Chapter 4
Ambisonic Amplitude Panning and Decoding in Higher Orders

...the second-order Ambisonic system offers improved imaging over a wider area than the first-order system and is suitable for larger rooms.

Jeffrey S. Bamford [1], Canadian Acoustics, 1994.

Abstract  Already in the 1970s, the idea of using continuous harmonic functions of scalable resolution was described by Cooper and then Gerzon, who introduced the name Ambisonics. This chapter starts by reviewing properties of first-order horizontal Ambisonics, using an interpretation in terms of panning functions. And the required mathematical formulations for 3D higher-order Ambisonics are developed here, with the idea to improve the directional resolution. Based on this formalism, ideal loudspeaker layouts can be defined for constant loudness, localization, and width, according to the previous models. The chapter discusses how Ambisonics can be decoded to less ideal, typical loudspeaker setups for studios, concerts, sound-reinforcement systems, and to headphones. The behavior is analyzed by a rich variety of listening experiments and for various decoding applications. The chapter concludes with example applications using free software tools.

Cooper [2] used higher-order angular harmonics to formulate circular panning of auditory events. Due to the work of Felgett [3], Gerzon [4], and Craven [5], the term Ambisonics became common for technology using spherical harmonic functions. Around the early 2000s, most notably Bamford [6], Malham [7], Poletti [8], Jot [9], and Daniel [10] pioneered the development of higher-order Ambisonic panning and decoding. Ward and Abhayapala [11], Dickens [12], and at the lab of the authors Sontacchi [13].

Another leap happened around 2010, when Ambisonic decoding to loudspeakers could be largely improved by considering regularization methods [14], singular-value decomposition [15], and all-round Ambisonic decoding (AllRAD) [15, 16],

© The Author(s) 2019
F. Zotter and M. Frank, Ambisonics, Springer Topics in Signal Processing 19, https://doi.org/10.1007/978-3-030-17207-7_4
a combination of vector-base panning techniques with Ambisonics, yielding the most robust and flexible higher-order decoding method known today.

For headphones, after the work of Jot [9] that outlined the basic problems of binaural decoding in the 1990s, Sun, Bernschütz, Ben-Hur, and Brinkmann [17–19] made important contributions to binaural decoding, and we consider TAC and MagLS decoders by Zaunschirm and Schörkhuber [20, 21] as the essential binaural decoders. Both remove HRTF delays or optimize HRTF phases at high frequencies to avoid spectral artifacts. By interaural covariance correction, MagLS/TAC manage to play back diffuse fields consistently, using the formalism of Vilkamo et al [22].

4.1 Direction Spread in First-Order 2D Ambisonics

In 2D first-order Ambisonics as discussed in Chap. 1, the directional mapping of a single sound source from the angle $\phi_s$ to the direction of each loudspeaker $\phi$ is described by the shape of panning function (or direction-spread function) in Eq. (1.17). The directional spreading is not infinitely narrow, but determined by what can be represented by first-order directivity patterns. Consequently, sound from the angle $\phi_s$ will be mapped by a dipole pattern aligned with the source and an additional omnidirectional pattern. We can involve a spread parameter $a$ to make the directional spread to the loudspeakers system adjustable and either cardioid-shaped $a = 1$, 2D-supercardioid-shaped $a = \sqrt{2}$, or 2D-hypercardioid-shaped $a = 2$, using:

$$g(\phi) = 1 + a \cos(\phi - \phi_s).$$ (4.1)

This function represents how first-order Ambisonic panning would distribute a mono signal to loudspeakers. With the loudspeaker positions described by the set of angles $\{\phi_l\}$, a vector of amplitude-panning gains with an entry for each loudspeaker could be determined by sampling the direction-spread function:

$$g = \begin{bmatrix} g_1 \\ \vdots \\ g_L \end{bmatrix} = 1 + a \begin{bmatrix} \cos(\phi_1 - \phi_s) \\ \vdots \\ \cos(\phi_L - \phi_s) \end{bmatrix}.$$ (4.2)

With these gain values, we evaluate models of perceived loudness, direction, and width, as introduced in Chap. 2, in order to enter a discussion of perceptual goals.

If the loudspeaker directions $\{\theta_l\}$ are chosen suitably, it is possible to obtain panning-independent loudness, direction, and width measures $E = \sum_l g_l^2$, $r_E = \frac{1}{E} \sum_l g_l^2 \theta_l$, and $\frac{\sqrt{3}}{2} \frac{180}{\pi} 2 \arccos ||r_E||$. How is it done?

For first-order 2D Ambisonics, it is theoretically optimal to use at least a ring of 4 loudspeakers with uniform angular spacing and $a = \sqrt{2}$, which is easily checked with the aid of a computer, cf. Fig. 4.1, and explained below and in Sect. 4.4.

Direction spread in FOA. The panning-function interpretation with its directional spread has some similarity to MDAP, with its attempt to directionally spread an amplitude-panned signal. Similar to the discrete virtual spread by $\pm \alpha = \arccos ||r_E||$
4.1 Direction Spread in First-Order 2D Ambisonics

Fig. 4.1 Width-related angle \( \arccos \| \mathbf{r}_E \| \) and angular error \( \angle \mathbf{r}_E - \phi_s \) for first-order Ambisonics on a ring with 90° spaced loudspeakers using a cardioid, or 2D super-/hyper-cardioid patterns; the measures are all panning-invariant and the super-cardioid weighting is the 2D max-\( \mathbf{r}_E \) weighting with \( \arccos \| \mathbf{r}_E \| = 45° \).

The virtual direction spread of first-order Ambisonics is described by its continuous panning function \( g(\varphi) \) in Eq. (4.1). To inspect the continuous function by the \( \mathbf{r}_E \) measure defined in Eq. (2.7), we may evaluate an integral over the panning function instead of the sum. Because of the symmetry around \( \varphi_s \), we may set for convenience \( \varphi_s = 0 \), which knowingly causes \( \mathbf{r}_E, y = 0 \), and evaluate

\[
\mathbf{r}_{E,x} = \frac{\int_0^{2\pi} g^2(\varphi) \cos \varphi \, d\varphi}{\int_0^{2\pi} g^2(\varphi) \, d\varphi} = \frac{\int_0^{\pi} [1 + 2a \cos \varphi + a^2 \cos^2 \varphi] \cos \varphi \, d\varphi}{\int_0^{\pi} [1 + 2a \cos \varphi + a^2 \cos^2 \varphi] \, d\varphi} = \frac{a}{1 + \frac{a^2}{2}}. \tag{4.3}
\]

The maximum of \( \mathbf{r}_{E,x} = \frac{2a}{2 + a^2} \) is found by \( \frac{d}{da} \mathbf{r}_{E,x} = \frac{4 + 2a^2 - 4a^2}{(2 + a^2)^2} = 0 \), hence at \( a = \sqrt{2} \). Consequently, the 2D max-\( \mathbf{r}_E \) weight is \( \mathbf{r}_{E,x} = \frac{\sqrt{2}}{2} = \frac{1}{\sqrt{2}} \) and yields the angle arccos \( \| \mathbf{r}_E \| = 45° \). This would resemble a 2D-MDAP-equivalent source spread to ±45°. Note that first-order Ambisonics cannot map to a smaller spread than this. Only higher orders permit to further reduce this spread to a desired angle below 90°.

**Ideal loudspeaker layouts.** Not only is the directional aiming of the virtual, continuous first-order Ambisonic panning function ideal and its width panning-invariant, also its loudness measure is panning-invariant. However, decoding to a physical loudspeaker setup can degrade the ideal behavior. For which loudspeaker layout are these properties preserved by sampling decoding?

The 2D first-order Ambisonic components \((W, X, Y)\) correspond to \( \{1, \cos \varphi, \sin \varphi\} \) patterns, a first-order Fourier series in the angle. Sampling the playback directions by \( L = 3 \) uniformly spaced loudspeakers on the horizon, the sampling theorem for this series is already fulfilled. Accordingly, Parseval’s theorem ensures panning-invariant loudness \( E \) for any panning direction.

For an ideal \( \mathbf{r}_E \) measure, however, one more loudspeaker is required \( L \geq 4 \) for a uniformly spaced horizontal ring. To explain this increase exhaustively, the concept of circular/spherical polynomials and \( t \)-designs will be introduced in this chapter. For a brief explanation, \( g^2(\varphi) \) is a second-order expression and therefore to represent the ideal constant loudness \( E = \frac{2\pi}{L} \sum_l g_l^2 \), it requires \( L = 3 \) uniformly...
spaced loudspeakers, as argued before. By contrast, the expressions $g^2(\varphi) \cos \varphi$ and $g^2(\varphi) \sin \varphi$ are third-order and appear in $\mathbf{r}_E \cdot \mathbf{E} = \int g^2(\varphi) [\cos \varphi, \sin \varphi]^T d\varphi$. Consequently, ideal mapping of $\mathbf{r}_E$ (direction and width) requires at least one more loudspeaker $L = 4$ for a uniformly spaced arrangement to make the continuous and the discretized form $\mathbf{r}_E \mathbf{E} = \frac{2\pi}{L} \sum_i g_i^2 [\cos \varphi_i, \sin \varphi_i]^T$ perfectly equal.

**Towards a higher-order panning function.** An $N$th-order cardioid pattern is obtained from the cardioid pattern by taking its $N$th power

$$g_N(\varphi) = \frac{1}{2N} (1 + \cos \varphi)^N,$$

which makes it narrower. With $N = 2$, this becomes, using $\cos^2 \varphi = \frac{1}{2} (1 + \cos 2\varphi)$,

$$g_2(\varphi) = \frac{1}{4} (1 + 2 \cos \varphi + \cos^2 \varphi) = \frac{1}{8} (3 + 4 \cos \varphi + \cos 2\varphi).$$

More generally, Chebyshev polynomials $T_m(\cos \varphi) = \cos m\varphi$, cf. [23, Eq. 3.11.6] can be used to argue that there is always a fully equivalent cosine series describing the higher-order 2D panning function in the azimuth angle

$$g(\varphi) = \sum_{m=0}^{N} a_m \cos m\varphi. \quad (4.4)$$

**Rotated panning function.** In first-order Ambisonics, panning functions consist of an omnidirectional part, $\cos(0\varphi) = 1$, and a figure-of-eight to $x$, $\cos \varphi$, but that was not all: Recording and playback also required a figure-of-eight pattern to $y$, $\sin \varphi$. The additional component allows to express rotated first-order directivities by a basis set of fixed directivities. For higher orders, a panning function rotated to a non-zero aiming $\varphi_s \neq 0$

$$g(\varphi - \varphi_s) = \sum_{m=0}^{N} a_m \cos[m(\varphi - \varphi_s)] \quad (4.5)$$

can be re-expressed by the addition theorem $\cos(\alpha + \beta) = \cos \alpha \cos \beta - \sin \alpha \sin \beta$ into a series involving the sinusoids (odd symmetric part of a Fourier series),

$$g(\varphi - \varphi_s) = \sum_{m=0}^{N} a_m [\cos m\varphi_s \cos m\varphi + \sin m\varphi_s \sin m\varphi] \quad (4.6)$$

$$= \sum_{n=0}^{N} a_m^{(c)} \cos m\varphi + \sum_{m=0}^{N} a_m^{(s)} \sin m\varphi.$$
4.2 Higher-Order Polynomials and Harmonics

The previous section required that direction and length of the \( r_E \) vector resulting from amplitude panning on loudspeakers matched the desired auditory event direction and width. Harmonic functions with strict symmetry around a panning direction \( \theta_s \) will help us in achieving this goal and in defining good sampling.

Regardless of the dimensions, be it in 2D or 3D, we desire to define continuous and resolution-limited axisymmetric functions around the panning direction \( \theta_s \) to fulfill our perceptual goals of a panning-invariant loudness \( E \), width \( \| r_E \| \), and perfect alignment between panning direction \( \theta_s \) and localized direction \( r_E \). Then we hope to find suitable directional discretization schemes for ideal loudspeaker layouts, so that the measures \( E \) and \( r_E \) are perfectly reconstructed in playback.

