

Chapter 6

Higher-Order Ambisonic Microphones and the Wave Equation (Linear, Lossless)



... a turning point has been the design of HOA microphones, opening an exciting experimental field in terms of real 3D sound field recording ...

Jérôme Daniel [1] at Ambisonics Symposium 2009.

Abstract Unlike pressure-gradient transducers, single-transducer microphones with higher-order directivity apparently turned out to be difficult to manufacture at reasonable audio quality. Therefore nowadays, higher-order Ambisonic recording with compact devices is based on compact spherical arrays of pressure transducers. To prepare for higher-order Ambisonic recording based on arrays, we first need a model of the sound pressure that the individual transducers of such an array would receive in an arbitrary surrounding sound field. The lossless, linear wave equation is the most suitable model to describe how sound propagates when the sound field is composed of surrounding sound sources. Fundamentally, the wave equation models sound propagation by how small packages of air react (i) when being expanded or compressed by a change of the internal pressure, and to (ii) directional differences in the outside pressure by starting to move. Based there upon, the inhomogeneous solutions of the wave equation describe how an entire free sound field builds up if being excited by an omnidirectional sound source, as a simplified model of an arbitrary physical source, such as a loudspeaker, human talker, or musical instrument. After addressing these basics, the chapter shows a way to get Ambisonic signals of high spatial and timbral quality from the array signals, considering the necessary diffuse-field equalization, side-lobe suppression, and trade off between spatial resolution and low-frequency noise boost. The chapter concludes with application examples.

Gary Elko and Jens Meyer are the well-known inventors of the first commercially available compact spherical microphone array that is able to record higher-order Ambisonics [2], the Eigenmike. There are several inspiring scientific works with

valuable contributions that can be recommended for further reading [3–12], above all Boaz Rafaely’s excellent introductory book [13].

This mathematical theory might appear extensive, but it cannot be avoided when aiming at an in-depth understanding of higher-order Ambisonic microphones. The theory enables processing of the microphone signals received such that the surrounding sound field excitation is retrieved in terms of an Ambisonic signal. Some readers may want to skip the physical introduction and resume in Sect. 6.5 on spherical scattering or Sect. 6.6 on the processing of the array signals.

6.1 Equation of Compression

Wave propagation involves *reversible* short-term temperature fluctuations becoming effective when air is being compressed by sound, causing the specific stiffness of air in sound propagation. The Appendix A.6.1 shows how to derive this adiabatic compression relation based on the first law of thermodynamics and the ideal gas law. It relates the relative volume change $\frac{V}{V_0}$ to the pressure change $p = -K \frac{V}{V_0}$ by the bulk modulus of air. After expressing the bulk modulus by more common constants¹ $K = \rho c^2$ and differentially formulating the volume change over time using the change of the sound particle velocity in space, e.g. in one dimension $\dot{p} = -\rho c^2 \frac{\partial v_x}{\partial x}$, cf. Appendix A.6.1, we get the three-dimensional compression equation:

$$\frac{\partial p}{\partial t} = -\rho c^2 \nabla^T \mathbf{v}. \quad (6.1)$$

Here the inner product of the Del symbol $\nabla^T = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z})$ with \mathbf{v} yields what is called divergence $\text{div}(\mathbf{v}) = \nabla^T \mathbf{v} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}$. The equation means: *Independently of whether the outer boundaries of a small package of air are traveling at a common velocity: If there are directions into which their velocity is spatially increasing, the resulting gradual volume expansion over time causes a proportional decrease of interior pressure over time.*

6.2 Equation of Motion

The equation of motion is relatively simple to understand from the Newtonian equation of motion, e.g. for the x direction, $F_x = m \frac{\partial v_x}{\partial t}$ equates the external force to mass m times acceleration, i.e. increase in velocity $\frac{\partial v}{\partial t}$. For a small package of air with constant volume $V_0 = \Delta x \Delta y \Delta z$, the mass is obtained by the air density $m = \rho V_0$, and the force equals the decrease of in pressure over the three space directions, times the corresponding partial surface, e.g. for the x direction

¹Typical constants are: density $\rho = 1.2 \text{ kg/m}^3$, speed of sound $c = 343 \text{ m/s}$.

$F_x = -[p(x + \Delta x) - p(x)]\Delta y\Delta z$. For the x direction, this yields after expanding by $\frac{\Delta x}{\Delta x}$

$$-\frac{\Delta p}{\Delta x} V_0 = \rho V_0 \frac{\partial v_x}{\partial t}.$$

Dividing by $-V_0$ and letting $V_0 \rightarrow 0$, we obtain the typical shape of the equation of motion for all three space directions

$$\nabla p = -\rho \frac{\partial \mathbf{v}}{\partial t}. \quad (6.2)$$

The equation of motion means: *Independently of the common exterior pressure load on all the outer boundaries of a small air package, an outer pressure decrease into any direction implies a corresponding pushing force on the package causing a proportional acceleration into this direction.*

6.3 Wave Equation

We can combine the compression equation $\frac{\partial p}{\partial t} = -\rho c^2 \nabla^T \mathbf{v}$ with the equation of motion $\nabla p = -\rho \frac{\partial \mathbf{v}}{\partial t}$ by deriving the first one with regard to time $\frac{\partial^2 p}{\partial t^2} = -\rho c^2 \nabla^T \frac{\partial \mathbf{v}}{\partial t}$ and the second one with the gradient ∇^T yielding the Laplacian $\nabla^T \nabla = \Delta$, hence $\Delta p = -\rho \nabla^T \frac{\partial \mathbf{v}}{\partial t}$. Division of the first result by c^2 and equating both terms yields the lossless wave equation $\Delta p = \frac{1}{c^2} \frac{\partial^2 p}{\partial t^2}$ that is typically written as

$$\left(\Delta - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) p = 0. \quad (6.3)$$

Obviously, the wave equation relates the curvature in space (expressed by the Laplacian) to curvature in time (expressed by the second-order derivative).

If p is a pure sinusoidal oscillation $\sin(\omega t + \phi_0)$, the second derivative in time corresponds to a factor $-\omega^2$, and by substitution with the wave-number $k = \frac{\omega}{c}$, we can write the frequency-domain wave equation as

$$(\Delta + k^2) p = 0, \quad \text{Helmholtz equation.} \quad (6.4)$$

6.3.1 Elementary Inhomogeneous Solution: Green's Function (Free Field)

The Green's function is an elementary prototype for solutions to inhomogeneous problems $(\Delta + k^2)p = -q$, which is defined as

$$(\Delta + k^2)G = -\delta.$$

A general excitation q of the equation can be represented by its convolution with the Dirac delta distribution $\int q(\mathbf{s}) \delta(\mathbf{r} - \mathbf{s}) dV(\mathbf{s}) = q(\mathbf{r})$. Consequently, as the wave equation is linear, the general solution must therefore also equal the convolution of the Green's function with the excitation function $p(\mathbf{r}) = \int q(\mathbf{s}) G(\mathbf{r} - \mathbf{s}) dV(\mathbf{s})$ over space; if formulated in the time domain: also over time. The integral superimposes acoustical responses of any point in time and space of the source phenomenon, weighted by the corresponding source strength in space and time.

The Green's function in three dimensions is derived in Appendix A.6.3, Eq. (A.91),

$$G = \frac{e^{-ikr}}{4\pi r}, \quad (6.5)$$

with the wave number $k = \frac{\omega}{c}$ and distance between source and receiver $r = \sqrt{\|\mathbf{r} - \mathbf{r}_s\|^2}$.

Acoustic source phenomena are characterized by the behavior of the Green's function: far away, the amplitude decays with $\frac{1}{r}$ and the phase $-kr = -\omega \frac{r}{c}$ corresponds to the radially increasing delay $\frac{r}{c}$. Both is expressed in Sommerfeld's radiation condition $\lim_{r \rightarrow \infty} r \left(\frac{\partial}{\partial r} p + ik p \right) = 0$.

