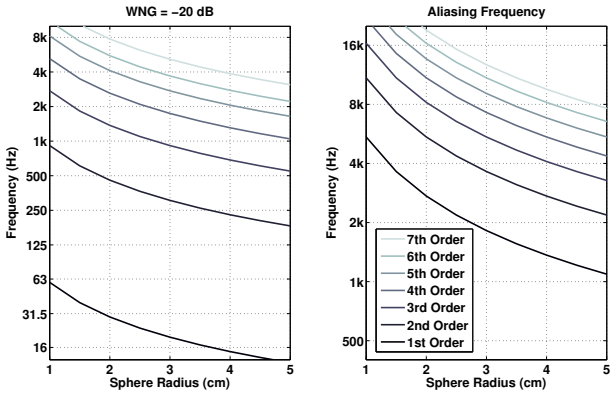

Figure 1: Left side: frequency at which WNG = -20 dB as a function of sphere radius for a 64-channel array at different spherical harmonic orders. Right side: frequency at which aliasing errors occur as a function of sphere radius for different spherical harmonic orders.

3.3. Aliasing

At higher frequencies, the aliasing of higher order spherical harmonic components into lower orders becomes an issue. We can see from Eq. 4 that a plane wave is not spherical harmonic order-limited and from Eq. 5 that higher-order components become more prominent at higher frequencies. Thus, aliasing errors affect the array for frequencies such that $ka > N$, where a is the radius of the sphere and N is the highest spherical harmonic order that can be measured by the array [11]. This frequency is shown in the right-hand portion of Fig. 1 for a range of sphere radii and spherical harmonic orders, with arrays of smaller radii and higher order (i.e. with a greater number of channels) offering higher thresholds for aliasing.

4. AMBISONICS

4.1. Basic Decoding

Ambisonics is a system that uses multiple loudspeakers to synthesize a sound field, where the gains of the loudspeakers are determined based on the spherical harmonic expansion of the sound field [12, 3, 4, 5].

Let us start with a plane wave of wavenumber $\mathbf{k_i}$ and incident from the direction (θ_i, ϕ_i), defined by its spherical harmonic coef-

ficients as:

$$p = e^{i\mathbf{k_i r}} = 4\pi \sum_{n=0}^{\infty} i^n j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta,\phi) Y_n^m(\theta_i,\phi_i)^*. \tag{10}$$

If we want to synthesize this plane wave with L plane wave sources located at (θ_l, ϕ_l) for $1 \leq l \leq L$, each with an amplitude of w_l, then our expression for the synthesized field $\hat{p}$ is:

$$\hat{p} = 4\pi \sum_{n=0}^{\infty} i^n j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta,\phi) \sum_{l=1}^{L} w_l Y_n^m(\theta_l,\phi_l)^*. \tag{11}$$

Then we solve for each n and m in order to have Eq. 11 be equal to Eq. 10:

$$\sum_{l=1}^{L} w_l Y_n^m(\theta_l,\phi_l)^* = Y_n^m(\theta_i,\phi_i)^*. \tag{12}$$

These amplitudes w_l can be solved for up to order N, subject to the requirement that $L \geq (N+1)^2$ for 3-D arrays and $L \geq 2N+1$ for horizontal arrays. This is known as basic decoding.

4.2. Wavefield Error Analysis

One method of evaluating the quality of the reconstruction is by calculating the normalized radial error between the synthesized sound field $\hat{p}$ and the sound field being reconstructed, p, given by:

$$\bar{\epsilon}(kr) = \frac{\int_0^{2\pi}\int_0^{\pi} |p-\hat{p}|^2 \sin\theta \, d\theta \, d\phi}{\int_0^{2\pi}\int_0^{\pi} |p|^2 \, \sin\theta \, d\theta \, d\phi}. \tag{13}$$

This quantity is a function of the distance from the center of the array as related to the wavelength one is examining, expressed as kr, and is shown for 1st through 7th order (moving from left to right) in the top portion of Fig. 2. A given value of kr will be a larger distance from the center of the array for a lower frequency (i.e. longer wavelength) than for a higher frequency. Thus, the sweet spot will be larger for lower frequencies than for higher frequencies.

One way to evaluate the performance of ambisonic systems in terms of human perception is to examine the frequency at which the radius of the sweet spot (as determined by some threshold in the normalized radial error) becomes smaller than the radius of the average human head (given in [4] as 8.9 cm). As the order of ambisonic reproduction increases, so will this frequency, as shown in the bottom portion of Fig. 2, for several different thresholds of normalized radial error. Thus, as higher spherical components are added to the reconstruction, higher frequencies are reconstructed accurately at the two ears.