The projection of a variable direction vector \( \theta \) onto the panning direction \( \theta_s \) always yields the cosine of the enclosed angle \( \theta_s^T \theta = \cos \phi \), no matter whether it is in two or three dimensions. Hereby constructing the panning function based on this projection readily meets the desired goals. The \( m \)th power thereof, \( (\theta_s^T \theta)^m = \cos^m \phi \) helps to build an \( N \)th-order power series \( g = \sum_{m=0}^{N} a_m (\theta_s^T \theta)^m \) to describe a virtual Ambisonic panning function.

For 2D, such a circular polynomial \( g = \sum_{m=0}^{N} a_m (\theta_s^T \theta)^m \) contains all \( (N + 1)(N + 2)/2 \) mixed powers by \( (\theta_s^T \theta)^m = (\theta_x \theta_x + \theta_y \theta_y)^m = \sum_{k=0}^{m} (\theta_x \theta_x)^k (\theta_y \theta_y)^{m-k} \) of the direction vectors’ entries \( \theta_s^T = [\theta_x, \theta_y] \) and \( \theta = [\theta_x, \theta_y]^T \). However, we could already recognize that it only takes \( 2N + 1 \) functions to express \( g = \sum_{m=0}^{N} a_m (\theta_s^T \theta)^m = \sum_{m=0}^{N} a_m \cos^m \phi \): First an initial polynomial with relative azimuth \( \phi = \varphi - \varphi_s \) relating to a harmonic series of \( N + 1 \) cosines or Chebyshev-polynomials \( g = \sum_{m=0}^{N} b_m \cos m \phi = \sum_{m=0}^{N} b_m \cos m \phi \). Then, in terms of absolute azimuth \( \varphi \), the trigonometric addition theorem re-expresses the series into one of \( N + 1 \) cosines and \( N \) sines, with \( T_n(\theta, \theta) = \cos[m(\varphi - \varphi_s)] = \cos m \varphi_s \cos m \varphi + \sin m \varphi_s \sin m \varphi \). As shown in the upcoming section, we can alternatively obtain such orthonormal harmonic functions by solving a second-order differential equation that is generally used to define harmonics, which bears the later benefit that we can use the approach to define spherical harmonics in three space dimensions.

Spherical polynomials are similar, \( g = \sum_{n=0}^{N} a_n (\theta_s^T \theta)^n \), involving the expressions \( (\theta_s^T \theta)^n = (\theta_x \theta_x + \theta_y \theta_y + \theta_z \theta_z)^n = \sum_{k=0}^{n} \sum_{l=0}^{n-k} \binom{n}{k} \binom{n-k}{l} (\theta_x \theta_x)^k (\theta_y \theta_y)^{n-k-l} (\theta_z \theta_z)^l \). Again, all these \((N + 1)(N + 2)(N + 3)/6\) combinations would be too many to form an orthogonal set of basis functions. Moreover, while the different cosine harmonics are orthogonal axisymmetric functions in 2D, they are not in 3D. On the sphere, the \( N + 1 \) orthogonal Legendre polynomials \( P_n(\cos \phi) \) replace the cosine series as a basis for \( g = \sum_{n=0}^{N} c_n P_n(\cos \phi) \), as shown below. All mathematical derivations for the sphere rely on the definition of harmonics. They result in \((N + 1)^2\) spherical harmonics and their addition theorem as a basis in terms of absolute directions \( 2^{\frac{n+1}{2}} P_n(\theta_s^T \theta) = \sum_{m=-n}^{n} Y_m^n(\theta_s) Y_m^n(\theta) \). Dickins’ thesis is interesting for further reading [12].

In both regimes, 2D and 3D, the circular or spherical polynomials concept will be used to determine optimal layouts, so-called \( t \)-designs. Such \( t \)-designs are directional
Ambisonic Amplitude Panning and Decoding in Higher Orders

sampling grids that are able to keep the information about the constant part of any either circular (2D) or spherical (3D) polynomials up to the order \( N \leq t \). This will be a mathematical key property exploited to determine requirements for preserving \( E \) and \( r_E \) measures during Ambisonic playback with optimal loudspeaker setups, but not only. Also \( t \)-designs simplify numerical integration of circular or spherical harmonics to define state-of-the-art Ambisonic decoders or mapping effects.

4.3 Angular/Directional Harmonics in 2D and 3D

The Laplacian is defined in the D-dimensional Cartesian space as

\[
\Delta = \sum_{j=1}^{D} \frac{\partial^2}{\partial x_j^2},
\]

and for any function \( f \), the Laplacian \( \Delta f \) describes the curvature. Any harmonic function is proportional to its curvature by an eigenvalue \( \lambda \),

\[
\Delta f = -\lambda f,
\]

and therefore is an oscillatory function. Generally, eigensolutions \( \Delta f = -\lambda f \) to the Laplacian are called harmonics. For suitable eigenvalues \( \lambda \), harmonics span an orthogonal set of basis functions that are typically used for Fourier expansion on a finite interval. It seems desirable to find such harmonics for functions only exhibiting directional dependencies, i.e. in the azimuth angle \( \varphi \) in 2D, and azimuth and zenith angle \( \varphi, \theta \) in 3D.

4.4 Panning with Circular Harmonics in 2D

For 2 dimensions Appendix A.3.2 uses the generalized chain rule to convert the Laplacian of a 2D coordinate system \( \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \) to a polar coordinate system with the radius \( r \) and the angle \( \varphi \) to the x axis, \( \Delta = \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial}{\partial r} \right) + \frac{1}{r^2} \frac{\partial^2}{\partial \varphi^2} \). And for functions \( \Phi = \Phi(\varphi) \) purely in the angle \( \varphi \), the radial derivatives of \( \Delta \Phi \) all vanish and it remains (\( \partial \rightarrow d \))

\[
\frac{d^2}{d\varphi^2} \Phi = -\lambda r^2 \Phi.
\]

It is only yielding useful solutions with \( \lambda r^2 = m^2, m \in \mathbb{Z} \), cf. Appendix A.3.4, Fig.4.2,
4.4 Panning with Circular Harmonics in 2D

Fig. 4.2 Circular harmonics with $m = -3, \ldots, 3$ plotted as polar diagram using the radius $R = 20 \lg |\sqrt{\pi} \Phi_m|$ and grayscale to distinguish between positive (gray) and negative (black) signs

\[
\Phi_m = \frac{1}{\sqrt{2\pi}} \begin{cases} 
\sqrt{2} \sin(|m| \varphi), & \text{for } m < 0, \\
1, & \text{for } m = 0, \\
\sqrt{2} \cos(m \varphi), & \text{for } m > 0,
\end{cases}
\] (4.10)

which defines how to decompose panning functions of limited order $|m| < N$. The harmonics are periodic in azimuth, orthogonal and normalized (orthonormal) on the period $-\pi \leq \varphi \leq \pi$. Due to their completeness, any square-integrable function $g(\varphi)$ can be expanded into a series of the harmonics using coefficients $\gamma_m$

\[
g(\varphi) = \sum_{m=-\infty}^{\infty} \gamma_m \Phi_m(\varphi).
\] (4.11)

For a known function $g(\varphi)$, the coefficients $\gamma_m$ are obtained by the transformation integral

\[
\gamma_m = \int_{-\pi}^{\pi} g(\varphi) \Phi_m(\varphi) \, d\varphi,
\] (4.12)

as shown in Appendix Eq. (A.14).

**2D panning function.** An infinitely narrow angular range around a desired direction $|\varphi - \varphi_s| < \varepsilon \to 0$ is represented by the transformation integral over a Dirac delta distribution $\delta(\varphi - \varphi_s)$, cf. Appendix Eq. (A.16), so that the coefficients of such a panning function are

\[
\gamma_m = \Phi_m(\varphi_s).
\] (4.13)

As the infinite circular harmonic series is complete, the panning function is

\[
g(\varphi) = \sum_{m=-\infty}^{\infty} \Phi_m(\varphi_s) \Phi_m(\varphi) = \delta(\varphi - \varphi_s),
\] (4.14)

and in practice we resolution-limit it to the Nth Ambisonic order, $|m| \leq N$, and use an additional weight $a_m$ that allows us to design its side lobes.
Fig. 4.3 2D unweighted $a_n = 1$ basic and weighted max-$r_E$ Ambisonic panning functions for the orders $N = 1, 2, 5$

$$g_N(\psi) = \sum_{m=-N}^{N} a_m \Phi_m(\psi_s) \Phi_m(\psi). \quad (4.15)$$

The max-$r_E$ panning function [24] uses the weights $a_m = \cos(\frac{\pi m}{2(N+1)})$, as derived in Appendix Eq. (A.20). The spread is now adjustable by the order to $\pm \frac{90^\circ}{N+1}$. The result is shown in Fig. 4.3, compared with no side-lobe suppression when $a_n = 1$ (basic).

It is easy to recognize: $\Phi_m(\psi_s)$ represents the recorded or encoded directions, and $\Phi_m(\psi)$ represents the decoded playback directions.

**Optimal sampling of the 2D panning function.** In the theory of circular/spherical polynomials in the variable $\zeta = \cos(\phi - \psi_s)$, so-called $t$-designs in 2D are optimal point sets of given angles $\{\psi_l\}$ with $l = 1, \ldots, L$ and size $L$. A $t$-design allows to perfectly compute the integral (constant part) over the polynomials $P_m(\zeta)$ of limited degree $m \leq t$ by discrete summation

$$\int_{-\pi}^{\pi} P_m(\cos \phi) \, d\phi = \sum_{l=1}^{L} P_m[\cos(\psi_l - \psi_s)] \frac{2\pi}{L}, \quad (4.16)$$

regardless of any angular shift $\psi_s$. In 2D, Chebyshev polynomials $T_m(\cos \phi) = \cos(m\phi)$ are orthogonal polynomials, therefore an Nth-order panning function composed out of $\cos(m\phi)$ is always a polynomial of Nth degree. Knowing this, it is clear that the integral over $g_N^2$ required to evaluate the loudness measure $E$ is a polynomial of the order $2N$. The integral to calculate $r_E$ is over $g_N^2 \cos(\phi)$ and thus of the order $2N + 1$. In playback, to get a perfectly panning-invariant loudness measure $E$ of the continuous panning function and also the perfectly oriented $r_E$ vector of constant spread $\arccos||r_E||$, the parameter $t$ must be $t \geq 2N + 1$. In 2D, all regular polygons are $t$-designs with $L = t + 1$ points

$$\psi_l = \frac{2\pi}{t+1} (l - 1). \quad (4.17)$$

We can use the smallest set of $2N + 2$ angles $\psi_l = \frac{180^\circ}{N+1} (l - 1)$ as optimal 2D layout.
4.5 Ambisonics Encoding and Optimal Decoding in 2D

To encode a signal $s$ into Ambisonic signals $\chi_m$, we multiply the signal with the encoder representing the direction of the signal at the angle $\varphi_s$ by the weights $\Phi_m(\varphi_s)$

$$\chi_m(t) = \Phi_m(\varphi_s) s(t),$$

or in vector notation

$$\chi_N = y_N(\varphi_s) s,$$

using the column vector $y_N = [\Phi_{-N}(\varphi_s), \ldots, \Phi_N(\varphi_s)]^T$ of $2N + 1$ components. The Ambisonic signals in $\chi_N$ are weighted by side-lobe suppressing weights $a_N = [a_{-N}, \ldots, a_N]^T$, expressed by the multiplication with a diagonal matrix $\text{diag}(a_N)$, and then decoded to the $L$ loudspeaker signals $x$ by a sampling decoder

$$D = \sqrt{\frac{2\pi}{L}} [y_N(\varphi_1), \ldots, y_N(\varphi_L)]^T = \sqrt{\frac{2\pi}{L}} Y_N^T,$$

using

$$x = D \text{diag}(a_N) \chi_N.$$  

In total, the system for encoding and decoding can also be written to yield a set of loudspeaker gains for one virtual source

$$g = D \text{diag}(a_N) y_N(\varphi_s),$$

or in particular for the 2D sampling decoder $g = \sqrt{\frac{2\pi}{L}} Y_N^T \text{diag}(a_N) y_N(\varphi_s)$.