Plane waves. The radius coordinate of the Green's function is the distance between two Cartesian position vectors \mathbf{r}_s and \mathbf{r} , the source and receiver location. Letting one of them become large is denoted by re-expressing it in terms of radius and direction vector $\mathbf{r}_s = r_s \boldsymbol{\theta}_s$. This permits far-field approximation

$$r_s = \|\mathbf{r}_s - \mathbf{r}\| = \sqrt{(r_s \boldsymbol{\theta}_s - \mathbf{r})^T (r_s \boldsymbol{\theta}_s - \mathbf{r})} = \sqrt{r_s^2 - 2r_s \boldsymbol{\theta}_s^T \mathbf{r} + r^2} \quad (6.6)$$

$$\lim_{r_s \rightarrow \infty} r_s = \lim_{r_s \rightarrow \infty} r_s \sqrt{1 - 2 \frac{\boldsymbol{\theta}_s^T \mathbf{r}}{r_s} + \frac{r^2}{r_s^2}} = r_s - \boldsymbol{\theta}_s^T \mathbf{r}. \quad (\text{with } \lim_{x \rightarrow 0} \sqrt{1 - 2x} = 1 - x).$$

For the *phase approximation*, for instance at a wave-length of 30 cm, we notice even for a relatively small distance difference, e.g. between 15 m and 15 m + 15 cm, we could change the sign of the wave. To approximate the phase of the Green's function, we must therefore at least use $r_s - \boldsymbol{\theta}_s^T \mathbf{r}$ as approximation. By contrast, this level of precision is irrelevant for the *magnitude approximation*, e.g., it would be negligible if we used $\frac{1}{15 \text{ m}}$ instead of the magnitude $\frac{1}{15 \text{ m} + 15 \text{ cm}}$.

At a large distance r_s assumed to be constant, the Green's function is proportional to a *plane wave* from the source direction $\boldsymbol{\theta}_s$

$$\lim_{r_s \rightarrow \infty} G = \frac{e^{-ikr_s}}{4\pi r_s} e^{ik \boldsymbol{\theta}_s^T \mathbf{r}}. \quad (6.7)$$

The plane-wave part is of unit magnitude $|p| = 1$

$$p = e^{ik \boldsymbol{\theta}_s^T \mathbf{r}} \quad (6.8)$$

and its phase evaluates the projection of the position vector onto the plane-wave arrival direction $\boldsymbol{\theta}_s$. Towards the direction $\boldsymbol{\theta}_s$, the phase grows positive, i.e. the signal arrives earlier. Towards the plane-wave propagation direction $-\boldsymbol{\theta}_s$ the phase grows negatively, implying an increasing time delay, which is constant on any plane perpendicular to $\boldsymbol{\theta}_s$.

Plane waves are an invaluable tool to locally approximate sound fields from sources that are sufficiently far away, within a small region.²

6.4 Basis Solutions in Spherical Coordinates

Figure 4.11 shows spherical coordinates [14, 15] using radius r , azimuth φ , and zenith ϑ . For simplification, zenith is replaced by $\zeta = \cos \vartheta = \frac{z}{r}$, here. We may solve the Helmholtz equation $(\Delta + k^2)p = 0$ in spherical coordinates by the radial and directional parts of the Laplacian $\Delta = \Delta_r + \Delta_{\varphi,\zeta}$, as identified in Appendix A.3

$$\Delta_r = \frac{\partial^2}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r}, \quad \Delta_{\varphi,\zeta} = \frac{1 - \zeta^2}{r^2} \frac{\partial^2}{\partial \zeta^2} - \frac{2}{r^2} \zeta \frac{\partial}{\partial \zeta} + \frac{1}{r^2(1 - \zeta^2)} \frac{\partial^2}{\partial \varphi^2}. \quad (6.9)$$

We already know the spherical harmonics as directional eigensolutions from Sect. 4.7

$$\Delta_{\varphi,\zeta} Y_n^m = -\frac{n(n+1)}{r^2} Y_n^m \quad (6.10)$$

and assume them to be a factor of the solution $p_n^m = R Y_n^m$ determining the value of $\Delta_{\varphi,\zeta}$ in $(\Delta_r + k^2 + \Delta_{\varphi,\zeta})p_n^m = 0$. We find a separated radial differential equation after insertion, multiplication by $\frac{r^2}{Y_n^m}$, and re-expressing the differentials $\frac{\partial}{\partial r} = k \frac{\partial}{\partial kr}$ and $\frac{\partial^2}{\partial r^2} = k^2 \frac{\partial^2}{\partial (kr)^2}$

$$\left[(kr)^2 \frac{\partial^2}{\partial (kr)^2} + 2(kr) \frac{\partial}{\partial (kr)} + (kr)^2 - n(n+1) \right] R = 0. \quad (6.11)$$

Appendix A.6.4 shows how to get physical solutions for R of this, so-called, *spherical Bessel differential equation*: spherical Hankel functions of the second kind $h_n^{(2)}(kr)$ able to represent radiation (radially outgoing into every direction), consistently with Green's function G , diverging with an $(n+1)$ -fold pole at $kr = 0$, a physical behavior that would also be observed after spatially differentiating G , see Fig. 6.1; spherical Bessel functions $j_n(kr) = \Re\{h_n^{(2)}(kr)\}$ are real-valued, converge everywhere, exhibit

²This is because, strictly speaking, an entire plane-wave sound field is unphysical and of infinite energy: either the exhaustive in-phase vibration of an infinite plane is required, or an infinite-amplitude point-source infinitely far away is required with infinite anticipation $t_s \rightarrow +\infty$ (non-causal).

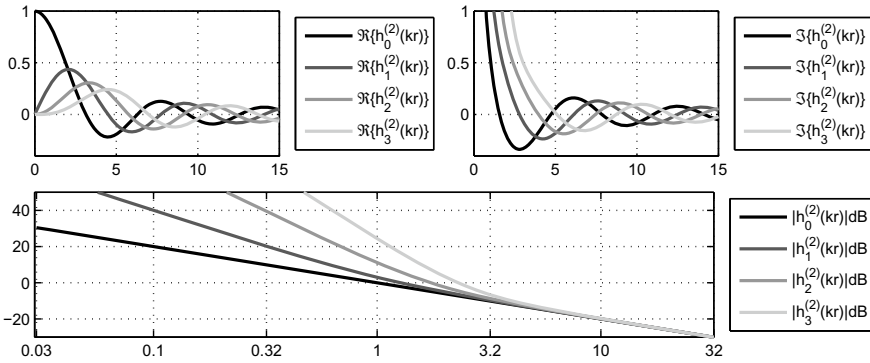


Fig. 6.1 Spherical Bessel functions $j_n(kr) = \Re\{h_n^{(2)}(kr)\}$ (top left), imaginary part of spherical Hankel functions $\Im\{h_n^{(2)}(kr)\}$ (top right), and magnitude/dB of $|h_n^{(2)}(kr)|$ (bottom), over kr

an n -fold zero at $kr = 0$, and can't represent radiation. Implementations typically rely on the accurate standard libraries implementing cylindrical Bessel and Hankel functions:

$$j_n(kr) = \sqrt{\frac{\pi}{2}} \frac{1}{kr} J_{n+\frac{1}{2}}(kr), \quad h_n^{(2)}(kr) = \sqrt{\frac{\pi}{2}} \frac{1}{kr} H_{n+\frac{1}{2}}^{(2)}(kr). \quad (6.12)$$

Wave spectra and spherical basis solutions. Any sound field evaluated at a radius r where the air is source-free and homogeneous in any direction can be represented by spherical basis functions for enclosed $j_n(kr)Y_n^m(\theta)$ and radiating fields $h_n(kr)Y_n^m(\theta)$

$$p = \sum_{n=0}^{\infty} \sum_{m=-n}^n [b_{nm}j_n(kr) + c_{nm}h_n(kr)]Y_n^m(\theta). \quad (6.13)$$

Here, b_{nm} are the coefficients for *incoming waves* that pass through and emanate from radii larger than r and c_{nm} are the coefficients of *outgoing waves* radiating from sources at radii smaller than r ; the coefficients are called *wave spectra* of the incoming and outgoing waves, cf. [16].