4.3. Max-r_E Decoding

For a given order of ambisonic reproduction, there will be some frequency above which the area of the sweet spot will be smaller than the size of a typical human head. Thus, audio content above this frequency will not be simulated accurately at the listener's ears. In [13], Gerzon proposed that high frequency localization can be predicted by the direction of the energy vector at the center of the array, $\hat{\mathbf{r}}_\mathbf{E}$, as given in:

$$r_E \hat{\mathbf{r}}_\mathbf{E} = \frac{\sum_{l=1}^{L} w_l^2 \hat{\mathbf{u}}_\mathbf{l}}{\sum_{l=1}^{L} w_l^2}, \tag{14}$$

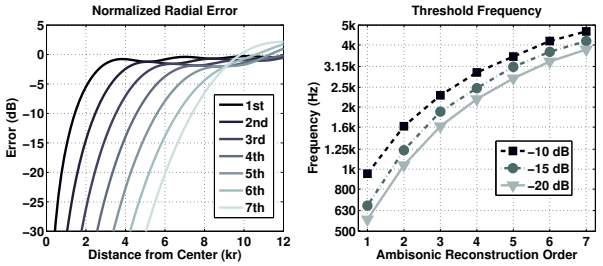

Figure 2: The left shows normalized radial error (in dB) as a function of distance from the center of the array (in units of kr), from 1st to 7th order basic decoding (moving left to right). The right shows the frequencies at which the -10, -15, and -20 dB error thresholds occur at a radius of 8.9 cm, the size of a typical human head.

where w_l is the gain of the lth loudspeaker radiating sound from a position denoted by the vector $\hat{\mathbf{u}}_\mathbf{l}$ from the loudspeaker's position to the center of the array. The magnitude of the vector, r_E, represents the concentration of source energy in the desired direction, and thus the accuracy of high-frequency localization from that direction. In order to maximize this value, one can apply correcting gains $(g_0, g_1, ..., g_N)$ to each order of spherical harmonics (i.e. to the right side of Eq. 12.) Methods for calculating the correcting gains are given in [3]. This is known as max-r_E decoding.

4.4. Binaural Cue Analysis

The localization accuracy of different decoding schemes can be examined using a model of the auditory system used previously in [14] to evaluate the localization of various stereo recording techniques. The auditory periphery is simulated with Head-Related Transfer Functions (HRTFs). The behavior of the basilar membrane and the hair cells is simulated with a gammatone filter bank with 72 bands and half-wave rectification. ITD analysis is performed using an interaural cross-correlation in each frequency band. ILD analysis is performed using an array of excitation/inhibition (EI) cells. This model allows for the creation of maps that correlate ITD and ILD values (expressed in milliseconds and decibels, respectively) with azimuthal directions.

This allows for another way to evaluate the accuracy of ambisonic reproduction - by calculating errors in the ITD and ILD cues rather than errors in the reproduced wavefield, as detailed earlier. First, the spherical harmonic signals are decoded to a virtual 24-channel horizontal loudspeaker array (with equiangular spacing, putting each channel 15 degrees apart in azimuth from its neighbors.) The signals reaching each ear are obtained by convolving the loudspeaker feeds with the appropriate HRTFs and summing over all loudspeakers. Plane waves can then be simulated from a variety of directions, and the ITD and ILD cues can be compared to the natural condition.

From these maps we can then look at average binaural cue error as a function of frequency by averaging over azimuthal angles and vice versa. This is shown for 1st through 5th order ambisonic rendering (using max-r_E decoding) for ITDs in Fig. 3 and for ILDs in Fig. 4, using an HRTF catalog measured for a HEAD Acoustics dummy head. The dotted lines indicate the average error across all frequency bands. For ITD cues, we see that there are significant errors with 1st-order ambisonic decoding, but that

these errors decrease significantly when moving up to 2nd-order decoding, and remain low through higher orders. For both ITD and ILD cues, errors are generally higher farther away from 0 degrees azimuth, indicating that these systems have a limited ability to produce the larger ITD and ILD values generated by lateral sound sources. However, increasing the spherical harmonic order can improve these cues. For ILD cues, errors are generally less at lower frequencies, and the higher the order, the higher the frequency at which significant cue errors occur.

5. PROCESSING MEASURED SOUND FIELDS FOR PLAYBACK

In this section, four hypothetical spherical microphone arrays are considered: two 16-channel and two 64-channel rigid spherical arrays, each set with radii of 2.5 and 5 cm. These arrays utilize a nearly-uniform sampling scheme, with the positions of the sensors given in [15]. Therefore, the 16-channel arrays are capable of decomposing the spherical harmonics up to third order, the 64-channel arrays up to seventh order.