4.6 Listening Experiments on 2D Ambisonics

There are several listening experiments discussing the features of Ambisonics, most of which are summarized in [25], which will be discussed complemented with those from [26] below.

The perceptually adjusted panning angle of 2nd-order max-$r_E$ Ambisonics panning on 6 horizontal loudspeakers matches quite well the acoustic reference direction as shown in Fig. 4.4, similar to MDAP in Fig. 3.8, but with a slightly more
Fig. 4.4 2nd-order max-$r_E$-weighted Ambisonic panning with pink-noise on horizontal rings of 60°-spaced loudspeakers adjusted to perceptually match reference loudspeaker directions (harmonic complex) every 15°. Markers and whiskers indicate 95% confidence intervals and medians, black curve the $r_E$ vector model.

Fig. 4.5 In Frank’s 2008 pointing experiment [27] on center and off-center listening seats for 3 virtual sources (A, B, C) using 1st-order (left) and 5th-order (right) Ambisonics on 12 horizontal loudspeakers (IEM CUBE) indicate a more stable localization with high orders. Moreover, for 5th-order, max-$r_E$ weighting and omission of delay compensation were preferred. Omission of max-$r_E$ weights (“basic”) or alternative “in-phase” weights that entirely suppresses any side lobe yield less precise localization at off-center listening positions.
4.6 Listening Experiments on 2D Ambisonics

Fig. 4.6 Experiments on an off-center position in a show that max-$r_E$ outperforms the basic, rectilinearly truncated Fourier series at off-center listening positions, b where it can avoid splitting of the auditory event. Stitt’s experiments c imply that localization with higher orders is more stable and that the localization deficiency at off-center listening seats seems to be proportional to the ratio between distance to the center divided by radius of the loudspeaker ring, and not the specific time-delays that are larger for large loudspeaker rings, cf. [28]

accurate median by 0.5° on average, and in particular at side and back panning directions.

Another aspect to investigate is how stable the results are for center and off-center listening seats as shown in Fig. 4.5. It illustrates that max-$r_E$ with the highest order achieves the best stability with regard to localization at off-center listening seats. Astonishingly, the delay compensation for non-uniform delay times to the center deteriorated the results, most probably because of the nearly linear frontal
arrangement of loudspeakers that is more robust to lateral shifts of the listening positions than a circular arrangement.

Figure 4.6a, b shows the direction histogram for two different weightings $a_m$, and it illustrates that proper sidelobe suppression of the panning function by using max-$r_E$ weights is decisive at shifted listening positions to avoid splitting of the auditory image, as it appears in Fig. 4.6b without the weights (basic).

Peter Stitt’s work shows that the localization offsets at off-center listening seats do not increase with the radius of the loudspeaker arrangement as long as the off-center seat stays in proportion to the radius, Fig. 4.6c. The result are predicted by the sweet area model from Sect. 2.2.9 for the first order (top row) and third order (bottom row) in Fig. 4.7, with both sizes small setup (left) and large setup (right).
4.6 Listening Experiments on 2D Ambisonics

(a) Listening area size experiment

(b) Sweet area for frontal sound

Fig. 4.8 The perceptual sweet spot size as investigated by Frank [29] is nearly covering the entire area enclosed by the IEM CUBE as a playback setup (black = 5th, gray = 3rd, light gray = 1st order Ambisonics). It is smallest for 1st-order Ambisonics.

Fig. 4.9 Frank’s 2013 experiments showed for max-$r_E$ Ambisonics on three frontal loudspeakers that large perceived widths are not entirely accurately modeled by the optimally-sampled, therefore constant $r_E$ vector for low orders, however reasonably well, and significantly for high orders/narrow width.

Frank’s 2016 experiments [29] used scales on the floor from which listeners read off where the sweet area ends in every radial direction, cf. Fig. 4.8a. For Fig. 4.8b, the criterion for listeners to indicate leaving the sweet area was when the frontally panned sound was mapped outside the loudspeaker pairs L, C, and R. It showed that a sweet area providing perceptually plausible playback measures at least $\frac{2}{3}$ of the radius of the loudspeaker setup if the order is high enough.

The perceived width of auditory events is investigated in the experimental results of Fig. 4.9, [25], in which pink noise was frontally panned in different orientations of the loudspeaker ring (with one loudspeaker in front, with front direction lying quarter- and half-spaced wrt. loudspeaker spacing). Listeners compared the width of multiple stimuli, and the results were expected to indicate constant width for the differently rotated loudspeaker ring, as the optimal arrangement with $L = 2N + 2$. 
Frank’s 2013 experiments on the variation of the sound coloration of virtual sources rotating at a speed of 100°/s imply that max-$r_E$ weighting outperforms the “basic” weighting with $a_m = 1$, and that adjusting the number of loudspeakers to the Ambisonic order seems to be reasonable. The perceived widths with 3rd order on 8 loudspeakers, for which the on-loudspeaker position is perceived as being significantly wider. By contrast, the high-order experiment with 7th order on 16 loudspeakers would perfectly validate the model.

Figure 4.10 shows experiments investigating the time-variant change in sound coloration for a pink-noise virtual source rotating at a speed of 100°/s, and for different Ambisonic panning setups. There is an obvious advantage of a reduced fluctuation in coloration at both listening positions, centered and off-center, when using the side-lobe-suppressing “max-$r_E$” weighting instead of the “basic” rectangular truncation of the Fourier series. At the off-center listening position, max-$r_E$ weights achieve good results with regard to constant coloration for both 3rd and 7th order arrangements with 8 and 16 loudspeakers that were investigated.

**How well would diffuse signals be preserved played back?** All the above experiments deal with how non-diffuse signals are presented. To complement what is shown in Fig. 1.21 of Chap. 1 with an explanation, the relation between Ambisonic order and its ability to preserve diffuse fields is estimated here by the covariance between uncorrelated directions. Assume a max-$r_E$-weighted Nth-order Ambisonic panning function $g(\theta^T \theta)$ that is normalized to $g(1) = 1$, encodes two sounds $s_{1,2}$ from two directions $\theta_1$ and $\theta_2$, with the sounds being uncorrelated and unit-variance $E\{s_1 s_2\} = \delta_{1,2}$. We can find that the Ambisonic representation mixes the sounds at their respective mapped directions and yields an increase of their correlation $x_1 = s_1 + g_{12} s_2$ and $x_2 = s_2 + g_{12} s_1$, using $g_{12} = g(\cos \phi)$,

$$R\{x_1 x_2\} = \frac{E\{x_1 x_2\}}{\sqrt{E\{x_1^2\} E\{x_2^2\}}} = \frac{E\{(1 + g_{12}^2) s_1 s_2 + g_{12} (s_1^2 + s_2^2)\}}{\sqrt{E\{s_1^2 + 2 g_{12} s_1 s_2 + g_{12}^2 s_2^2\} E\{s_2^2 + 2 g_{12} s_1 s_2 + g_{12}^2 s_1^2\}}} = \frac{2 g_{12}}{1 + g_{12}^2}. \quad (4.23)$$
4.6 Listening Experiments on 2D Ambisonics

This result was presented in Fig. 1.21 and was used to argue that the directional separation of first-order Ambisonics by its high crosstalk term $g_{12}$ might be too weak. Higher-order Ambisonics decreases this directional crosstalk and therefore improves the representation of diffuse sound fields.

4.7 Panning with Spherical Harmonics in 3D

In three space dimensions, the spherical coordinate system has a radius $r$ and two angles, azimuth $\varphi$ indicating the polar angle of the orthogonal projection to the $xy$ plane, and the zenith angle $\vartheta$ indicating the angle to the $z$ axis, according to the right-handed spherical coordinate systems in ISO31-11, ISO80000-2, [30, 31], Fig. 4.11.

By the generalized chain rule, Appendix A.3 re-writes the Laplacian to spherical coordinates in 3D with $r$ signifying the radius, $\varphi$ the azimuth angle, and the zenith angle $\vartheta$ re-expressed as $\zeta = \frac{\vartheta}{r} = \cos \vartheta$, yielding the operator $\Delta = \frac{2}{r} \frac{\partial}{\partial r} + \frac{\partial^2}{\partial \varphi^2} + \frac{1}{r^2(1-\zeta^2)} \frac{\partial^2}{\partial \vartheta^2} - \frac{2}{r^2} \frac{\partial}{\partial \zeta} + \frac{1-\zeta^2}{r^2} \frac{\partial^2}{\partial \zeta^2}$. Any radius-dependent part is removed to define an eigenproblem yielding the basis for panning functions, taking only $r^2 \Delta_{\varphi,\zeta,3D},$

$$\begin{bmatrix} \frac{1}{1-\zeta^2} \frac{\partial^2}{\partial \varphi^2} - \frac{2\zeta}{\partial \zeta} + (1-\zeta^2) \frac{\partial^2}{\partial \zeta^2} \end{bmatrix} Y = -\lambda Y \quad (4.24)$$

whose solution with $\lambda = n(n+1)$ defines the spherical harmonics

$$Y_n^m(\theta) = Y_n^m(\varphi, \zeta) = \Theta_n^m(\vartheta) \Phi_m(\varphi). \quad (4.25)$$

The pre-requisites are (i) periodicity in $\varphi$ and (ii) that the function $Y_n^m$ is finite on the sphere. In addition to the circular harmonics $\Phi_m$ expressing the dependency on azimuth $\varphi$ according to Eq. (4.10), the spherical harmonics contain the associated Legendre functions $P_n^m$ and their normalization term

$$\Theta_n^m(\vartheta) = N_n^{|m|} P_n^{|m|}(\cos \vartheta) \quad (4.26)$$

Fig. 4.11 The spherical coordinate system
to express the dependency on the zenith angle \( \vartheta \). The index \( n \geq 0 \) expresses the order and the directional resolution can be limited by requiring \( 0 \leq n \leq N \). The index \( m \) is the degree and for each \( n \) it is limited by \( -n \leq 0 \leq n \).

The spherical harmonics, Fig. 4.12, are orthonormal on the sphere \( -\pi \leq \varphi \leq \pi \) and \( 0 \leq \vartheta \leq \pi \), and for unbounded order \( N \to \infty \) they are complete; see also Appendix A.3.7.

The spherical harmonics permit a series representation of square-integrable 3D directional functions by the coefficients \( \gamma_{nm} \),

\[
g(\theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \gamma_{nm} Y_n^m(\theta). \tag{4.27}
\]

From a known function \( g(\theta) \), the coefficients are obtained by the transformation integral over the unit sphere \( S^2 \), cf. appendix Eq. (A.38)

\[
\gamma_{nm} = \int_{S^2} g(\theta) Y_n^m(\theta) \, d\theta. \tag{4.28}
\]

Note that the above \( N3D \) normalization \( \int_{\vartheta \in S^2} |Y_n^m(\theta)|^2 \, d\theta = 1 \) defines each spherical harmonic except for an arbitrary-phase it might be multiplied with. Legendre functions for the zenith dependency might be defined differently in literature, and for azimuth, some implementations use \( \sin(m \varphi) \) instead of \( \sin(|m| \varphi) \).

In Ambisonics, real-valued functions and the \( SN3D \) normalization \( \sqrt{\frac{1}{2} \frac{(n-|m|)!}{(n+|m|)!}} \) are preferred, and positive signs of the first-order dipole components in the directions of the respective coordinate axes, \( x, y, z \), are preferred. This might require to involve

![Fig. 4.12 Spherical harmonics indexed by Ambisonic channel number \( ACN = n^2 + n + m \); rows show spherical harmonics for the order \( 0 \leq n \leq 3 \) with the \( 2n + 1 \) harmonics of the degree \( -n \leq m \leq n \). What is plotted is a polar diagram with the radius \( R = 20 \log |Y_n^m| \) normalized to the upper 30 dB of each pattern, with positive (gray) and negative (black) color indicating the sign. The order \( n \) counts the circular zero crossings, and \( |m| \) counts those running through zenith and nadir.](image-url)
the Condon-Shortley phase \((-1)^m\) to correct the signs of the Legendre functions, or \(-1\) for \(m < 0\) to correct the sign of azimuthal sinusoids, depending on the implementation of the respective functions. It is often helpful to employ converters and directional checks to ensure compatibility!