Ambisonic plane-wave spectrum, plane wave. Plane waves only use the coefficients b_{nm} , while $c_{nm} = 0$ in Eq. (6.13). The sum of incoming plane waves from all directions, whose amplitudes are given by the spherical harmonics coefficients χ_{nm} as a set of Ambisonic signals are described by the *incoming wave spectrum*, see Appendix A.6.5, Eq. (A.119)

$$b_{nm} = 4\pi i^n \chi_{nm}. \quad (6.14)$$

Figure 6.2 shows a single plane wave incoming from the direction θ_s represented by

$$b_{nm} = 4\pi i^n Y_n^m(\theta_s) \quad (6.15)$$

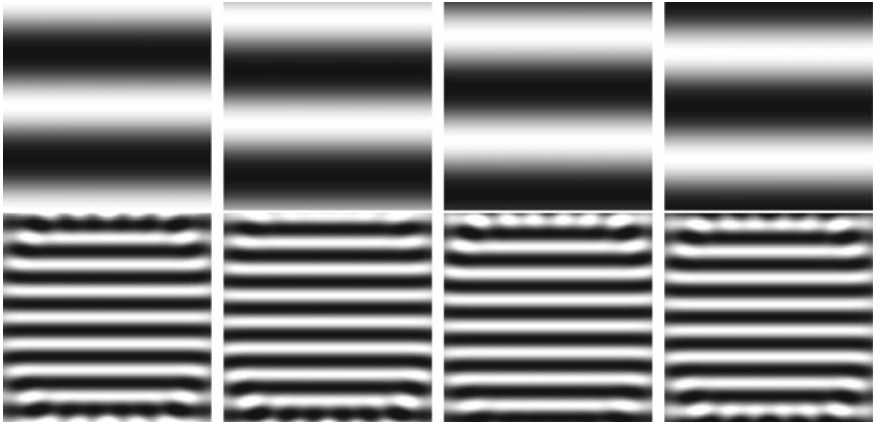


Fig. 6.2 Plane wave from y axis $\varphi = \vartheta = \frac{\pi}{2}$ in horizontal cross section; time steps correspond to $0^\circ, 60^\circ, 120^\circ,$ and 180° phase shifts ϕ in the plot $\Re\{p e^{i\phi}\}$ showing p from Eq. (6.13) with $c_{nm} = 0$ and b_{nm} of Eq. (6.15) with $b_{nm} = 4\pi i^n Y_n^m(\frac{\pi}{2}, \frac{\pi}{2})$; long wave (top), short wave (bottom); simulation uses $N = 25$ and area shows $|kx|, |ky| < 2\pi$ and 8π

at four different time steps corresponding to $0^\circ, 60^\circ, 120^\circ$ and 180° time shifts for the two wave lengths shown.

6.5 Scattering by Rigid Higher-Order Microphone Surface

Higher-order Ambisonic microphone arrays are typically mounted on a rigid sphere of some radius $r = a$, such as the Eigenmike EM32, see Fig. 6.3. The physical boundary of the rigid spherical surface is expressed as a vanishing radial component of the sound particle velocity. The radial sound particle velocity is obtained via the

Fig. 6.3 32-channel higher-order Ambisonic mic. Eigenmike EM32



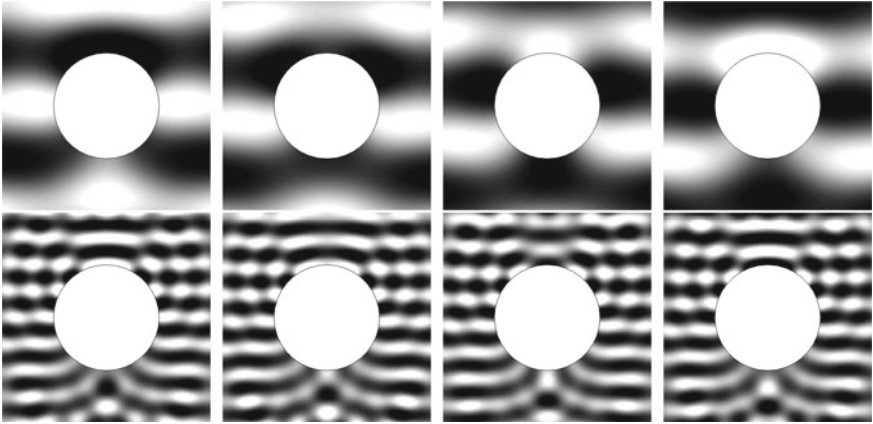


Fig. 6.4 Plane waves scattered by rigid sphere $ka = \pi$ (top) or $ka = 4\pi$ (bottom); time steps correspond to 0° , 60° , 120° , and 180° phase shifts ϕ in the plot $\Re\{p e^{i\phi}\}$ showing p from Eq. (6.13) with b_{nm} and c_{nm} from Eq. (6.15) with $b_{nm} = 4\pi i^n Y_n^m(\frac{\pi}{2}, \frac{\pi}{2})$ and Eq. (6.16); simulation uses $N = 25$

equation of motion Eq. (6.2) by deriving Eq. (6.13). This requires to evaluate differentiated spherical radial solutions $j'_n(x)$ as well as $h_n^{(2)}(x)$, which is implemented by $f'_n(x) = \frac{n}{x} f_n(x) - f_{n+1}(x)$ for either of the functions, cf. e.g. [16]. A sound-hard boundary condition at the radius a requires

$$v_r|_{r=a} = \frac{i}{\rho c} \sum_{n=0}^{\infty} \sum_{m=-n}^n [b_{nm} j'_n(kr) + c_{nm} h_n^{(2)}(kr)]_{r=a} Y_n^m(\theta) = 0,$$

which is fulfilled by a vanishing term in square brackets. The rigid boundary responds to incoming surround-sound by velocity-canceling outgoing waves $h_n^{(2)}(ka) c_{nm} = -j'_n(ka) b_{nm}$. The coefficients ψ_{nm} yield the sound pressure in Fig. 6.4,

$$p = \sum_{n=0}^{\infty} \sum_{m=-n}^n \psi_{nm} Y_n^m(\theta), \quad \text{with } \psi_{nm} = \left[j_n(kr) - h_n^{(2)}(kr) \frac{j'_n(ka)}{h_n^{(2)}(ka)} \right]_{r=a} b_{nm}. \quad (6.16)$$

The two terms of the bracket are typically further simplified by a common denominator and recognizing the Wronskian Eq. (A.97) in the numerator $\frac{j_n(x)h'_n(x) - j'_n(x)h_n(x)}{h_n'(x)} = \frac{i}{x^2 h_n'(x)}$

$$\psi_{nm}|_{r=a} = \frac{i}{(ka)^2 h_n^{(2)}(ka)} b_{nm}. \quad (6.17)$$

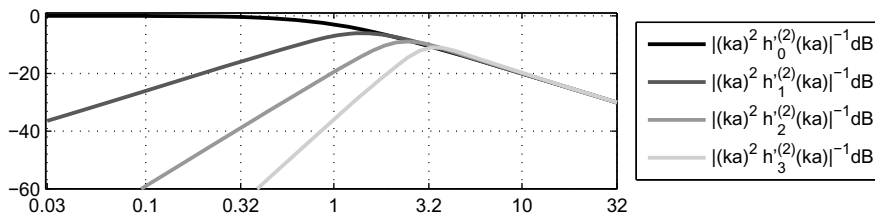


Fig. 6.5 Attenuation/dB of Ambisonic signals of different orders for varying values of ka

Relation of recorded sound pressure to Ambisonic signal. The scattering equation relates the recorded sound pressure expanded in spherical harmonics to the Ambisonic signal of surround sound scene, see frequency responses in Fig. 6.5,

$$\psi_{nm}|_{r=a} = \frac{4\pi i^{n+1}}{(ka)^2 h_n^{(2)}(ka)} \chi_{nm}. \quad (6.18)$$

It is formally convenient that as soon as the sound pressure is given in terms of its spherical harmonic coefficient signals ψ_{nm} , the Ambisonic signals χ_{nm} of a concentric playback system are obviously just an inversely filtered version thereof, with no need for further unmixing/matrixing.

Recognizable from Fig. 6.6 and following our intuition, waves of lengths larger than the diameter $2a$ of the sphere will only weakly map to complicated high-order patterns. It is therefore easily understood that the transfer function $i^{n+1}[(ka)^2 h_n^{(2)}(ka)]^{-1}$ attenuates the reception of high-order Ambisonic signals at low frequencies, see Fig. 6.5.

6.6 Higher-Order Microphone Array Encoding

The block diagram of Ambisonic encoding of higher-order microphone array signals is shown in Fig. 6.7. The first processing step is about decomposing the pressure samples $p(t)$ from the microphone array into its spherical harmonics coefficients $\psi_N(t)$: To which amount do the samples contain omnidirectional, figure-of-eight, and other spherical harmonic patterns, up to which the microphone arrangement allows decomposition. The frequency-independent matrix $(Y_N^T)^\dagger$ does the conversion. It is the left-inverse to the spherical harmonics sampled at the microphone positions, as shown in the upcoming section.