Fig. 5 shows the White Noise Gain thresholds (the lower line corresponding to a WNG value of -30 dB, the upper to -20 dB) and aliasing frequencies, plotted together with the thresholds for transitioning from basic to max-r_E decoding. (Note that the decoding thresholds are based on the size of a typical human head, and thus are not affected by the properties of the microphone array.) Overlaid on these plots are the decoding schemes chosen based on the principles outlined previously, particularly with respect to the WNG values of the array, max-r_E ambisonic decoding, and the spherical harmonic aliasing frequency.

As before, where we compared different orders of ambisonics, we can evaluate the accuracy of the ITD and ILD cues of these various decoding schemes (of "mixed" order) across multiple frequency bands, as shown in Fig. 6 (with the plots of 1st and 5th order ambisonics shown for reference). The average errors across all frequency bands shown in Table 1.

Table 1: *Average ITD and ILD cue error and aliasing frequency for each spherical microphone array.*

Array	Avg. ITD error (ms)	Avg. ILD error (dB)	Aliasing Frequency (Hz)
a=2.5 cm, Q=16	0.24	4.3	6551
a=5 cm, Q=16	0.21	4.1	3275
a=2.5 cm, Q=64	0.20	4.1	15285
a=5 cm, Q=64	0.17	3.8	7643

There are two main points that we can gather from Fig. 6. The first is that increasing the number of channels for a spherical microphone array of a given radius is a "win-win" scenario: higher spherical harmonic orders become available, the WNG threshold frequencies for lower orders are pushed lower in frequency, and the aliasing frequency moves higher, yielding a wider frequency range for reconstruction. Thus, we see improvements in both localization accuracy and bandwidth. Of course, increasing the number of channels will increase the cost and complexity of building the array. In addition, going from a 3rd order to a 7th order microphone array does not yield the same gains in localization accuracy as moving from 3rd order to 7th order ambisonic reproduction of a simulated sound field, as the highest order components are sometimes available for only 1 or 2 octaves before encountering the aliasing frequency.

Increasing the spherical array radius, however, involves a tradeoff. On the positive side, the WNG thresholds move lower, making higher order spherical harmonic components available at lower frequencies. However, the aliasing frequency also moves lower, meaning that we are trading localization accuracy for bandwidth.

6. CONCLUSION

The common basis in spherical harmonics makes ambisonics a natural fit for reproducing sound fields measured with spherical microphone arrays. Incorporating higher order spherical harmonic components offers the opportunity for more precise localization, but at the same time introduces complexities at both the recording stage and the playback stage that need to be dealt with. The goal of this paper is to illuminate the sources of those issues and develop a framework to resolve them, particularly with respect to auditory perception.

7. ACKNOWLEDGMENTS

The project reported here was supported by a Rensselaer HASS fellowship and the National Science Foundation (NSF, #12293911).

8. REFERENCES

[1] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution," *Journal of the Acoustical Society of America*, vol. 116, no. 4, pp. 2149–2157, October 2004.

[2] Z. Li and R. Duraiswami, "Flexible and optimal design of spherical microphone arrays for beamforming," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 2, pp. 702–714, February 2007.

[3] J. Daniel, J.-B. Rault, and J.-D. Polack, "Ambisonic encoding of other audio formats for multiple listening conditions," in *Proc. of the 105th Convention of the Audio Eng. Soc.*, San Francisco, California, USA, September 26–29, 1998.

[4] M. A. Poletti, "Three-dimensional surround sound systems based on spherical harmonics," *Journal of the Audio Engineering Society*, vol. 53, no. 11, pp. 1004–1025, November 2005.

[5] A. J. Heller, E. M. Benjamin, and R. Lee, "A toolkit for the design of ambisonic decoders," in *Linux Audio Conference*, CCRMA, Stanford University, CA, USA, April 12-15 2012.

[6] E. G. Williams, *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press, 1999.

[7] P. M. Morse and K. U. Ingard, *Theoretical Acoustics*. McGraw-Hill, Inc., 1968.

[8] J. Meyer and G. W. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, Florida, USA, May 13–17, 2002.