3D panning function. An infinitely narrow direction range around a desired direction \(\theta_s^T \theta > \cos \varepsilon \rightarrow 1\) is represented by the transformation integral over the Dirac delta \(\delta(1 - \theta_s^T \theta)\), cf. Eq. (A.41), so that the coefficients of the panning function are

\[
y_{nm} = Y_n^m(\theta_s) .
\] (4.29)

As infinitely many spherical harmonics are complete, the panning function is

\[
g(\theta) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} Y_n^m(\theta_s) Y_n^m(\theta) = \delta(1 - \theta_s^T \theta) ,
\] (4.30)

and in practice, the finite-resolution Nth-order panning function with \(n \leq N\) employs a weight \(a_n\) to reduce side lobes and optimize the spread

\[
g_N(\theta) = \sum_{n=0}^{N} \sum_{m=-n}^{n} a_n Y_n^m(\theta_s) Y_n^m(\theta).
\] (4.31)

The max-\(r_E\) panning function uses the weights \(a_n = P_n[\cos(\frac{137.9\degree}{N+1.51})]\), as derived in Appendix Eq. (A.46). The spread is now adjustable by the order to \(\pm \frac{137.9\degree}{N+1.51}\).

Figure 4.13 shows a comparison to the basic weighting \(a_n = 1\). An alternative expression that uses Legendre polynomials \(P_n\) and only depends on the angle \(\phi\) to the panning direction \(\theta_s\) is obtained by replacing the sum over \(m\) by the spherical harmonics addition theorem \(\sum_{m=-n}^{n} Y_n^m(\theta_s) Y_n^m(\theta) = \frac{2n+1}{4\pi} P_n(\cos \phi)\),

\[
g_N(\phi) = \sum_{n=0}^{N} \frac{2n+1}{4\pi} a_n P_n(\cos \phi).
\] (4.32)

Comparison to first-order Ambisonics shows: now \(Y_n^m(\theta_s)\) represents the recorded or encoded directions, and \(Y_n^m(\theta)\) represents the decoded playback directions.

Optimal sampling of the 3D panning function. In the theory of spherical polynomials in the variable \(\zeta = \theta_s^T \theta\), so-called \(t\)-designs describe point sets of given directions \(\{\theta_l\}\) with \(l = 1, \ldots, L\) and size \(L\) that allow to perfectly compute the integral (constant part) over the polynomials \(P_n(\zeta)\) of limited order \(n \leq t\) by discrete summation

\[
\int_{-\pi}^{\pi} d\phi \int_{-1}^{1} P_n(\zeta) d\zeta = \sum_{l=1}^{L} P_n(\theta_s^T \theta_l) \frac{4\pi}{L}
\] (4.33)
relative to any axis $\theta$, the point set is projected onto. In 3D, the Legendre polynomials $P_n(\zeta)$ are orthogonal polynomials, therefore an Nth-order panning function composed thereof is a polynomial of Nth order. The loudness measure $E$ is calculated by the integral over $g^2_N$, therefore over a polynomial of the order 2N. The integral to calculate $r_E$ runs over $g^2_N \zeta$, therefore over a polynomial of the order 2N + 1. In playback, to get a perfectly panning-invariant loudness measure $E$ of the continuous panning function and also the perfectly oriented $r_E$ vector of constant spread $\arccos ||r_E||$, the parameter $t$ must be $t \geq 2N + 1$. In 3D there are only 5 geometrically regular layouts

- the tetrahedron, $L = 4$ corners, is a 2-design,
- the octahedron, $L = 6$ corners, is a 3-design,
- the hexahedron (cube), $L = 8$ corners, is a 3-design,
- the icosahedron, $L = 12$ corners, is a 5-design,
- the dodecahedron, $L = 20$ corners, is a 5-design.

For instance, for $N = 1$, the octahedron is a suitable spherical design, for $N = 2$, the icosahedral or dodecahedral layouts are suitable.

Exceeding the geometrically regular layouts, there are designs found by optimization to be regular under the mathematical rule to approximate $\int_{S^2} Y^m_n(\theta) \ d\theta = \sqrt{4\pi}\delta_n$ accurately by $\frac{4\pi}{L} \sum_i Y^m_n(\theta_i)$ for all $n \leq t$ and $|m| \leq n$. A large collection can be found by Hardin and Sloane [32], Gräf and Potts [33], and Womersley [34] available on the following websites
http://neilsloane.com/sphdesigns/dim3/
http://homepage.univie.ac.at/manuel.graef/quadrature.php
(Chebyshev-type Quadratures on $S^2$), and
https://web.maths.unsw.edu.au/~rsw/Sphere/EffSphDes/ss.html.

Figure 4.14 gives some graphical examples.
4.8 Ambisonic Encoding and Optimal Decoding in 3D

To encode a signal $s$ into Ambisonic signals $\chi_{nm}$, we multiply the signal with the encoder representing the direction $\theta_s$ of the signal by the weights $Y^m_n(\theta_s)$

$$\chi_{nm}(t) = Y^m_n(\theta_s) s(t), \quad (4.34)$$

or in vector notation

$$\chi_N = y_N(\theta_s) s, \quad (4.35)$$

using the column vector $y_N = [Y^0_0(\theta_s), Y^{-1}_1(\theta_s), \ldots, Y^N_N(\theta_s)]^T$ of $(N+1)^2$ components. The Ambisonic signals in $\chi_N$ are weighted by side-lobe suppressing weights $a_N = [a_0, a_1, a_1, a_1, a_2, \ldots, a_N]^T$, expressed by the multiplication with a diagonal matrix $\text{diag}\{a_N\}$, and then decoded to the $L$ loudspeaker signals $x$ by a sampling decoder

$$D = \sqrt{\frac{4\pi}{L}} \begin{bmatrix} y_N(\theta_1), & \ldots, & y_N(\theta_L) \end{bmatrix}^T = \sqrt{\frac{4\pi}{L}} Y_N^T, \quad (4.36)$$

using

$$x = D \text{diag}\{a_N\} \chi_N. \quad (4.37)$$

In total, the system for encoding Eq. (4.35) and decoding Eq. (4.36) can also be written to yield loudspeaker gains for one signal

$$g = D \text{diag}\{a_N\} y_N(\theta_s), \quad (4.38)$$

or in particular for the 3D sampling decoding $g = \sqrt{\frac{4\pi}{L}} Y_N^T \text{diag}\{a_N\} y_N(\theta_s)$. 

Fig. 4.14 $t$-designs from Gräf’s website (Chebyshev-type quadrature)
4.9 Ambisonic Decoding to Loudspeakers

Ambisonic decoding to loudspeakers has been dealt with by numerous researchers, in the past, particularly because result are not very stable for first-order Ambisonics, and later because they strongly depend on how uniform the loudspeaker layout is for higher-order Ambisonics. Moreover, Solvang found that even the use of too many loudspeakers has a degrading effect [35].

For first-order decoding, the Vienna decoders by Michael Gerzon [36] are often cited, and for higher-order Ambisonic decoding, one can, e.g. find works by Daniel with max-$r_E$ [37] and pseudo-inverse decoding [10], also by Poletti [14, 38, 39].

What turned out to be the most practical solution, is the All-Round Ambisonic Decoding approach (AllRAD) due to its feature of allowing imaginary loudspeaker insertion and downmix as described in the sections above, cf. [40]. It moreover does not have restrictions on the Ambisonics order, which for other decoders often yields poor controllability of panning-dependent fluctuations in loudness and directional mapping errors.

The playable set of directions $\theta_l$ or $\varphi_l$ is usually finite and discrete, and it is represented by the surrounding loudspeakers’ directions. The directional distribution of the surrounding loudspeakers is typically neither a $t$-design (with $t \geq 2N + 1$ in general, sometimes not even regular polygons with $L \geq 2N + 2$ loudspeakers for 2D, in particular). In such cases, it is extremely helpful to be aware of the properties of the various decoder design methods.

4.9.1 Sampling Ambisonic Decoder (SAD)

The sampling decoder as introduced above is the simplest decoding method. For dimensions ($D = 2$) and three ($D = 3$), it uses the matrix $Y_N = [y_N(\theta_1), \ldots, y_N(\theta_L)]$ containing the respective circular or spherical harmonics $y_N(\theta)$ sampled at the loudspeaker directions $\{\theta_l\}$,

$$D = \sqrt{\frac{S_{D-1}}{L}} Y_N^T ,$$

(4.39)

with the circumference of the unit circle denoted as $S_1 = 2\pi$ or the surface of the unit sphere written as $S_2 = 4\pi$. The factor $\sqrt{\frac{S_{D-1}}{L}}$ expresses that each loudspeaker synthesizes a fraction of the $E$ measure on the circle or sphere of the surrounding directions. However, the sampling decoder would neither yield perfectly constant loudness and width measures, $E$, $\|r_E\|$, nor a correct aiming of the localization measure $r_E$ if the loudspeaker layout wasn’t optimal. For instance concerning loudness, for panning towards directional regions of poor loudspeaker coverage, sampling misses out the main lobe of the panning function, yielding a noticeably reduced loudness.
4.9.2 Mode Matching Decoder (MAD)

The mode-matching method is used in [10, 39] and yields a fundamentally different decoder design. Its concept is to re-encode the gain vector $g$ of the loudspeakers for any panning direction $\theta_s$ by the encoding matrix $Y_N = [y_N(\theta_1), \ldots, y_N(\theta_L)]$ for all loudspeaker directions $\{\theta_l\}$. Ideally, the re-encoded result should match the encoding of the panning direction with sidelobes suppressed

$$Y_N g = \text{diag}(a_N) y_N(\theta_s).$$

Using the definition $g = D \text{diag}(a_N) y_N(\theta_s)$ of the panning gains, we obtain

$$Y_N D \text{diag}(a_N) y_N(\theta_s) = \text{diag}(a_N) y_N(\theta_s),$$

so that the decoder $D$ is required to be right-inverse to the matrix $Y_N$, i.e. $Y_N D = Y_N Y_N^T (Y_N Y_N^T)^{-1} = I$, see Eq. (A.63) in Appendix A.4. For the inverse of $Y_N Y_N^T$ to exist, it is necessary to have at least as many loudspeakers as harmonics, i.e. $L \geq (N + 1)^2$ with $D = 3$ or $L \geq 2N + 1$ for $D = 2$. However, this is not a sufficient criterion yet: In directions poorly covered with loudspeakers, the inversion will boost the loudness, so that the result is often numerically ill conditioned for $(Y_N Y_N^T)^{-1}$ unless the loudspeaker layout is uniformly designed, at least. Mode matching decoding is ill-conditioned on hemispherical or semicircular loudspeaker layouts. The solution is equivalently described by the more general pseudo inverse $Y_N^\dagger$, which is right-inverse for fat matrices.

4.9.3 Energy Preservation on Optimal Layouts

For instance, for an order of $N = 2$, 2D Ambisonics should work optimally with a ring of 45° spaced loudspeakers on the horizon, a circular $(2N + 1)$-design, or for 3D, a spherical $(2N + 1)$-design. On a $t$-design selected by $t \geq 2N$, the loudness measure $E$ is panning-invariant, in general,

$$E = \|g\|^2 = y_N^T(\theta_s) \text{diag}(a_N) D^T D \text{diag}(a_N) y_N(\theta_s) = \|\text{diag}(a_N) y_N(\theta_s)\|^2 = \text{const.}$$

This is because a $t \geq 2N$-design discretization preserves orthonormality

$$\int y_N(\theta) y_N^T(\theta) \, d\theta = \frac{S_{D-1}}{L} \sum_{l=1}^{L} y_N(\theta_l) y_N^T(\theta_l) = \frac{S_{D-1}}{L} Y_N Y_N^T = I,$$
which implies for the sampling decoder $D^T D = \frac{S_{D-1}}{L} Y_N Y_N^T = I$, and we notice the panning invariant norm of $g(\theta)$ within its coefficients $\gamma_N = \text{diag}(a_N) y_N(\theta_s)$ by the Parseval theorem $\int g^2(\theta) d\theta = \|\gamma_N\|^2$. The panning invariant $E$ measure also holds for the mode-matching decoder using a $t \geq 2N$-design, as it becomes equivalent to a sampling decoder $D = \sqrt{\frac{L}{S_{D-1}}} Y_N^T (Y_N Y_N^T)^{-1} = \sqrt{\frac{L}{S_{D-1}}} Y_N^T S_{D-1}^{-1} = \sqrt{\frac{S_{D-1}}{L}} Y_N^T$. Under these ideal conditions, both decoders are energy-preserving.