The second step then sharpens the sound pressure image to an Ambisonic signal by filtering the spherical harmonic coefficient signals. The basic relation between sound pressure coefficients and Ambisonic signals is given in Eq. (6.18) and describes a filter for every coefficient signal, differing only in filter characteristics for different spherical harmonic orders. Robustness to noise, microphone matching and position-

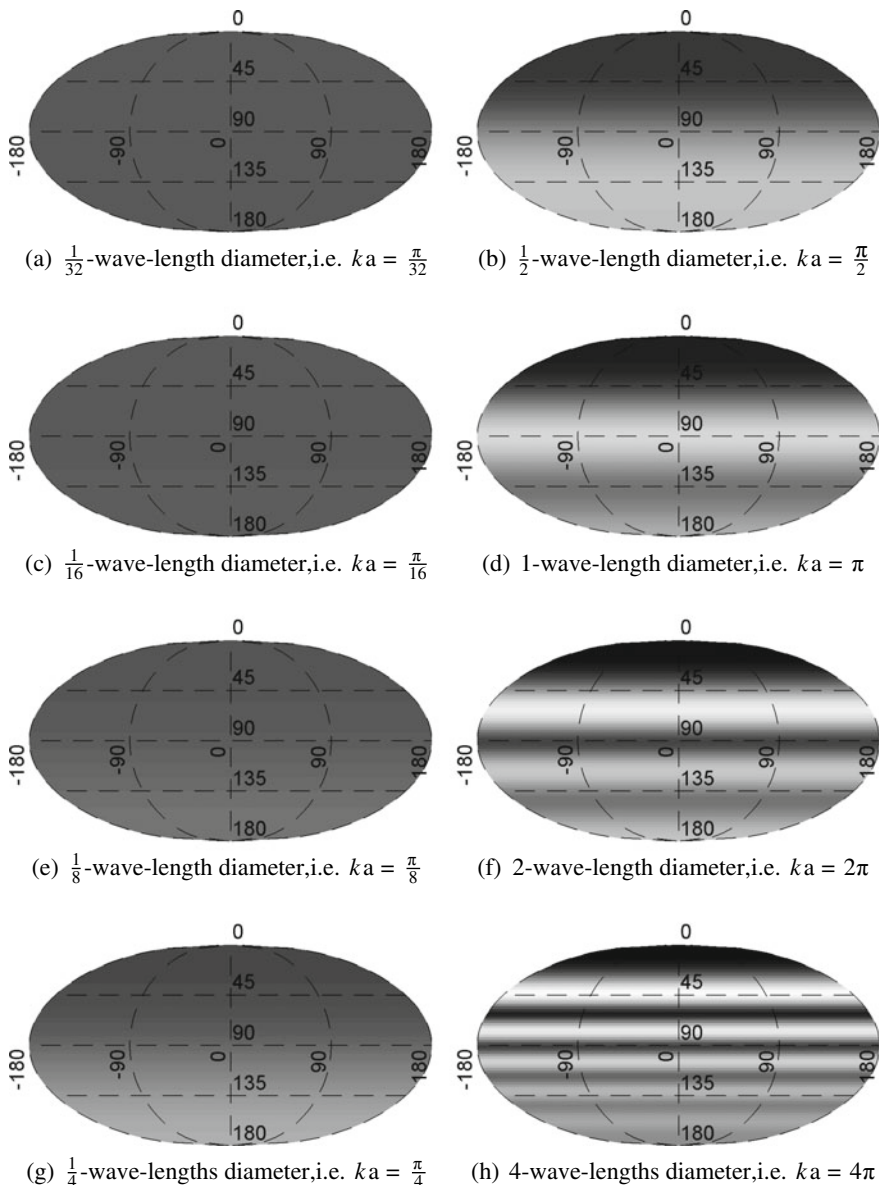
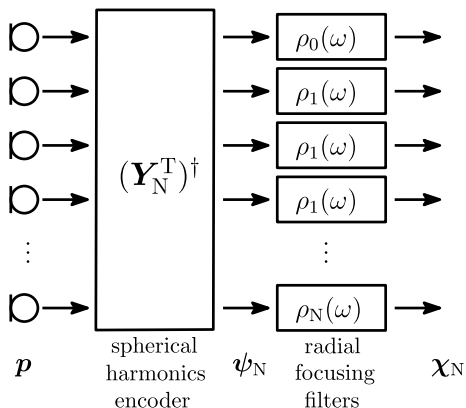


Fig. 6.6 Plane-wave sound pressure image $\Re\{p e^{-ika}\}$ on rigid sphere with varying ka using ψ_{nm} from Eq. (6.17) expanded over the spherical harmonics $p = \sum \psi_{nm} Y_n^m$ and $\chi_{nm} = Y_n^m(0, 0)$ for a plane wave from z . With the wave length $\lambda = \frac{c}{f}$, the value ka is related to a diameter $2a$ of the $\frac{ka}{\pi} = \frac{2\pi f a}{\pi c} = \frac{2a}{\lambda}$ in wave lengths to express frequency dependency; simulation uses $N = 50$; for $a = 4.2$ cm, ka values correspond to $f = 125, 250, 500, 1000, 2000, 4000, 8000, 16000$ Hz

Fig. 6.7 Higher-order Ambisonic microphone encoding: sound pressure samples $\mathbf{p}(t)$ are spherical-harmonics decomposed by the matrix $(\mathbf{Y}_N^T)^\dagger$, and the resulting coefficient signals $\boldsymbol{\psi}_N(t)$ are converted to Ambisonic signals $\boldsymbol{\chi}_N(t)$ by the sharpening filters $\rho_n(\omega)$



ing is the key here, and only achieved by the careful design of these filters, as shown in a further sections below. The design considers a gradually increasing sharpening over frequency, for which it moreover employs a filter bank with separate, $\max-r_E$ weighted and E normalized bands, in order to provide (i) limitation of noise and errors, (ii) a frequency response perceived as flat, and (iii) optimal suppression of the sidelobes.

6.7 Discrete Sound Pressure Samples in Spherical Harmonics

To determine the Ambisonics signals χ_{nm} , we obviously need to find ψ_{nm} based on all sound pressure samples $p(\boldsymbol{\theta}_i)$ recorded by the microphones distributed on the rigid-sphere array. To accomplish this, we set up a system of model equations equating the pressure samples to the unknown coefficients ψ_{nm} expanded over the spherical harmonics $Y_n^m(\boldsymbol{\theta}_i)$ sampled at every microphone position. A vector and matrix notation $\mathbf{p} = [p(\boldsymbol{\theta}_i)]_i$ and $\mathbf{Y}_N^T = [\mathbf{y}(\boldsymbol{\theta}_i)^T]_{i, nm}$ is helpful

$$\begin{bmatrix} p(\boldsymbol{\theta}_1) \\ \vdots \\ p(\boldsymbol{\theta}_M) \end{bmatrix} = \begin{bmatrix} Y_0^0(\boldsymbol{\theta}_1) & \dots & Y_N^N(\boldsymbol{\theta}_1) \\ \vdots & \vdots & \vdots \\ Y_0^0(\boldsymbol{\theta}_M) & \dots & Y_N^N(\boldsymbol{\theta}_M) \end{bmatrix} \begin{bmatrix} \psi_{00} \\ \vdots \\ \psi_{NN} \end{bmatrix}$$

$$\mathbf{p}_N = \mathbf{Y}_N^T \boldsymbol{\psi}_N. \tag{6.19}$$

Left inverse (MMSE). The equation can be (pseudo-)inverted if the matrix \mathbf{Y}_N is well conditioned. Typically more microphones are used than coefficients searched $M \geq (N + 1)^2$. Inversion is a matter of mean-square error minimization: As the M dimensions may contain more degrees of freedom than $(N + 1)^2$, the coefficient

vector $\boldsymbol{\psi}_N$ giving the closest model \boldsymbol{p}_N to the measurement \boldsymbol{p} is searched,

$$\min_{\boldsymbol{\psi}_N} \|\boldsymbol{e}\|^2, \quad \text{with } \boldsymbol{e} = \boldsymbol{p}_N - \boldsymbol{p} = \boldsymbol{Y}_N^T \boldsymbol{\psi}_N - \boldsymbol{p}. \quad (6.20)$$

The minimum-mean-square-error (MMSE) solution is, see Appendix A.4, Eq. (A.65),

$$\boldsymbol{\psi}_N = (\boldsymbol{Y}_N \boldsymbol{Y}_N^T)^{-1} \boldsymbol{Y}_N \boldsymbol{p} = (\boldsymbol{Y}_N^T)^\dagger \boldsymbol{p}. \quad (6.21)$$

The resulting left inverse $(\boldsymbol{Y}_N \boldsymbol{Y}_N^T)^{-1} \boldsymbol{Y}_N$ inverts the thin matrix \boldsymbol{Y}_N^T from the left. $(\boldsymbol{Y}_N^T)^\dagger$ symbolizes the pseudo inverse; it is left-inverse for thin matrices.