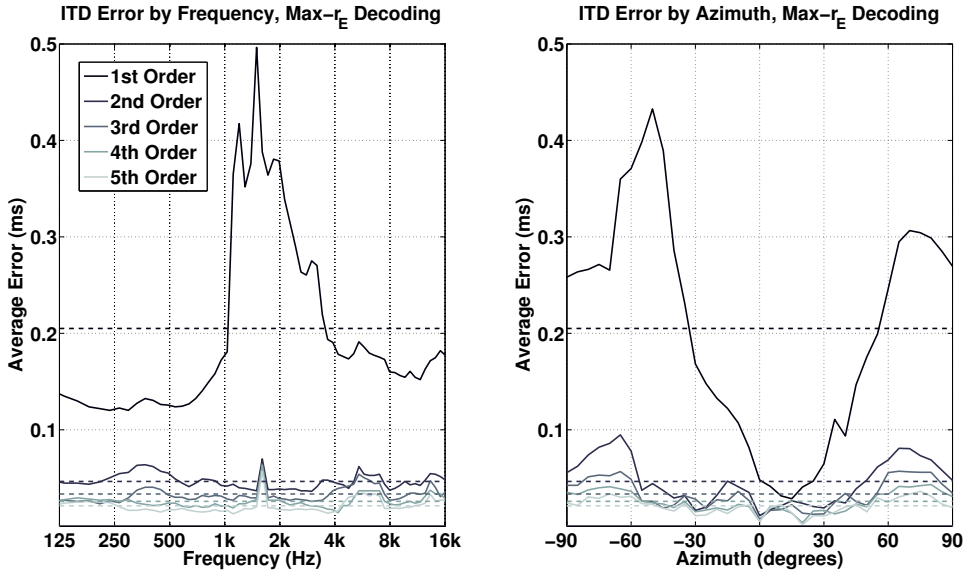

Figure 3: ITD error for 1st through 5th order ambisonics (max-r_E decoding). The left-hand portion shows the error by frequency band averaged over all azimuthal angles, the right-hand side vice versa. Average error over all frequency bands and azimuthal angles shown by the dotted lines.

[9] J. Meyer, G. W. Elko, and T. Agnello, "Spherical microphone array for spatial sound recording," in *Proc. of the 115th Convention of the Audio Eng. Soc.*, New York, New York, USA, October 10–13, 2003.

[10] J. Meyer and G. W. Elko, *Audio Signal Processing for Next-Generation Multimedia Communication Systems*. Hingham, Massachusetts, USA: Kluwer Academic Publishers, 2004, ch. 3, "Spherical Microphone Arrays for 3D Sound Recording", pp. 67–89.

[11] B. Rafaely, B. Weiss, and E. Bachmat, "Spatial aliasing in spherical microphone arrays," *IEEE Transactions on Signal Processing*, vol. 55, no. 3, pp. 1003–1010, March 2007.

[12] M. A. Gerzon, "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, January/February 1973.

[13] ——, "General metatheory of auditory localisation," in *Proc. of the 92nd Convention of the Audio Eng. Soc.*, Vienna, Austria, March 24-27 1992.

[14] J. Braasch, "A binaural model to predict position and extension of spatial images created with standard sound recording techniques," in *Proc. of the 119th Convention of the Audio Eng. Soc.*, New York, NY, USA, October 7–10, 2005.

[15] J. Fliege and U. Maier, "The distribution of points on the sphere and corresponding cubature formulae," *IMA Journal of Numerical Analysis*, vol. 19, no. 2, pp. 317–334, 1999.