### 4.9.4 Loudness Deficiencies on Sub-optimal Layouts

For 2D layouts, Fig. 4.15 shows what happens if a decoder is calculated for a $t \geq 2N + 1$-design with one loudspeaker removed: While, for panning across the gap, the sampling Ambisonic decoder (SAD) yields a quieter signal, moderate localization errors and width fluctuation, the mode-matching decoder (MAD) yields a strong loudness increase and severe jumps in the localization/width. MAD is therefore not very practical with sub-optimal layouts, SAD only slightly more so.

### 4.9.5 Energy-Preserving Ambisonic Decoder (EPAD)

To establish panning-invariant loudness for decoding to non-uniform surround loudspeaker layouts one can ensure a constant loudness measure $E$ by enforcing $D^T D = I$, which is otherwise only achieved on $t \geq 2N$-designs. We may search for
a decoding matrix \( D \) whose entries are closest to the sampling decoder under the constraint to be column-orthogonal:

\[
\| D - \sqrt{\frac{N}{L}} Y_N^T \|_{\text{Fro}}^2 \rightarrow \min \tag{4.42}
\]

subject to \( D^T D = I \).

The singular value decomposition of

\[
Y_N^T = U [\text{diag}\{s\}, 0]^T V^T
\tag{4.43}
\]

can be used to create

\[
D = U [I, 0]^T V^T. \tag{4.44}
\]

by replacing the singular values \( s \) with ones. Such a decoder is column-orthogonal, as the singular-value decomposition delivers \( U^T U = I \) and \( V V^T = I \), and as a consequence\(^1\) \( D^T D = I \). The energy-preserving decoder in this basic version requires \( L \geq 2N + 1 \) loudspeakers in 2D or \( L \geq (N + 1)^2 \) in 3D to work.

Note that if the loudspeaker setup directions are already a \( t \geq 2N \) design, the sampling, mode-matching, and energy-preserving decoders are equivalent.

### 4.9.6 All-Round Ambisonic Decoding (AllRAD)

In Chap. 3 on vector-base amplitude panning methods, a well-balanced panning result in terms of loudness, width, and localization was achieved by MDAP that distributes a signal to an arrangement of several superimposed VBAP virtual sources. Hereby \( E = \text{const.}, r_E = r_E \theta_s, \) and \( r_E \approx \text{const.} \). This works for nearly any loudspeaker layout.

While, to calculate loudspeaker gains, MDAP superimposes an arrangement of discrete virtual sources within a range of \( \pm \alpha \) around the panning direction \( \theta_s \), one could also think of superimposing a quasi-continuous distribution of virtual sources that are weighted by a continuous panning function \( g(\theta) \).

The ideal continuous panning function \( g(\theta) \) of axisymmetric directional spread around the panning direction \( \theta_s \) is described by \( g(\theta) = y_N^T(\theta) \text{diag}\{a_N\} y_N(\theta_s) \), the Ambisonic panning function. This rotation-invariant continuous function is optimal in terms of loudness, width, and localization measures, which are all evaluated by continuous integrals: \( E = \int g^2(\theta) \, d\theta = \text{const.} \) expresses panning-invariant loudness, \( r_E = \frac{1}{E} \int g^2(\theta) \theta \, d\theta = r_E \theta_s \) indicates a perfect alignment \( r_E \parallel \theta_s \) with the panning direction and a panning-invariant width \( r_E = \text{const.} \). However, the optimal values of

\(^1\)In detail, this follows from \( D^T D = V[I, 0][U^T U][I, 0]^T V^T = V[I, 0][I, 0]^T V^T = V V^T = I \).
these integrals are only preserved by discretization with optimal $t \geq 2N + 1$-design loudspeaker layouts.

**All-round Ambisonic decoding (AllRAD)** is preceded by the work of Batke and Keiler [16]. They describe Ambisonic panning $g_{\text{AllRAD}}(\theta) = D y_N(\theta)$ by a decoder $D$, whose result matches best with VBAP $g_{\text{VBAP}}(\theta)$. Without max-$r_E$ weights yet, we use this here to define AllRAD by the integral expressing a minimum-mean-square-error problem using the integral over all panning directions $\theta$

$$
\min_D \int_{S^2} \| g_{\text{VBAP}}(\theta) - D y_N(\theta) \|^2 \, d\theta.
$$

(4.45)

Equivalently, as described by Zotter and Frank [40] who coined the name, we may define AllRAD as VBAP synthesis on the physical loudspeakers when using as multiple-virtual-source inputs the Ambisonic panning function $g_{\text{AMBI}}(\theta) = y_N^T(\theta) \text{diag}[a_N] y_N(\theta_s)$ sampled at an optimal layout of virtual loudspeakers. Here, we write the synthesis as the integral over infinitely many virtual loudspeakers $\theta$,

$$
g = \int g_{\text{VBAP}}(\theta) g_{\text{AMBI}}(\theta) \, d\theta = \int g_{\text{VBAP}}(\theta) y_N^T(\theta) \text{diag}[a_N] y_N(\theta_s) \, d\theta
$$

$$
= \int g_{\text{VBAP}}(\theta) y_N^T(\theta) \, d\theta \text{ diag}[a_N] y_N(\theta_s).
$$

We can obviously pull the term $\text{diag}[a_N] y_N(\theta_s)$ out of the integral. The remaining integral defines the AllRAD matrix $D$. We may interpret it as a transformation of the VBAP loudspeaker gain functions $g_{\text{VBAP}}(\theta)$ into spherical harmonic coefficients. In the original paper [40], AllRAD is evaluated by an optimal layout of discrete virtual loudspeakers

$$
D = \int g_{\text{VBAP}}(\theta) y_N^T(\theta) \, d\theta = \frac{s_{d-1}}{L} \sum_{l=0}^{L} g_{\text{VBAP}}(\hat{\theta}_l) y_N^T(\hat{\theta}_l) = \frac{s_{d-1}}{L} \hat{G} \hat{Y}_N^T,
$$

(4.46)

using the directions $\{\hat{\theta}_l\}$ of a $t$-design. As VBAP’s gain functions aren’t smooth (derivatives are non-continuous), they are order-unlimited, and a $t$-design of sufficiently high $t$ should be used. In the 3D practice, the 5200 pts. Chebyshev-type design from [33] is dense enough. Note that the VBAP part permits improvements by insertion and downmix of imaginary loudspeakers to adapt to asymmetric or hemispherical layouts, as suggested in the original paper [40], cf. Sect. 3.3.

*Note that the decoder needs to be scaled properly. For instance, the norm of the omnidirectional component (first column) could be equalized to one, as it would typically be with a sampling decoder; there are alternative strategies to circumvent the scaling problem [41].*
4.9.7 EPAD and AllRAD on Sub-optimal Layouts

Figure 4.16 shows the improvement achieved with EPAD and AllRAD on an equiangular arrangement that is suboptimal by the missing loudspeaker at $-90^\circ$. Both decoders manage to handle either the loudness stabilization perfectly well (EPAD) or keep the directional and spread mapping errors small (AllRAD). We notice that for EPAD, with the constraint that $L \geq (2N + 1)$ just fulfilled for $N = 3$ and $L = 7$ of the simulation, it would not simply be possible to remove any further loudspeakers without degradation.

4.9.8 Decoding to Hemispherical 3D Loudspeaker Layouts

In typical loudspeaker playback situations for large audience, a solid floor and no loudspeakers below ear level are considered practical for several reasons. However, this does not permit decoding by sampling with optimal $t$-design layouts covering all directions. As shown above, EPAD and AllRAD do not require such arrays. And yet, they still require some care when used with hemispherical loudspeaker layouts, see [15, 40] for further reading.

**EPAD with hemispherical loudspeaker layouts.** Even for a hemispherical layout, the energy-preserving decoding method requires $L \geq (N + 1)^2$ loudspeakers to achieve a perfectly panning-invariant loudness. However, this is counter-intuitive: Why should one need at least as many loudspeakers on a hemisphere as are required for same-order playback on a full sphere? Shouldn’t the number be half as many?
Table 4.1 Integration ranges $0 \leq \vartheta \leq \vartheta_{\text{max}}$ to obtain $(N+1)(N+2)/2$ Slepian functions with minimum loudness fluctuation $\max_{\vartheta_{\text{min}} E} E$ for panning on the hemisphere

| $N$ | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  | 9  |
|-----|----|----|----|----|----|----|----|----|----|
| $\vartheta_{\text{max}}$ | $115^\circ$ | $114^\circ$ | $113^\circ$ | $111^\circ$ | $108^\circ$ | $106^\circ$ | $104^\circ$ | $102^\circ$ | $101^\circ$ |
| $\max_{\vartheta_{\text{min}} E}$ | 0.5 dB | 0.4 dB | 0.4 dB | 0.3 dB | 0.3 dB | 0.3 dB | 0.2 dB | 0.3 dB | 0.2 dB |

We can show that while the spherical harmonics are orthonormal on the sphere $S^2$, i.e. $\int_{S^2} y_N(\theta) y_N^T(\theta) \, d\theta = I$, they aren’t orthogonal on the hemisphere $S = S^2$:

$$\int_{S} y_N(\theta) y_N^T(\theta) \, d\theta = G. \quad (4.47)$$

Here, $G$ is called Gram matrix, and it is evaluated by $\frac{4\pi}{L} \sum_{l,\ell,\gamma,\beta \geq 0} y_N(\theta) y_N^T(\theta)$ using a high-enough $t$-design. By singular-value decomposition of the positive semi-definite matrix $G = Q \, \text{diag}(s) \, Q^T$, with $Q^T Q = \text{diag}(s) = \int_S Q^T y_N(\theta) y_N^T(\theta) Q \, d\theta$, we diagonalize $G$ and find new basis functions $\tilde{y}_N(\theta)$, the so-called Slepian functions [42], that are orthogonal on $S$

$$Q^T G Q = \text{diag}(s) = \int_S Q^T y_N(\theta) y_N^T(\theta) Q \, d\theta, \quad \Rightarrow \tilde{y}_N(\theta) = Q^T y_N(\theta).$$

Typically, the singular values in $s$ are sorted descendingly $s_1 \geq s_2 \geq \cdots \geq s_{(N+1)^2}$ so that it is possible to cut out basis functions of significantly large contribution to the upper hemisphere $S$ by

$$\tilde{y}_N(\theta) = [I, \, 0] \, Q^T y_N(\theta). \quad (4.48)$$

Typically, the numerical integral is extended to slightly below the horizon, see Table 4.1, so that truncation to the $(N+1)(N+2)/2$ most significant basis functions, see Fig. 4.17, produces a minimum fluctuation in the loudness measure $\tilde{E} = \|\tilde{y}_N(\theta)\|^2$ for panning on the hemisphere.

With $\tilde{y}_N(\theta)$, EPAD is calculated in the same way as for the ordinary harmonics

$$\tilde{Y}_N = \tilde{U} [\text{diag}(s), \, 0]^T \tilde{V}^T, \quad \tilde{D} = \tilde{U} [I, \, 0]^T \tilde{V}^T, \quad (4.49)$$

with the main difference that the lower limit for the number of loudspeakers decreases to $L \geq (N+1)(N+2)/2$. Interfaced to the spherical harmonics by $[I, \, 0] \, Q^T$, the hemispherical energy-preserving decoder becomes

$$D = \tilde{D} [I, \, 0] \, Q^T. \quad (4.50)$$
Fig. 4.17  Slepian basis functions for the upper hemisphere, composed of the spherical harmonics up to 3rd order, using $0^\circ \leq \theta \leq 113^\circ$ as integration interval and $(N + 1)(N + 2)/2$ functions. To get these nice shapes, the Slepian functions were found separately for every degree $m$ where they were moreover de-mixed by $QR$ decomposition

AllRAD with hemispherical loudspeaker layouts. Because of the vector-base amplitude panning involved, all-round Ambisonic decoding (AllRAD) is comparatively robust to irregular loudspeaker setups. Still, a hemispherical layout does not contain any loudspeaker direction vector pointing to the lower half space, therefore one could just omit information of the lower half space. However, the Ambisonic panning function implies a directional spread, so that panning to exactly the horizon also produces content below, whose omission causes: (i) a loss in loudness, (ii) a slight elevation of the perceived direction, cf. Fig. 4.18.