If the microphones are arranged in a t -design and the order N is chosen suitably, then the transpose matrix times $\frac{4\pi}{L}$ is equivalent to the left inverse. A more thorough discussion on spherical point sets can be found in [17–19].

The *maximum determinant points* [20] are a particular kind of *critical* directional sampling scheme that allows to use exactly as few microphones $M = (N + 1)^2$ as spherical harmonic coefficients obtained, yielding a well-conditioned square matrix \boldsymbol{Y}_N , so that it can be inverted directly without left/pseudo-inversion. The 25 maximum-determinant points for $N = 4$ are used in the simulation example below.³

Finite-order assumption and spatial aliasing. An important implication of estimating ψ_{nm} is that we need to assume that the distribution of the sound pressure is of limited spherical harmonic order on the measurement surface. This could be done by restricting the frequency range, as high-order harmonics are attenuated well-enough according above suitable frequency limits, cf. Fig. 6.5. However, low-pass filtered signals are unacceptable in practice. Instead, one has to accept *spatial aliasing* at high frequencies, i.e. directional mapping errors and direction-specific comb filters. Figure 6.8 shows spatial aliasing of $\boldsymbol{\psi}_N = (\boldsymbol{Y}_N^T)^{-1} \boldsymbol{p}$ in the angular domain $\boldsymbol{p} = \sum \psi_{nm} \boldsymbol{Y}_n^m$.

6.8 Regularizing Filter Bank for Radial Filters

The filters $i^n [(ka)^2 h_n^{(2)}(ka)]^{-1}$ of Fig. 6.5 exhibit an n th-order zero at 0 Hz, $ka = 0$. To retrieve the Ambisonic signals χ_{nm} from the sound pressure signals ψ_{nm} , their inverse would have a n -fold (unstable) pole at 0 Hz. Considering that microphone self noise and array imperfection cause erroneous signals louder than the acoustically expected n th-order vanishing signals around 0 Hz, filter shapes will moreover cause an excessive boost of erroneous signals unless implemented with precaution. Filters of the different orders n must be stabilized by high-pass slopes of at least the order n , see also [6, 9, 21–25], and with $(n + 1)$ th-order high-pass slopes, see Fig. 6.9, such errors are being cut off by first-order high-pass slopes at exemplary cut-on frequencies at 90, 680, 1650, 2600 Hz for the Ambisonic orders 1, 2, 3, 4, yielding a

³md04.0025 on https://web.maths.unsw.edu.au/~rsw/Sphere/Images/MD/md_data.html.

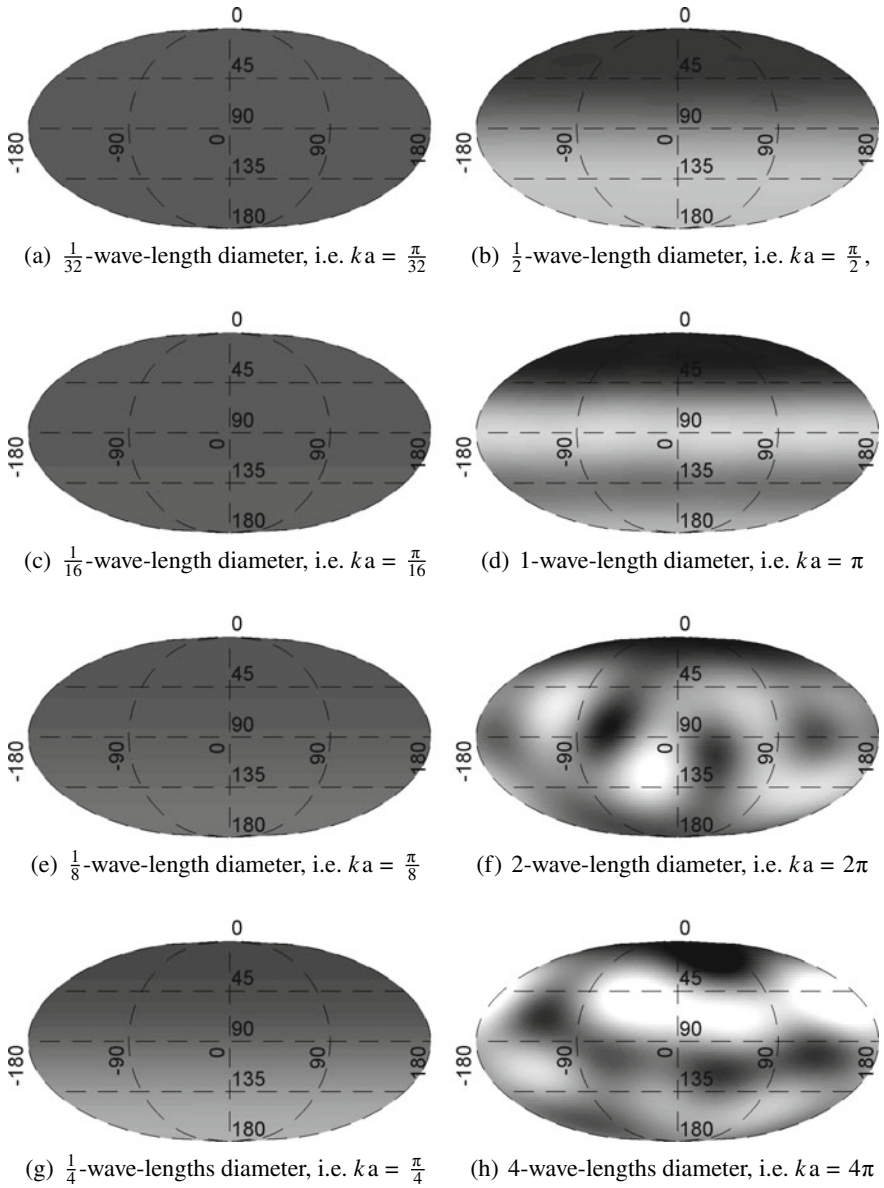


Fig. 6.8 Interpolated plane-wave sound pressure image $\Re\{p e^{-ika}\}$ on rigid-sphere array with 25 microphones allowing decomposition up to the order $N = 4$; simulation uses orders up to 25, and the aliasing-free operation can only be expected within $kr < N$

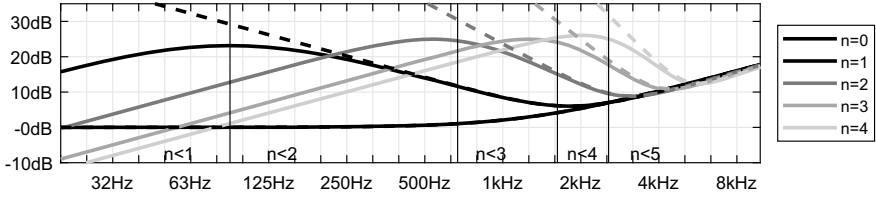


Fig. 6.9 Filters $(ka)^2 h_n^{(2)}(ka)$ /dB over frequency/Hz, regularized with $(n + 1)$ th-order high-pass filters

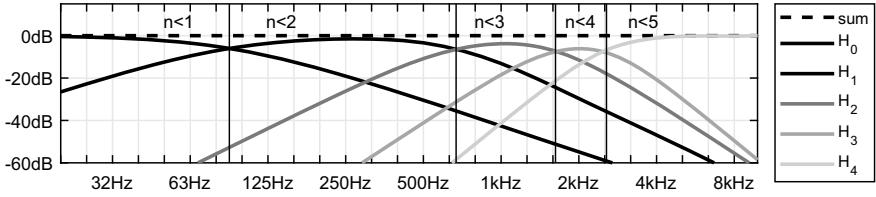


Fig. 6.10 Stabilizing filter bank/dB over frequency/Hz: signal orders $n > b$ are excluded from the band b

noise boost of 20 dB for a 4th-order microphone with $a = 4.2$ cm, at most. However, just cutting on the frequencies of each order is not enough: every cut-on frequency causes a noticeable loudness drop below due to the discarded signal contributions. It is better to design a filter bank with crossovers instead, which allows compensation for the loudness loss in every band. A zero-phase, n th-order Butterworth high-pass response is defined by $H_{hi} = \frac{\omega^n}{1+\omega^n}$ and amplitude-complementary to the low pass $H_{lo} = \frac{1}{1+\omega^n}$, so that $H_{hi} + H_{lo} = 1$.

