Chapter 5
Signal Flow and Effects in Ambisonic Productions

This system offers significantly enhanced spatialisation technologies [...] with new creative possibilities opening up to anyone with the appropriate number of audio channels available on their computer systems.

Dave G. Malham [1], ICMC, Beijing, 1999.

Abstract This chapter presents the internal working principles of various Ambisonic 3D audio effects. No matter which digital audio workstation or processing software is used in a production, the general Ambisonic signal infrastructure is outlined as an important overview of the signal processing chain. The effects presented are frequency-independent effects such as directional re-mapping (mirror, rotation, warping) and re-weighting (directional level modification), and frequency-dependent effects such as widening/distance/diffuseness, diffuse reverberation, and resolution-enhanced convolution reverberation.

The typical audio processing steps for Ambisonic surround-sound signal manipulation are shown in the block diagram Fig. 5.1 from [2]. The description of the multi-venue application in [3] and one for live effects [4] might be encouraging.

Ambisonic encoding and Ambisonic bus. From the previous section we know that representing single-channel signals $s_c(t)$ together with their direction $\theta_c$ is a matter of encoding, of multiplying the signal by the coefficients $y_N(\theta_c)$ obtained by evaluating the spherical harmonics at the direction from which the signal should appear to come. In productions, there will be multiple signals, either representing spot microphones, virtual playback spots of embedded channel-based content (beds), e.g. stereo or 5.1, material. With all input signals encoded and summed up on an Ambisonic bus, we obtain the multi-channel Ambisonic signal representation of an entire audio production.
Ambisonic surround-sound signal. Without decoding to a specific loudspeaker layout, the signal \( \chi_N(t) \) of the Ambisonic bus might appear somewhat virtual. Nevertheless, it allows to be drawn as a surround-sound signal \( x(\theta, t) \) whose amplitude can be evaluated and metered at any direction \( \theta \), anytime \( t \), using the expansion into spherical harmonics

\[
x(\theta, t) = y_T^N(\theta) \chi_N(t).
\]  

(5.2)

**Upmixing.** As first-order recordings are not highly resolved, there are several works on algorithms with resolution enhancement strategies that re-assign time-frequency bins more sharply to directions. A good summary on such input-specific insert effects has been given in the book [5, 6]. Available solutions are DirAC, HOA-DirAC, COMPASS, Harpex.

**Higher order.** Higher-order microphones require more of the acoustic holophonic and holographic basics than presented above, yielding pre-processing filters as input-specific insert effect. Higher-order recording is dealt with in the subsequent Chap. 6 after the derivation of the wave equation and the solutions in the spherical coordinate system.

**Insert effects: Generic re-mapping and leveling.** One can imagine that it should be possible to manipulate the surround-sound signal \( x(\theta, t) \) in various ways. For
instance, effects based on directional re-mapping can take signals out of their original directional range and place them back into the Ambisonic signal at manipulated directions. Also, directions can be altered in amplitude levels so that, for instance, signals at directions with unwanted content undergo attenuation. Many more useful effects are presented below.

**Decoding to loudspeakers/headphones.** To map the modified Ambisonic signal $\tilde{\chi}_N$ to loudspeakers or headphones, an Ambisonic decoder is needed as discussed in the previous chapter. For decoding to headphones, it should be considered to either take only as few HRIR directions to decode to as possible [7, 8], before signals get convolved and mixed to avoid coloration at frontal directions where delays in the HRIRs change too strongly over the direction to get resolved properly [9, 10]. Alternatively, the approach in [11] proposed removal of the HRIR delay at high frequencies and diffuse-field covariance equalization by a $2 \times 2$ filter system, cf. Sect. 4.11.

### 5.1 Embedding of Channel-Based, Spot-Microphone, and First-Order Recordings

Microphone arrays for near-coincident higher-order Ambisonic recording based on holography will be discussed in the subsequent chapter. Nevertheless it possible to use (i) spot and close microphones and encode their direction into the directional panorama, (ii) first-order microphone arrays to fill the Ambisonic channels only up to the first order, (iii) more classical non-coincident or equivalence-stereophonic microphone arrays whose typical playback directions are encoded in Ambisonics.