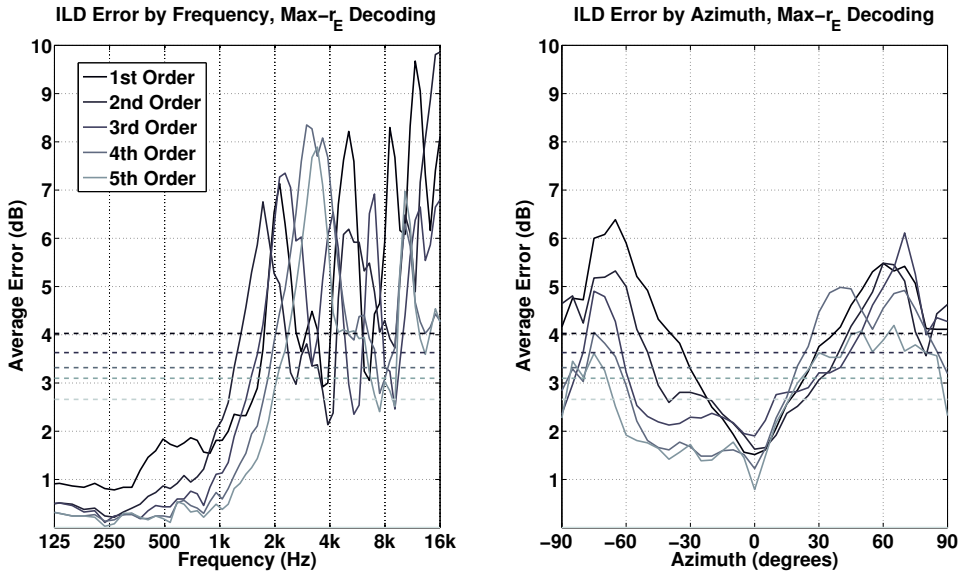

Figure 4: ITD error for 1st through 5th order ambisonics (max-r_E decoding). The left-hand portion shows the error by frequency band averaged over all azimuthal angles, the right-hand side vice versa. Average error over all frequency bands and azimuthal angles shown by the dotted lines.

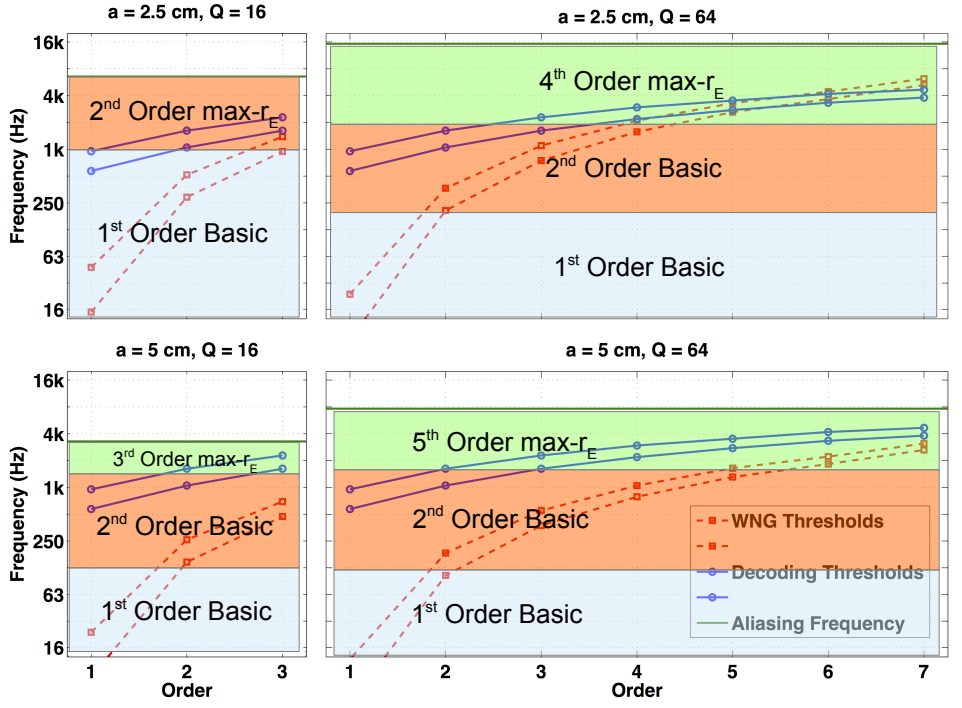

Figure 5: WNG thresholds and aliasing frequencies for 4 spherical arrays, plotted together with basic/max-r_E ambisonic decoding thresholds, overlaid with decoding schemes.

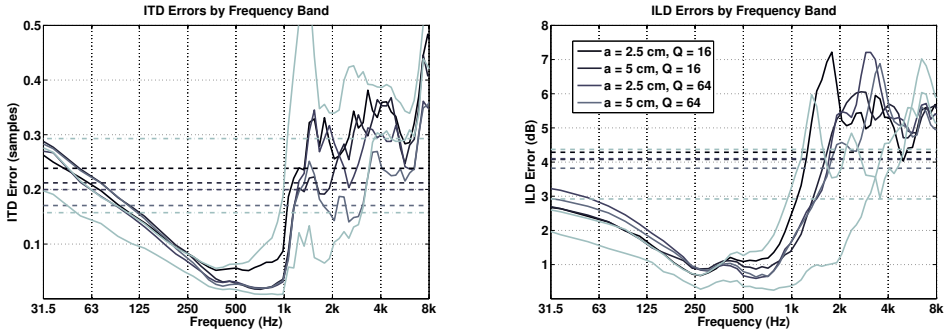

Figure 6: ITD and ILD errors for the 4 different decoding schemes, with the errors for 1st and 5th order ambisonic decoding shown in the lightest color for reference.