Using this filter type, the filter bank in Fig. 6.10 can be constructed as follows: The band-pass filters $H_b(\omega)$ are composed of a $(b + 1)$ th-order high- and $(b + 2)$ th-order low-pass skirt at ω_b , and ω_{b+1} , respectively, except for the band $b = 0$ (low-pass) and $b = N$ (high-pass)

$$\hat{H}_0(\omega) = \frac{1}{1 + (\frac{\omega}{\omega_1})^2}, \quad \hat{H}_b(\omega) = \frac{(\frac{\omega}{\omega_b})^{b+1}}{1 + (\frac{\omega}{\omega_b})^{b+1}} \frac{1}{1 + (\frac{\omega}{\omega_{b+1}})^{b+2}}, \quad \hat{H}_N(\omega) = \frac{(\frac{\omega}{\omega_N})^{N+1}}{1 + (\frac{\omega}{\omega_N})^{N+1}}. \tag{6.22}$$

To make the bands perfectly reconstructing, filters are normalized by the sum response

$$H_b = \frac{\hat{H}_b}{\sum_{b=0}^N \hat{H}_b(\omega)}. \tag{6.23}$$

By adjusting the cut-on frequencies ω_b of the different orders $b = 1, \dots, N$, the noise and mapping behavior of the microphone array is adjusted; only the zeroth order is present in every band down to 0 Hz.

This filter bank design moreover allows to adjust loudness and sidelobe suppression in every frequency band, separately.

6.9 Loudness-Normalized Sub-band Side-Lobe Suppression

The filter bank design shown above would only yield Ambisonic signals whose order increases with the frequency band. Ideally, this variation of the order comes with the necessity of individual max- r_E sidelobe suppression in every band. Moreover, Ambisonic signals of different orders are differently loud, so also diffuse-field equalization of the E measure is desirable in every band.

To fulfill the above constraints, we propose to use the following set of FIR filter responses as given in [26, 27], that are modified by a filter bank employing diffuse-field normalized max- r_E -weights in separate frequency bands $b = 0, \dots, N$, cf. Fig. 6.11, with the n th order discarded for bands below $b < n$:

$$\rho_n(\omega) = \left[\sum_{b=n}^N a_{n,b} H_b(\omega) \right] i^{-n-1} (ka)^2 h_n^{(2)}(ka) e^{ika}. \quad (6.24)$$

Here, e^{ika} removes the linear phase of $h_n^{(2)}$, and $a_{n,b}$ is the set of diffuse-field (\sqrt{E}) equalized max- r_E weights for the band b in which the Ambisonic orders retrieved are $0 \leq n \leq b$

$$a_{n,b} = \begin{cases} P_n \left(\cos \frac{137.9^\circ}{b+1.51} \right) \sqrt{\frac{\sum_{n=0}^N (2n+1) \left[P_n \left(\cos \frac{137.9^\circ}{N+1.51} \right) \right]^2}{\sum_{n=0}^b (2n+1) \left[P_n \left(\cos \frac{137.9^\circ}{b+1.51} \right) \right]^2}}, & \text{for } n \leq b \\ 0, & \text{otherwise.} \end{cases} \quad (6.25)$$

Figure 6.12 shows the polar patterns of the corresponding direction-spread functions.

For the implementation of $\rho_n(\omega)$ by fast block filtering, $\omega = 2\pi f$ and $k = \omega/c$ are uniformly sampled with frequency, and the inverse discrete Fourier transform yields the associated impulse responses (attention: the value at 0 Hz must be replaced for stable results, and cyclic time-domain shifts and windows are necessary).

The direction-spread function of a plane-wave sound pressure mapped to a directional Ambisonic signal becomes frequency-dependent as shown in Fig. 6.13, and it has minimal side lobes.

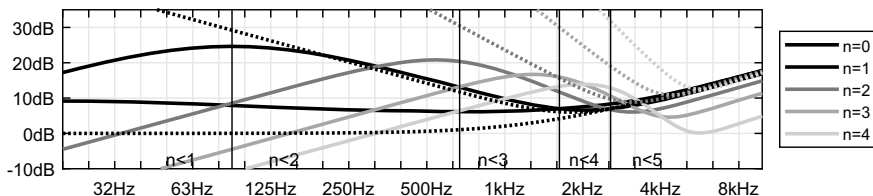


Fig. 6.11 Filter-bank-regularized/dB over frequency/Hz, diffuse-field equalized max- r_E weighted spherical microphone array responses using $i^n \rho_n(\omega) = \sum_{b=n}^N a_{n,b} H_b(\omega) (ka)^2 h_n^{(2)}(ka)$

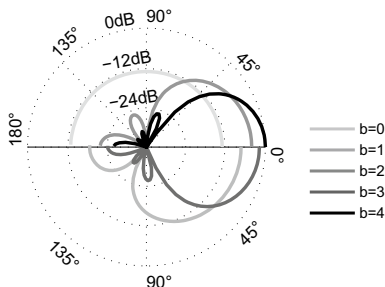


Fig. 6.12 Diffuse-field equalized (to $E = 1$) max- r_E direction-spread functions; even orders are plotted on upper, odd orders on lower semi-circle

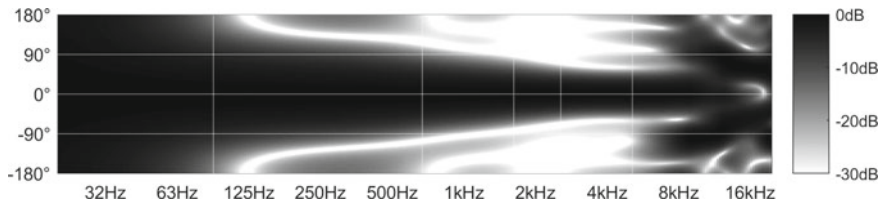


Fig. 6.13 Direction spread/dB over frequency/Hz in zenithal cross section/degrees through Ambisonic signal of simulated microphone processing response to plane wave from zenith and the parameters $a = 4.2$ cm, $M = 25$ mics., max- r_E -weighted in bands 90, 680, 1650, 2600 Hz for the cut on of the orders 1, 2, 3, 4. Simulation is done with the order $N_{sim} = 30$ and spatial aliasing will occur above 5.2 kHz. Gain matching was assumed to be up to $< \pm 0.5$ dB accurate; the map shows the direction spread normalized to its value at 0° for every frequency to make its shape easier to read

6.10 Influence of Gain Matching, Noise, Side-Lobe Suppression

Typical gain mismatch between the microphones is not always more accurate than 0.5 dB. The result is that the physically dominant omnidirectional signal will leak into the higher-order signals by directionally random gain variations. However, acoustically, higher-order components are expected to be weak and to require amplification.