As discussed in the section on triangulation Sect. 3.3, the insertion of imaginary loudspeakers fixes this behavior. In the case of hemispherical loudspeaker layouts, it is not necessary to downmix the signal of the imaginary loudspeaker at nadir to stabilize both loudness and localization for panning to the horizon.

Signal contributions below but close to the horizon largely contribute to the horizontal loudspeakers, and it is therefore safe to dispose the signal that would feed the imaginary loudspeaker at nadir without loss of loudness. Moreover, this contribution from below also reinforces signals on the horizontal loudspeakers so that localization is pulled back down. Both can be observed in Fig. 4.18 that shows the loudness measure $E$ as well as mislocalization and width by the measure $r_E$ using max-$r_E$-weighted AllRAD with 5th-order Ambisonics along a vertical panning circle on the IEM mobile Ambisonics Array (mAmbA). It consists of 25 loudspeakers set up in rings of 8, 8, 4, 4, and 1 loudspeakers at 0, 20, 40, 60, 90 degrees elevation. Rings two and four start at 0 degree, the others are half-way rotated.

Performance comparison on hemispherical layouts. Figure 4.18 shows a comparison of AllRAD and EPAD decoding to the 25-channel mAmbA hemispherical loudspeaker layout.
Perceptual measures for 5th max-$r_E$-weighted AllRAD on the IEM mAmbA layout [43] with (black) and without insertion of the bottom imaginary loudspeaker (black dotted) whose signal is disposed, and max-$r_E$-weighted EPAD (gray), for panning on a vertical circle: $E$ in dB (top), orientation error of $r_E$ in degrees (middle), and width expressed by arccos $\|r_E\|$ in degrees (bottom). The thin dashed line shows AllRAD without imaginary loudspeakers.

While (top in Fig. 4.18) AllRAD produces a loudness fluctuation roughly spanning 1 dB for panning on the hemisphere, EPAD only exhibits 0.3 dB, as specified in Table 4.1. While in monophonic playback of noise, loudness differences of less than 0.5 dB can be heard, it is safe to assume that a weak directional loudness fluctuation of less than 1 dB is normally inaudible. In this regard, loudness fluctuation should be no problem with both EPAD and AllRAD.

Concerning the directional mapping, EPAD produces a more strongly pronounced ripple, with $r_E$ indicating sounds on the horizon $\vartheta_s = \pm 90^\circ$ to be pulled upwards towards 0° more with EPAD ($7^\circ$) than with AllRAD ($3^\circ$). In terms of width, both EPAD and AllRAD exhibit the $\approx 20^\circ$ average associated with max-$r_E$ weighting. However, EPAD also produces a greater fluctuation, and it widens up to about 30° degree for panning to the horizon $\vartheta_s = \pm 90^\circ$.

With the 9 loudspeakers of the ITU [44] $4 + 5 + 0$ layout (horizontal ring: $\varphi = 0, \pm 30^\circ, \pm 120^\circ$, upper ring at 40° elevation with $\varphi = \pm 30^\circ, \pm 120^\circ$), it is not possible anymore to use EPAD with 5th order, which would be the optimal resolution for the front loudspeaker triplet. EPAD only supports orders up to $N = 2$, and to lose level towards below-horizon directions, we can use the reduced set of 6 Slepian functions; alternatively all 9 spherical harmonics of $N = 2$ would also be thinkable. For AllRAD, imaginary loudspeakers are inserted at the sides at azimuth/elevation $\pm 75^\circ/27^\circ$, up at 0°/78°, back 180°/35°, and below 0°/−90°. It is reasonable to downmix the imaginary loudspeakers with a factor one for up, sides, back, and re-normalize the VBAP gain matrix, while disposing the signal of the imaginary loudspeaker below. AllRAD permits to use the order $N = 5$, which resolves the frontal loudspeaker triplet much better for horizontal panning.

Figure 4.19 shows the result of max-$r_E$-weighted 2nd-order EPAD and 5th-order AllRAD for the $4 + 5 + 0$ layout using a vertical panning curve. While the perfectly constant loudness measure of EPAD might be favored over the almost $+3$ dB loudness
increase of front and back for AllRAD, AllRAD’s lower directional error, narrower width mapping, greater flexibility, and simplicity has often proven to be clearly superior in practice.

### 4.10 Practical Studio/Sound Reinforcement Application Examples

This section analyzes the application of 3D Ambisonic amplitude panning consisting of encoding and AllRAD to studio (with typical setups of 2 m radius) and sound reinforcement applications (for an audience of, e.g., 250 people). Application scenarios are sketched in [43], and various other examples are given below. Requirements of a constant loudness and width are analyzed below, and as sound reinforcement requires a particularly large sweet area, the $r_E$ vector model for off-center listening positions from Sect. 2.2.9 is used to depict the sweet area size.

The analysis of decoders above described loudness measures for panning on a circle. To observe them with panning across all directions in Figs. 4.20 and 4.22, world-map-like mappings using a gray-scale representation of the loudness and width measures are more reasonable. For several loudspeaker layouts, its axes are azimuth horizontally and zenith vertically, and the gray-scale map displays the loudness measure $E$ in dB (left column) and the width measure $\arccos\|r_E\|$ in degrees (right column). As 5th-order max-$r_E$-weighted AllRAD typically produces minor directional mapping errors, they aren’t explicitly shown in Figs. 4.20 and 4.22. However, the mappings of the sweet area size of plausible localization in Figs. 4.21 and 4.23 illustrate the usefulness of the systems for the listening areas hosting the number of listeners targeted for either the studio or the sound reinforcement application.

Figure 4.20 illustrates AllRAD’s tendency of attenuated signals in too closely spaced loudspeaker ensembles as in the front section of the ITU [44] 4 + 5 + 0. By
Ambisonic Amplitude Panning and Decoding in Higher Orders

Fig. 4.20 Comparison of 5th-order max-$r_E$-weighted AllRAD for panning across all directions on hemispherical loudspeaker layouts in studios. The left column the loudness measure $E$ in dB and the right-most column the width measure $\arccos ||r_E||$ in degree, and the loudspeaker position are marked with a white + sign.

Fig. 4.21 Comparison of the calculated sweet area size for 5th-order max-$r_E$-weighted AllRAD for panning across all directions on hemispherical loudspeaker layouts in studios. As a plausibility definition, the directional mapping errors depending on the listening position should stay within angular bounds (e.g. $10^\circ$).

contrast, for instance the mAmbA layout in Fig. 4.22 only has 8 loudspeakers on the horizon, and signals panned to the largely spaced below-horizon triangles tend to get louder. Moreover, it is easier for loudspeaker systems of many channels such as IEM CUBE, mAmbA, Lobby, and Ligeti Hall in Fig. 4.22 to yield smooth loudness and width mappings. Still, also with only a few loudspeakers, slight direction adjustment in the layout can fix some of the behavior, as with the IEM Production Studio, whose $\pm 45^\circ$ loudspeakers in the elevated layer is superior to a $\pm 30^\circ$ spacing.
Fig. 4.22 Comparison of 5th-order max-$r_E$-weighted AllRAD for panning across all directions on various hemispherical loudspeaker layouts for sound reinforcement. The left column the loudness measure $E$ in dB and the right-most column the width measure $\arccos \|r_E\|$ in degree, and the loudspeaker position are marked with a white + sign.

A hint for designing good decoders sometimes is idealization: often it is better to disregard the true loudspeaker setup locations and feed the decoder design with idealized positions instead. Hereby can one trade slight directional distortions for a more uniform loudness distribution. For instance at the IEM CUBE, loudspeaker
Fig. 4.23 Comparison of the calculated sweet area size for 5th-order $max-r_E$-weighted AllRAD for panning across all directions on various hemispherical loudspeaker layouts for sound reinforcement. As a plausibility definition, the directional mapping errors depending on the listening position should stay within angular bounds (e.g. $10^{\circ}$)
locations of the horizontal ring could be idealized to $30^\circ$ to get a smoother loudness mapping as the one shown in Fig. 4.22.

4.11 Ambisonic Decoding to Headphones

Typically, Ambisonic decoding to headphones can be done similarly as with loudspeakers, except that the loudspeaker signals are rendered to headphones by convolution with the head-related impulse responses (HRIRs) of the corresponding playback directions. Various databases of such HRIRs can be found, e.g., on the website SOFA-conventions. This headphone decoding approach is classically using a small set of so-called virtual loudspeakers, as it is found in many places in technical literature, e.g. in the pioneering works of Jean-Marc Jot et al. [9] or Jérôme Daniel [10]. It is relevant in many important other works [18, 45, 46], the SADIE project, and it is employed in Sect. 1.4.2 on first-order Ambisonics.

Coarse. However, as outlined in some research papers [9, 18, 46], these approaches have in common that low-order Ambisonic synthesis is problematic. It can either happen when inserting a dense grid of virtual-loudspeaker HRIRs that the Ambisonic smoothing attenuates high-frequency at frontal and dorsal directions. Or, what had been the solution for a long time, a coarse grid of virtual-loudspeaker HRIRs does not attenuate high frequencies, but still yields that spatial quality strongly depends on the particular grid layout or orientation [46]. An early paper by Jot [9] proposed to remove the time delays of the HRIR before Ambisonic decomposition, and then to re-insert the otherwise missing interaural time-delay afterwards, for any sound panned in Ambisonics, which unfortunately yields an object-based panning system rather than a scene-based Ambisonic system.

Dense. Some dense-grid approaches propose to keep the HRIR time delays, or if formulated in the frequency domain: the HRTF phases (head-related transfer function), and hereby stay in a scene-based Ambisonic format, while correcting spectral deficiencies by diffuse-field or interaural-covariance equalization [18, 47]. Finally, most recent solutions proposed by Jin, Sun, and Epain, [17, 48] or Zaunschirm, Schörkhuber, and Höldrich [20, 21] modify the HRIR time delays/HRTF phases but only above, e.g., 3 kHz, without any object-based re-insertion afterwards. The omission of high-frequency interaural time-delay/phase information is a reasonable trade off done in favor of a more important accuracy in spectral magnitude.

What does directional HRIR smoothing do to high frequencies? The geometrical theory of diffraction [49] suggests that HRIRs must always contain at least the delay to the ear of either the shortest direct path or the shortest indirect path via the surface of the head. For a spherical head model with the radius $R = 0.0875$ m and speed of sound $c = 343 \, \frac{m}{s}$, the Woodworth-Schlosberg formula [50] is composed of this

---

2 https://www.sofaconventions.org.
3 https://www.york.ac.uk/sadie-project/database.html.
Fig. 4.24  Rigid-sphere model of delay to ear depending on angle $\phi$ (here: head rotation)

![Diagram](image)

Fig. 4.25  Time delay to the ear depending on azimuth of horizontal sound (a) and (b) $360^\circ$-measured KU100 dummy head HRIRs set from TH Köln (color displays dB levels)

consideration, see Fig. 4.24. The left ear receives a distant horizontal sound from the azimuth interval $0 \leq \phi \leq \frac{\pi}{2}$ as direct sound anticipated by $\tau = -\frac{R}{c} \sin \phi$, or for $-\frac{\pi}{2} < \phi \leq 0$ as an indirect sound delayed by $\tau = -\frac{R}{c} \phi$,  

$$\tau(\phi) = -\frac{R}{c} \min \{\sin \phi, \phi\}, \quad (4.51)$$

as plotted in Fig. 4.25a, and recognizable from dummy-head measurements\(^4\) in Fig. 4.25b.