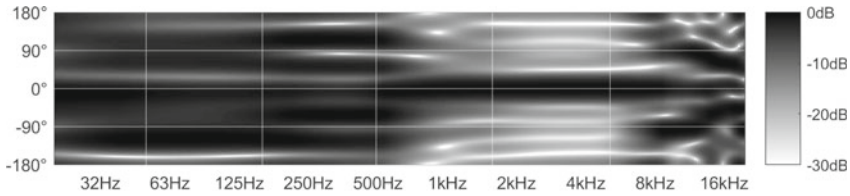
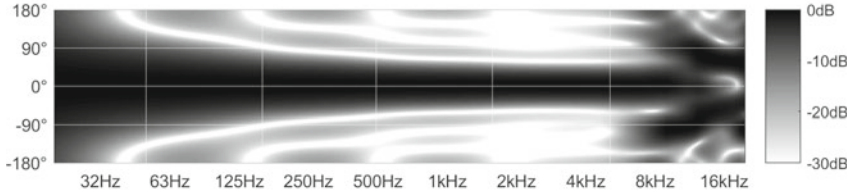
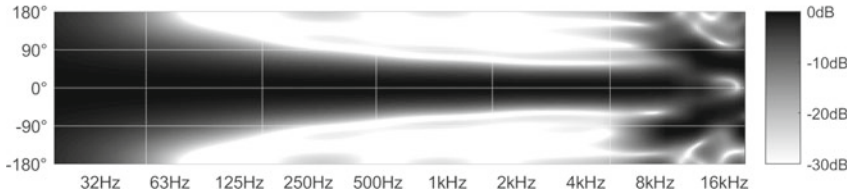
(a) 50, 160, 500, 1600 cut on with $< \pm 0.5$ dB gain matching, no sidelobe suppression(b) 50, 160, 500, 1600 cut on with only 4th-order sidelobe suppression, assuming perfect gain match(c) 50, 160, 500, 1600 cut on with individual max- r_E sidelobe suppression per band, assuming perfect gain match

Fig. 6.14 Influence of carelessly selected cut-on frequencies for regularization (top), and of non-individual sidelobe suppression per band (middle), in contrast to ideal results (bottom); the maps show direction spreads normalized to their values at 0° for every frequency to make side lobes easier to read

The effect on mapping is equivalent to one of microphone self noise, however gain mismatch yields a correlated signal exciting the microphones, whereas self-noise yields low-frequency noise.

If regularization filters were set to 50, 160, 500, 1600 and sidelobe suppression turned off for testing, one would get the poor image as in Fig. 6.14a, where high-order signals at low frequencies are highly boosted.

If a noise-free case is assumed, and only the max- r_E side-lobe suppression of the highest band is used for all bands, one gets the image in Fig. 6.14b, which improves with individual max- r_E weights in Fig. 6.14c.

Self-noise behavior. Assuming that self-noise of the microphones is uncorrelated, it will also remain uncorrelated and of equal strength after decomposing the M microphone signals $p_i = \mathcal{N}$ into the $(N + 1)^2$ spherical harmonic coefficient signals $\psi_{nm} = \frac{(N+1)^2}{M} \mathcal{N}$, if $M \approx (N + 1)^2$ and the microphone arrangement permits a well-conditioned pseudo inversion \mathbf{Y}_N^\dagger . The spectral change of the microphone self noise due to the radial filters $\rho_n(\omega)$ can be described by the noise of the $(2n + 1)$

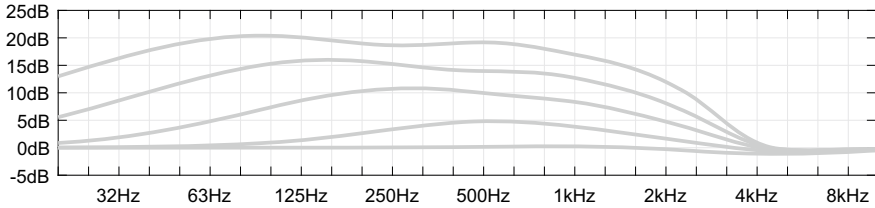


Fig. 6.15 Self-noise modification $|G(\omega)|^2/\text{dB}$ over frequency/Hz for the filter bank configurations using the cut on frequencies $2k, 3k, 4k, 5k$ (no noise amplification), $600, 2k, 3.5k, 4.2k$ (5 dB noise amplification), $280, 1.3k, 2.6k, 3.6k$ (10 dB noise amplification), $150, 950, 2k, 3.15k$ (15 dB noise amplification), and $90, 680, 1.65k, 2.6k$ (20 dB noise amplification)

signals of the same order, amplified by $|\rho_n(\omega)|^2$, in comparison to the zeroth-order signal:

$$|G(\omega)|^2 = \frac{\sum_{n=0}^N (2n+1) |\rho_n(\omega)|^2}{|(ka)^2 h_0^{(2)}(ka)|^2}. \quad (6.26)$$

Figure 6.15 analyzes the noise amplification for the simulation example (max- r_E weighting in each sub band, $a = 4.2$ cm) and shows the dependency on exemplary cut on frequencies configured to tune the filterbank to 0, 5, 10, 15, and 20 dB noise boosts. The trade here is: the more noise boost one can allow, the more directional resolution one gets, see Fig. 6.16.

Open measurement data (SOFA format) characterizing the directivity patterns of the 32 Eigenmike em32 transducers are provided under the link <http://phaidra.kug.ac.at/o:69292>. They are measured on a $12^\circ \times 11.25^\circ$ azimuth \times zenith grid, yielding 480×256 pt impulse responses for each of the 32 transducers.

6.11 Practical Free-Software Examples

6.11.1 Eigenmike Em32 Encoding Using Mcfx and IEM Plug-In Suites

We give a practical signal processing example for the Eigenmike em32 which is applicable e.g. in digital audio workstations. First the 32 signals are encoded by matrix multiplication (IEM `MatrixMultiplier`), cf. Fig. 6.17a, yielding 25 fourth-order signals. The preset (json file) is provided online <http://phaidra.kug.ac.at/o:79231>. The radial filtering that sharpens the surround sound image uses `mcfx-convolver`, see Fig. 6.17b, with 25 SISO filters, one for each Ambisonic signal, using the 5 different filter curves for the orders $n = 0, \dots, 4$ as defined above. The convolver presets (`wav`

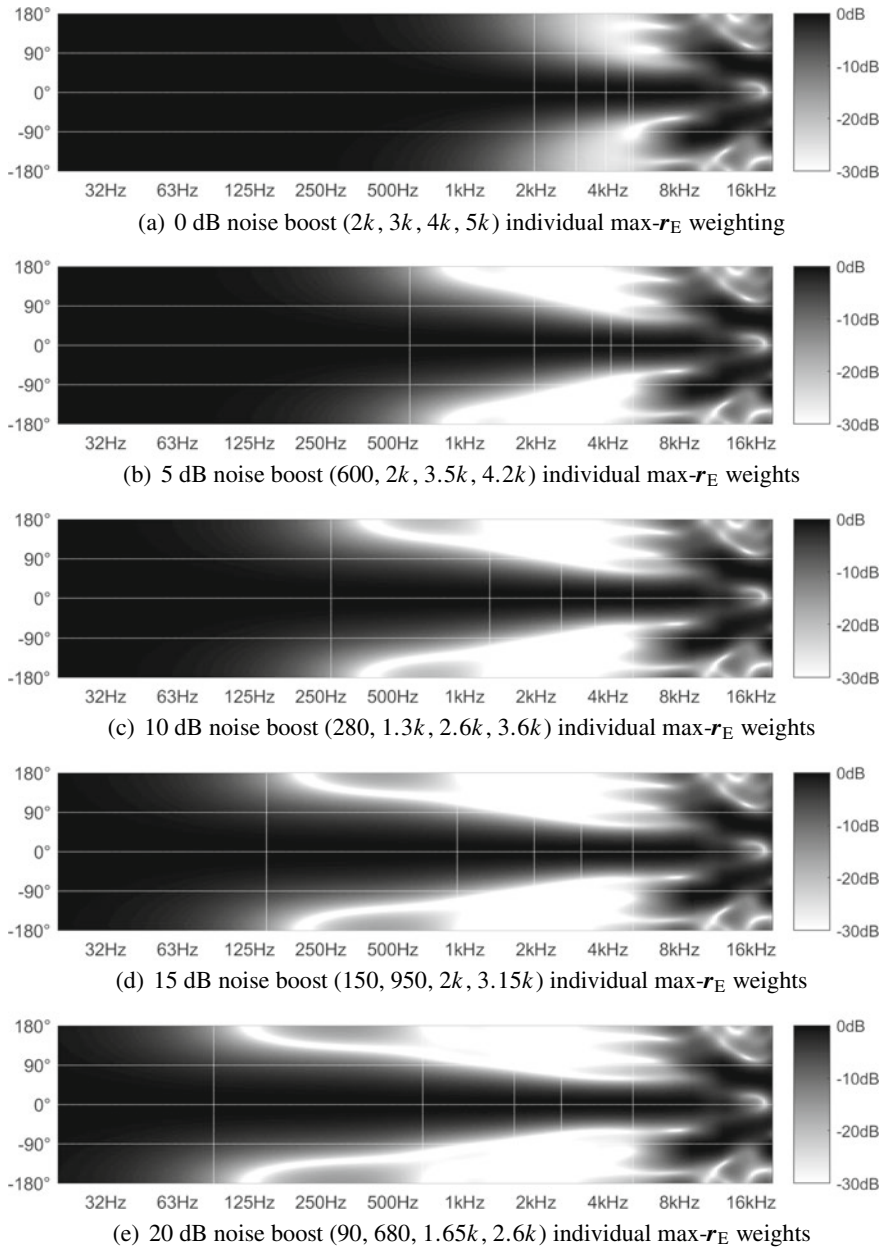


Fig. 6.16 Direction spread/dB for over frequency/Hz and zenith/degrees of filterbank with different settings to achieve 0, 5, 10, 15, 20 dB noise boosts; the maps show direction spreads normalized to their values at 0° at every frequency as above

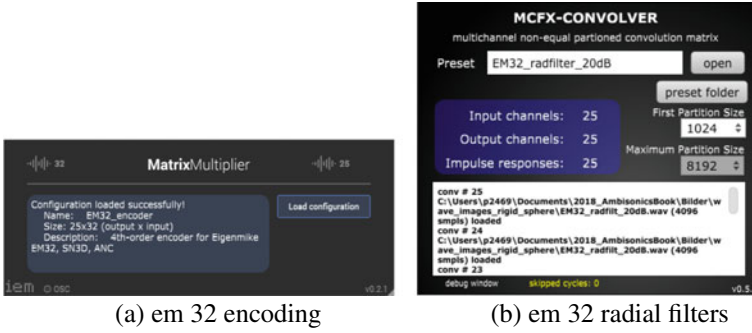


Fig. 6.17 IEM MatrixMultiplier encoding the Eigenmike em32 signals and mcfx-convolver applying radial filters to encoded em32 recording

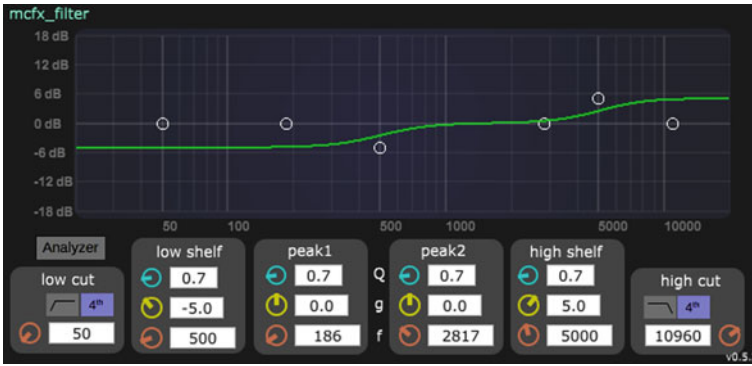


Fig. 6.18 Practical equalization of the em32 transducer characteristics by two parametric shelving filters of the mcfx_filter, cf. [28]

files and config files for mcfx-convolver) are provided online <http://phaidra.kug.ac.at/o:79231> and are available for the different noise boosts 0, 5, 10, 15, 20 dB. As found in [28], the em32 transducers exhibit a frequency response that favors low frequencies and attenuates high frequencies. This behavior is sufficiently well equalized in practice using two parametric shelving filters, a low shelf at 500 Hz with a gain of -5 dB, and a high shelf at 5 kHz using a gain of $+5$ dB, see Fig. 6.18.

6.11.2 SPARTA Array2SH

The SPARTA suite by Aalto University includes the Array2SH plug-in shown in Fig. 6.19 to convert the transducer signals of a microphone array into Ambisonics. It provides both encoding of the signals, as well as calculation and application of radial-focusing filters based on the geometry of the array. It supports rigid and open arrays



Fig. 6.19 SPARTA Array2SH encoding for, e.g., em32

and comes with presets for several arrays, such as the Eigenmike em32. The plug-in allows to adjust the radial filters in terms of regularization type and maximum gain. The Reg. Type called Z-Style corresponds to the linear-phase design of Sect. 6.9.

References

1. J. Daniel, Evolving views on HOA: from technological to pragmatic concerns, in *Proceedings of the 1st Ambisonics Symposium* (Graz, 2009)
2. G.W. Elko, R.A. Kubli, J. Meyer, Audio system based on at least second-order eigenbeams, in *PCT Patent*, vol. WO 03/061336, no. A1 (2003)
3. G.W. Elko, Superdirectional microphone arrays, in *Acoustic Signal Processing for Telecommunication*, ed. by J. Benesty, S.L. Gay (Kluwer Academic Publishers, Dordrecht, 2000)
4. J. Meyer, G. 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, 2002. Proceedings. (ICASSP'02)*, vol. 2 (Orlando, 2002)
5. G.W. Elko, Differential microphone arrays, in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, ed. by Y. Huang, J. Benesty (Springer, Berlin, 2004)
6. J. Daniel, S. Moreau, Further study of sound ELD coding with higher order ambisonics, in *116th AES Convention* (2004)
7. S.-O. Petersen, Localization of sound sources using 3d microphone array, M. Thesis, University of South Denmark, Odense (2004). www.oscarnpetersen.dk/speciale/Thesis.pdf
8. B. Rafaely, Analysis and design of spherical microphone arrays. *IEEE Trans. Speech Audio Process.* (2005)
9. S. Moreau, Étude et réalisation d'outils avancés d'encodage spatial pour la technique de spatialisation sonore Higher Order Ambisonics: microphone 3d et contrôle de distance, Ph.D. Thesis, Université du Maine (2006)
10. H. Teutsch, *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition* (Springer, Berlin, 2007)
11. Z. Li, R. Duraiswami, Flexible and optimal design of spherical microphone arrays for beamforming. *IEEE Trans. ASLP* **15**(2) (2007)

12. W. Song, W. Ellermeier, J. Hald, Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise. *J. Acoust. Soc. Am.* **123**(2) (2008)
13. B. Rafaely, *Fundamentals of Spherical Array Processing*, 2nd edn. (Springer, Berlin, 2019)
14. ISO 31-11:1978, Mathematical signs and symbols for use in physical sciences and technology (1978)
15. ISO 80000-2, quantities and units? Part 2: Mathematical signs and symbols to be used in the natural sciences and technology (2009)
16. E.G. Williams, *Fourier Acoustics* (Academic, Cambridge, 1999)
17. B. Rafaely, B. Weiss, E. Bachmat, Spatial aliasing in spherical microphone arrays. *IEEE Trans. Signal Process.* **55**(3) (2007)
18. F. Zotter, Sampling strategies for acoustic holography/holophony on the sphere, in *NAG-DAGA, Rotterdam* (2009)
19. P. Lecomte, P.-A. Gauthier, C. Langrenne, A. Berry, A. Garcia, A fifty-node Lebedev grid and its applications to ambisonics. *J. Audio Eng. Soc.* **64**(11) (2016)
20. I.H. Sloan, R.S. Womersley, Extremal systems of points and numerical integration on the sphere. *Adv. Comput. Math.* **21**, 107–125 (2004)
21. B. Bernschütz, C. Pörschmann, S. Spors, Soft-limiting bei modaler amplitudenverstärkung bei sphärischen mikrofonarrays im plane-wave decomposition verfahren, in *Fortschritte der Akustik - DAGA* (2011)
22. T. Rettberg, S. Spors, On the impact of noise introduced by spherical beamforming techniques on data-based binaural synthesis, in *Fortschritte der Akustik - DAGA* (2013)
23. T. Rettberg, S. Spors, Time-domain behaviour of spherical microphone arrays at high orders, in *Fortschritte der Akustik - DAGA* (2014)
24. B. Rafaely, *Fundamentals of Spherical Array Processing*, 1st edn. (Springer, Berlin, 2015)
25. D.L. Alon, B. Rafaely, Spatial decomposition by spherical array processing, in *Parametric Time-Frequency Domain Spatial Audio*, ed. by V. Pulkki, S. Delikaris-Manias, A. Politis (Wiley, New Jersey, 2017)
26. S. Lösler, F. Zotter, Comprehensive radial filter design for practical higher-order ambisonic recording, in *Fortschritte der Akustik – DAGA Nürnberg* (2015)
27. F. Zotter, M. Zaunschirm, M. Frank, M. Kronlachner, A beamformer to play with wall reflections: The icosahedral loudspeaker. *Comput. Music J.* **41**(3) (2017)
28. F. Zotter, M. Frank, C. Haar, Spherical microphone array equalization for ambisonics, in *Fortschritte der Akustik - DAGA* (Nürnberg, 2015)

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.