If the HRIR is smoothed across an angular range, the time-delay curve gets spread across time as well, see Fig. 4.26. In this way, depending on whether the smoothing uses a continuous or discrete set of directions, one either obtains something like a comb filter or a sinc-shaped frequency response. This smoothing is least disturbing around the direct-ear side as shown left in Fig. 4.26, and, as the indirect ear also encounters high-frequency shadowing effects, it is most disturbing mainly for frontal and rear sounds at $0^\circ$ or $180^\circ$, as shown right in Fig. 4.26. The corresponding frequency responses are roughly exemplified with what third-order Ambisonics equivalent smoothing would do to either $45^\circ$-spaced HRIRs in Fig. 4.27a or $15^\circ$-spaced ones in Fig. 4.27b.

\(^4\)Data HRIR_CIRC360.sofa from http://sofacoustics.org/data/database/thk.
4.11 Ambisonic Decoding to Headphones

Fig. 4.26 Directionally smoothed playback to multiple HRIRs within a window, e.g. $30^\circ$, causes different impulse response shapes at $90^\circ$ and $0^\circ$.

Fig. 4.27 Differences of directionally smoothed HRTF frequency responses for a horizontal sound from either $0^\circ$ and $90^\circ$, smoothed within a $\pm22.5^\circ$ window, which roughly corresponds to the Ambisonics order $N = 3$; a and b either use a grid of $45^\circ$ or $7.5^\circ$ spaced HRIRs. The dashed line shows the theoretical frequency limit $1.87 \, \text{kHz}$ for $N = 3$.

To get an upper frequency limit, it is insightful to work in the frequency domain where the HRIR is denoted head-related transfer function (HRTF). A simplified linearized-phase version around $\phi = 0$ uses $\tau \approx \frac{R}{c} \phi$, and the resulting in the Fourier transform with $\omega = 2\pi f$ is

$$H \approx e^{-i\frac{\omega}{c} R \phi} = e^{i\frac{\omega}{c} R \phi}.$$ (4.52)

To represent it by circular or spherical harmonics transformation limited to the order $N$, a maximum phase change represented by the harmonic $e^{iN\phi}$ implies that we can only resolve the phase up to $\frac{\omega}{c} R \leq N$, hence the range of accurate operation is limited in frequency

$$f_N \leq \frac{c}{2\pi R} N = N \cdot 624 \, \text{Hz}.$$ (4.53)

As high-frequency HRTF phase evolves more rapidly over the angle as what the finite order can represent, this typically yields attenuation of the high frequencies when obtaining circular/spherical harmonics coefficients by transformation integral.

Directional smoothing of the discrete directional HRTFs causes relevant spectral problems, regardless of whether directional smoothing is done by Ambisonics, VBAP, MDAP. Mainly the geometric delay in the HRIRs is responsible for the emerging comb-filter or low-pass behavior. One could pull out the linear phase trend above the frequency limit and re-insert it, but is re-insertion necessary?
4.11.1 High-Frequency Time-Aligned Binaural Decoding (TAC)

As a pre-requisite for their binaural Ambisonic decoders, Schörkhuber et al. [21] tested, above which frequency the removal of the HRTF linear phase trend remains inaudible in direct HRTF-based rendering without panning or smoothing. In fact, most of their listeners could not distinguish the absence of the linear phase trend when removed above 3 kHz for various sound examples (drums, speech, pink noise, rendered at directions 10°, −45°, 80°, −130°). They had their subjects compare the result to a reference with unaltered HRTFs, and the result is analyzed in Fig. 4.28.

By this finding, it is possible to split up each of the $2 \times 1$ HRIRs $h(t, \theta)$ into an unaltered low-pass band and a time-aligned high-pass band to unify the high-frequency HRIR delay

$$\hat{h}(t, \theta) = h_{\leq 3kHZ}(t, \theta) + \left[ h_{\text{left} > 3kHZ}[t - \tau(\arcsin(\theta_y), \theta)] \right].$$

(4.54)

The time delay model $\tau(\phi)$ uses the angle to the left/right ear on the positive/negative y axis, so arccos $\pm \theta_y$, but shifted by 90°, hence $\phi = \pm \arcsin(\theta_y)$.

This removal allows use all available HRIRs of dense measurement sets for binaural synthesis of high accuracy, using a suitable linear Ambisonic decoder such as AllRAD. Assuming the resulting modified left and right HRIR for all directions are denoted as $2 \times L$ matrix $\hat{H}(t) = [\hat{h}(t, \theta_1), \ldots, \hat{h}(t, \theta_L)]^T$, the $2 \times (N + 1)^2$ filter set for decoding every of the Ambisonic channels to the ears becomes:

$$\hat{H}_{SH}(t) = \hat{H}(t) \ D\ \text{diag}\{a\}.$$  

(4.55)

![Fig. 4.28](image-url)  

Experiment on audibility of removal of the linear phase trend from HRTF above a varied cutoff frequency from [21] showing medians and 95% confidence intervals.
Results achieved by a pseudo-inverse decoding to hereby time-aligned HRIRs using $R = 0.085\, \text{cm}$ with $N = 3$ from the 2702-directions Cologne HRIRs\(^5\) is shown in Fig. 4.29. The resulting polar patterns (ta) clearly outperform the linear decomposition (lin) at frequencies above 2kHz in representing the original HRTFs (max).

### 4.11.2 Magnitude Least Squares (MagLS)

Alternative to high-frequency time delay disposal, Schörkhuber et al. present an optimum-phase approach [21] that disregards phase match in favor of an improved magnitude match above cutoff. Formulated exemplarily for the left ear, across every HRTF direction $\theta_l$, and for every discrete frequency $\omega_k$, with $h_{l, k} = h(\theta_l, \omega_k)$, this becomes

\[
\min_{\hat{h}_{SH, k}} \sum_{l=1}^{L} \left[ |y_{N}(\theta_l)^\top \hat{h}_{SH, k} - |h_{l, k}| \right]^2. \tag{4.56}
\]

Typically, one would need to solve magnitude least squares or magnitude squares least squares tasks with semidefinite relaxation, see Kassakian [51].

In practice, however, results turn out to be perfect already with an iterative combination of the reconstructed phase $\hat{\phi}_{l,k-1}$ from the previous frequency $\omega_{k-1}$ with the HRTF magnitude $|h_{k,l}|$ of the current frequency $\omega_k$, before a linear decomposition thereof into spherical harmonic coefficients $\hat{h}_{SH,k}$.

Every frequency below cutoff $\omega_k < 2\pi f_N$ just uses the linear least-squares spherical harmonics decomposition with the left-inverse of the spherical harmonics $Y_N$ sampled at the HRTF measurement nodes,

---

\(^5\)Data HRIR_L2702.sofa from [http://sofacoustics.org/data/database/thk](http://sofacoustics.org/data/database/thk).
\[ \hat{h}_{\text{SH},k} = (Y_N^T Y_N)^{-1} Y_N^T [h_{l,k}]_l. \] (4.57)

Continuing with the first frequency above/equal to cutoff \( \omega_k \geq 2\pi f_N \), the algorithm proceeds as:

\[
\hat{\phi}_{l,k-1} = \angle \left\{ y_N(\theta_l)^T \hat{h}_{\text{SH},k-1} \right\}, \tag{4.58}
\]

\[
\hat{h}_{\text{SH},k} = (Y_N^T Y_N)^{-1} Y_N^T [|h_{l,k}| e^{i\hat{\phi}_{l,k-1}}]_l, \tag{4.59}
\]

and then moves to the next frequency \( k \leftarrow k + 1 \). The results are typically transformed back to time domain to get a real-valued impulse response for every spherical harmonic to the regarded ear.

The results of the MagLS approach (mls) outperform the time-alignment approach (ta) in the exemplary results shown for \( N = 3 \) in Fig. 4.29, in particular at the highest frequencies, where sphere-model-based delay simplification is not sufficiently helpful, anymore.

### 4.11.3 Diffuse-Field Covariance Constraint

Also for both the above approaches that modify the high-frequency phase, Zaunschirm et al. [20] note that low order rendering degrades envelopment in diffuse fields, so that they introduce an additional covariance constraint as defined by Vilkamo [22]. It can be implemented as a \( 2 \times 2 \) filter matrix equalizing the resulting frequency-domain diffuse-field covariance matrix to the one of the original HRTF datasets. On the main diagonal, this covariance matrix shows the diffuse-field ear sensitivities (left and right), and off-diagonal it contains the diffuse-field inter-aural cross correlation.

At every frequency, the \( 2 \times 2 \) diffuse-field covariance matrix of the original, very-high-order spherical harmonics HRTF dataset \( H^H_{\text{SH}} \) of the dimensions \( 2 \times (M + 1)^2 \) with \( M \gg N \) is given by

\[
R = H^H_{\text{SH}} H_{\text{SH}}. \tag{4.60}
\]

The derivation why this inner product of spherical harmonic coefficients represents the diffuse-field covariance is given in Appendix A.5. The low-order high-frequency modified HRTF coefficient set \( \hat{H}_{\text{SH}} \) of the dimensions \( 2 \times (N + 1)^2 \) also has a \( 2 \times 2 \) covariance matrix \( \hat{R} \) that will differ from the more accurate \( R \),

\[
\hat{R} = \hat{H}^H_{\text{SH}} \hat{H}_{\text{SH}}. \tag{4.61}
\]

Its diffuse-field reproduction improves after equalizing \( R = \hat{R} \) by a \( 2 \times 2 \) filter matrix,
Fig. 4.30 Covariance constraint filters enhance the binaural decorrelation of MagLS by negative crosstalk $M_{12}$ and $M_{21}$, under corresponding correction of the diffuse-field sensitivities $M_{11}$ and $M_{22}$ at playback orders $N < 3$

$$\hat{H}_{SH,\text{corr}} = \hat{H}_{SH}M.$$ (4.62)

Appendix A.5 shows the derivation of $M$ based on [20, 22]. In summary, it is composed of factors obtained by Cholesky and SVD matrix decompositions

$$\hat{H}_{\text{corr,SH}} = \hat{H}_{SH}^{-1}VU^HX,$$ (4.63)

Cholesky factors: SVD:

$$H_{SH}^HH_{SH} = X^HX,$$ $$\hat{X}^HX = USV^H,$$

$$\hat{H}_{SH}^HH_{SH} = \hat{X}^H\hat{X}.$$  

While MagLS binaural decoding with orders higher than 2 or 3 does not require covariance correction, the correction enhances the decorrelation of the ear signals for 1st to 2nd order reproduction, as shown in Fig. 4.30.

### 4.12 Practical Free-Software Examples

#### 4.12.1 Pd and Circular/Spherical Harmonics

Similar as in the example section on first-order encoding and decoding in pure data (Pd), Fig. 4.31 shows 3rd-order 2D Ambisonic encoding and decoding for an octagon loudspeaker layout. The implementation [mtx_circular_harmonics] of the circular harmonics is used from the iemmatrix library, and the numbers for $\frac{180}{\pi} = 57.29$ and $a_m = \cos\left(\frac{\pi m}{2(N+1)}\right)$ were pre-calculated. Note the similarity to the first-order 2D example of Fig. 1.13, to which the main change is the use of the circular harmonics matrix object.

For decoding to headphones, programming in Pd also looks rather similar as in the first-order example in Fig. 1.14, only more HRIRs matching the respective
loudspeaker positions need to be employed. To work in 3 dimensions, programming in Pd would also be similar as in the corresponding first-order example of Fig. 1.15, using the matrix object \([\text{mtx}_\text{spherical}_\text{harmonics}]\). Typically, pre-calculated decoders including AllRAD and max-\(r_E\) are used and loaded by, e.g., \([\text{mtx D.mtx}]\) into Pd to keep programming simple.

4.12.2 Ambix Encoder, IEM MultiEncoder, and IEM AllRADecoder

For encoding single- or multi-channel signals into Ambisonics, there are the \text{ambix}\text{\_encode\text{\_}o<\text{\_}N>}, or \text{ambix}\text{\_encode\text{\_}i<\text{\_}L>\text{\_}o<\text{\_}N>} VST plugins available from Kronlachner’s ambix plugin suite or the IEM MultiEncoder from the IEM plugin suite. As exemplarily shown in Fig. 4.32, the multi encoder allows to encode channel-based multi-channel audio material, where \text{channel-based} \[52\] typically refers to each channel of the multi-channel material meant to be played back on a separate loudspeaker of clearly defined direction, cf. \[44\]. Elsewhere, the embedding of virtual playback directions can also be found referred to as \text{beds} or \text{virtual panning spots}.
4.12 Practical Free-Software Examples

Fig. 4.32 MultiEncoder plug-in: encoding of a $4 + 5 + 0$ recording

Fig. 4.33 AllRADecoder plug-in: $5 + 7 + 0$ layout from IEM Production Studio
Fig. 4.34  AllRADecoder plug-in: $5 + 7 + 0$ layout from IEM Production Studio

Fig. 4.35  AllRADecoder plug-in: $5 + 7 + 0$ layout from IEM Production Studio
The IEM AllRADecoder permits to manually enter or import the loudspeaker coordinates and channel indices, with the coordinates specified by the azimuth and elevation angle in degrees, as exemplified for the IEM production studio in Fig. 4.33. The figure also shows that just entering the pure $5 + 7 + 0$ layout would produce an error message *Point of origin not within convex hull. Try adding imaginary loudspeakers.*

By adding an imaginary loudspeaker below whose signal is typically omitted, see Fig. 4.34, it becomes geometrically valid to calculate and employ the resulting decoder, however it is better to also insert an imaginary loudspeaker at the rear whose signal is preserved by specifying the gain value 1, as shown in Fig. 4.35.

### 4.12.3 *Reaper, IEM RoomEncoder, and IEM BinauralDecoder*

Particularly relevant for head-phone-based listening, rendering of anechoic sounds will typically not *externalize* well, as it does not match the mental expectation of ordinary listening environments [53–56]. To avoid that this would rather cause an in-head localization than the desired external sound image, one can, e.g., use the IEM *RoomEncoder* plugin, see Fig. 4.36. It is based on an image-source room model.

![RoomEncoder plug-in](image)
and encodes first-order wall-reflections involving reflection factors and propagation delays together with the desired direct sound.

The MagLS approach for Ambisonic decoding, using the KU100 measurements from Cologne Applied Science University and (optionally) their headphone equalization curves is implemented by the IEM BinauralDecoder, see Fig. 4.37.

In combination of both, IEM RoomEncoder and IEM BinauralDecoder with an Ambisonics-encoded single-channel sound (e.g. using ambix_encoder), one can simply try to place the source and receiver together in the symmetry plane of the room, and then to slightly shift one of both sideways to see how externalization improves by slight asymmetry in the ear signals.

References

1. J.S. Bamford, Ambisonic sound for the masses (1994)
2. D.H. Cooper, T. Shiga, Discrete-matrix multichannel stereo. J. Audio Eng. Soc. 20(5), 346–360 (1972)
3. P. Felgett, Ambisonic reproduction of directionality in surround-sound systems. Nature 252, 534–538 (1974)
4. M.A. Gerzon, The design of precisely coincident microphone arrays for stereo and surround sound, in prep: L-20 of 50th Audio Engineering Society Convention (1975)
5. P. Craven, M.A. Gerzon, Coincident microphone simulation covering three dimensional space and yielding various directional outputs, U.S. Patent, no. 4,042,779 (1977)
6. J.S. Bamford, An analysis of ambisonic sound systems of first and second order, Master’s thesis, University of Waterloo, Ontario (1995)
7. D.G. Malham, A. Myatt, 3D sound spatialization using ambisonic techniques. Comput. Music J. 19(4), 58–70 (1995)
8. M.A. Poletti, The design of encoding functions for stereophonic and polyphonic sound systems. J. Audio Eng. Soc. 44(11), 948–963 (1996)
9. J.-M. Jot, V. Larcher, J.-M. Pernaux, A comparative study of 3-d audio encoding and rendering techniques, in 16th AES Conference (Rovaniemi, 1999)
10. J. Daniel, Représentation des champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia, Ph.D. dissertation, Université Paris 6 (2001)
11. D.B. Ward, T.D. Abhayapala, Reproduction of a plane-wave sound field using an array of loudspeakers. IEEE Trans. Speech Audio Process. 9(6), 697–707 (2001)
12. G. Dickins, Sound field representation, reconstruction and perception, Master’s thesis, Australian National University, Canberra (2003)
13. A. Sontacchi, Neue Ansätze der Schallfeldreproduktion, Ph.D. dissertation, TU Graz (2003)
14. M.A. Poletti, Robust two-dimensional surround sound reproduction for nonuniform loudspeaker layouts. J. Audio Eng. Soc. 55(7/8), 598–610 (2007)
15. F. Zotter, H. Pomerantz, M. Noisternig, Energy-preserving ambisonic decoding. Acta Acust. United Acust. 98(11), 37–47 (2012)
16. J.-M. Batke, F. Keiler, Using vbap-derived panning functions for 3d ambisonics decoding, in 2nd International Symposium on Ambisonics and Spherical Acoustics (Paris, 2010)
17. D. Sun, Generation and perception of three-dimensional sound fields using higher order ambisonics (University of Sydney, School of Electrical and Information Engineering, 2013). Ph.D thesis
18. Z. Ben-Hur, F. Brinkmann, J. Sheaffer, S. Weinzierl, B. Rafaely, Spectral equalization in binaural signals represented by order-truncated spherical harmonics. J. Acoust. Soc. Am. 141(6) (2017)
19. F. Brinkmann, S. Weinzierl, Comparison of head-related transfer functions pre-processing techniques for spherical harmonics decomposition, in AES Conference AVAR (Redmont, 2018)
20. M. Zaunschirm, C. Schörkhuber, R. Höldrich, Binaural rendering of ambisonic signals by head-related impulse response time alignment and a diffuseness constraint. J. Acoust. Soc. Am. 143(6), 3616–3627 (2018)
21. C. Schörkhuber, M. Zaunschirm, R. Höldrich, Binaural rendering of ambisonic signals via magnitude least squares, in Fortschrritte der Akustik - DAGA (Munich, 2018)
22. J. Vilkamo, T. Bäckström, A. Kuntz, Optimized covariance domain framework for time-frequency processing of spatial audio. J. Audio Eng. Soc. 61(6), 403–411 (2013)
23. F.W.J. Olver, R.F. Boisvert, C.W. Clark (eds.), NIST Handbook of Mathematical Functions (Cambridge University Press, Cambridge, 2000), http://dlmf.nist.gov. Accessed June 2012
24. J. Daniel, J.-B. Rault, J.-D. Polack, Acoustic properties and perceptive implications of stereophonic phenomena, in AES 6th International Conference: Spatial Sound Reproduction (1999)
25. M. Frank, How to make ambisonics sound good, in Forum Acusticum, Krakow (2014)
26. M. Frank, F. Zotter, Extension of the generalized tangent law for multiple loudspeakers, in Fortschrritte der Akustik - DAGA (Kiel, 2017)
27. M. Frank, A. Sontacchi, F. Zotter, Localization experiments using different 2d ambisonic decoders, in 25th Tonmeistertagung (2008)
28. P. Stitt, S. Bertet, M.v. Walstijn, Off-centre localisation performance of ambisonics and hoa for large and small loudspeaker array radii. Acta Acust. United Acust. 100(5), 937–944 (2014)
29. M. Frank, F. Zotter, Exploring the perceptual sweet area in ambisonics, in AES 142nd Convention (2017)
30. ISO 31-11:1978, Mathematical signs and symbols for use in physical sciences and technology (1978)
31. ISO 80000-2, quantities and units? Part 2: Mathematical signs and symbols to be used in the natural sciences and technology (2009)
32. R.H. Hardin, N.J.A. Sloane, McLaren’s improved snub cube and other new spherical designs in three dimensions. Discret. Comput. Geom. 15, 429–441 (1996), http://neilsloane.com/sphdesigns/dim3/
33. M. Gräf, D. Potts, On the computation of spherical designs by a new optimization approach based on fast spherical fourier transforms. Numer. Math. 119 (2011), http://homepage.univie.ac.at/manuel.grae/f/quadature.php
34. R.S. Womersley, Chapter, Efficient spherical designs with good geometric properties in Contemporary Computational Mathematics - A Celebration of the 80th Birthday of Ian Sloan (Springer, Berlin, 2018), pp. 1243–1285
35. A. Solvang, Spectral impairment of 2d higher-order ambisonics. J. Audio Eng. Soc. 56(4) (2008)
36. M.A. Gerzon, G.J. Barton, Ambisonic decoders for HDTV, in prepr. 3345, 92nd AES Convention, Vienna, 1992
37. J. Daniel, J.-B. Rault, J.-D. Polack, Ambisonics encoding of other audio formats for multiple listening conditions, in prepr. 4795, 105th AES Convention (San Francisco, 1998)
38. M.A. Poletti, A unified theory of horizontal holographic sound systems. J. Audio Eng. Soc. 48(12), 1155–1182 (2000)
39. M.A. Poletti, Three-dimensional surround sound systems based on spherical harmonics (J. Audio Eng. Soc, 2005)
40. F. Zotter, M. Frank, All-round ambisonic panning and decoding (J. Audio Eng. Soc, 2012)
41. F. Zotter, M. Frank, Ambisonic decoding with panning-invariant loudness on small layouts (allrad2), in 144th AES Convention, prepr. 9943 (Milano, 2018)
42. R. Pail, G. Plank, W.-D. Schuh, Spatially restricted data distributions on the sphere: the method of orthonormalized functions and applications. J. Geodesy 75, 44–56 (2001)
43. M. Frank, A. Sontacchi, Case study on ambisonics for multi-venue and multi-target concerts and broadcasts. J. Audio Eng. Soc. 65(9) (2017)
44. ITU, Recommendation BS.2051: Advanced sound system for programme production. ITU (2018)
45. M. Noisternig, A. Sontacchi, T. Musil, R. Höldrich, “A 3d ambisonic based binaural sound reproduction system, in 24th AES Conference (Banff, 2003)
46. B. Bernschütz, A. V. Giner, C. Pörschmann, J. Arend, Binaural reproduction of plane waves with reduced modal order. Acta Acust. United Acust. 100(5) (2014)
47. S. Delikaris-Manias, J. Vilkamo, Adaptive mixing of excessively directive and robust beamformers for reproduction of spatial sound, in Parametric Time-Frequency Domain Spatial Audio, ed. by V. Pulikki, S. Delikaris-Manias, A. Politis (Wiley, New Jersey, 2017)
48. C.T. Jin, N. Epain, D. Sun, Perceptually motivated binaural rendering of higher order ambisonic sound scenes, in Fortschrritte der Akustik AIA-DAGA (Merano, 2013). March
49. J.B. Keller, Geometrical theory of diffraction. J. Acoust. Soc. Am. 52(2), 116–130 (1962)
50. R.S. Woodworth, H. Schlosberg, Experimental Psychology (Holt, Rinehart and Winston, 1954)
51. P.W. Kassakian, Convex approximation and optimization with applications in magnitude filter design and radiation pattern synthesis, Ph.D. dissertation, EECS Department, University of California, Berkeley (2006)
52. ITU, Recommendation BS.2076: Audio Definition Model (ITU, 2017)
53. S. Werner, G. Götz, F. Klein, Influence of head tracking on the externalization of auditory events at divergence between synthesized andf listening room using a binaural headphone system, in prepr. 9690, 142nd AES Convention (Berlin, 2017)
54. F. Klein, S. Werner, T. Mayenfels, Influences of tracking on externalization of binaural synthesis in situations of room divergence. J. Audio Eng. Soc. 65(3), 178–187 (2017)
55. J. Cubick, Investigating distance perception, externalization and speech intelligibility in complex acoustic environments, Ph.D. dissertation, DTU Copenhagen (2017)
56. G. Plenge, Über das Problem der Im-Kopf-Lokalisation. Acustica 26(5), 241–252 (1972)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.