The study by Kurz et al. [12] investigated how recordings by first-order encoding of the soundfield microphone ST450 and the Oktava MK4012 tetrahedral microphone arrays compare to the equivalence-stereophonic ORTF, see Fig. 5.2. In addition, ORTF-like mapping of the Oktava MK4012’s frontal signals to the $\pm 30^\circ$ directions in 5th order was tested instead of its first-order encoding. Figure 5.3 shows the results of the study in terms of the perceptual attributes localization and spatial depth. It seems that a mixture between ORTF-like 5th-order encoding and first-order encoding of the MK4012 microphone achieves preferred results, while the first-order encoded output of the ST450 Soundfield microphone is rated fair in both attributes, the ORTF microphone only ranked well terms of localization. The results of the ST450 were independent from its orientation, whereas the localization of the first-order-encoded MK4012 was found to be dependent on the orientation. This dependency of the MK4012 is because its microphones are not sufficiently coincident.

As a bottom line of the detailed analysis, one should be encouraged to keep using classical microphone techniques where known to be appropriate and encode their output in higher-order beds or virtual playback directions. However, this should be done with the awareness that stereophonic recording won’t necessarily work for a
Fig. 5.2 Ensemble of the Ambisonic and reference microphones of the study by Kurz et al. [12]; the pixelized microphone prototype by AKG was excluded from the study.

Fig. 5.3 Median values and 95% confidence intervals for each attribute from experiments in [12] for different microphones, orientations, and playback processing.

large audience area, for which the robustness in directional mapping of equivalence-based techniques seem to be attractive.

An interesting layout is, e.g., specified in Hendrickx et al’s work [13], in which they use an equivalence-stereophonic six-channel microphone array. Another interesting idea was used in the ICSA Ambisonics Summer School 2017. A height layer of suitably inclined super-cardioid microphones was added at small vertical distance to the horizontal microphone layer, similarly as the upwards-pointing directional microphones suggested in Lee’s and Wallis’ work [14, 15] to provide sufficiently attenuated horizontal sounds to the height layer.
5.1 Embedding of Channel-Based, Spot-Microphone, and First-Order Recordings

Fig. 5.4  Median values and 95% confidence intervals of listening experiment comparing channel-based orchestra recordings on headphone playback, either directly rendered using the corresponding HRIRs or via binaural Ambisonic decoding of different orders.

Binaural rendering study using surround-with-height material. In another study by Lee, Frank, and Zotter [16], static headphone-based rendering of channel-based recordings was compared using direct HRIR-based rendering or Ambisonics-based binaural rendering, cf. Sect. 4.11. The aim was to find whether differently recorded material could be rendered at high quality via binaural Ambisonics renderers, or under which settings this would imply quality degradation when compared to channel-based binaural rendering.

The results from the half of the listening experiment done in Graz is analyzed in Fig. 5.4, and the renderers compared were channel-based “ref”, a low-passed mono anchor designed to have poor quality “0”, a first-order binaural Ambisonic renderer “1c” based on a cube layout with loudspeakers at ±90°, ±270° azimuth and ±35.3° elevation, and MagLS binaural Ambisonic renderers at the orders “1”, “2”, “3”, “4”, and “5”. Obviously, for orders 2 and above, there is not much quality degradation compared to the reference channel-based binaural rendering. The spatial quality cannot be distinguished from the reference for MagLS with Ambisonic orders 3 and above, and the timbral qualities cannot be distinguished for Ambisonic orders 2 and above.

While this result simplifies the practical requirements for headphone playback remarkably, it can be supposed that due to the limited sweet spot size, loudspeaker playback would still require higher orders, typically.

5.2 Frequency-Independent Ambisonic Effects

Many frequency- and time-independent Ambisonic effects are based on the aforementioned re-mapping of directions and manipulation of directional amplitudes, see e.g. Kronlachner’s thesis, [2, 17]; advanced effects can be found in [18]. In general, the surround-sound signal allows to be manipulated by any thinkable transformation that modifies the directional mapping and amplitude of its contents. The formulation
\[ \tilde{x}(\vec{\theta}, t) = g(\theta) x(\theta, t) \]  

expresses an operation that is able to pick out every direction \( \theta \) of the input signal, weight its signal by a directional gain \( g(\theta) \), and re-map it to a new direction \( \vec{\theta} = \tau(\theta) \) within a transformed signal \( \tilde{x} \). To find out how this affects Ambisonic signals, we write both \( x \) and \( \tilde{x} \) as Ambisonic signals

\[
x(\theta, t) = y^T_N(\theta) \chi_N(t)
\]

expanded in spherical/circular harmonics,

\[
y^T_N(\vec{\theta}) \tilde{\chi}_N(t) = g(\theta) y^T_N(\theta) \chi_N(t),
\]

and use \( \int_{S_d} y_N(\vec{\theta}) y^T_N(\vec{\theta}) d\vec{\theta} = I \) by integrating over \( y_N(\vec{\theta})d\vec{\theta} \) to get \( \tilde{\chi}_N(t) \) on the left

\[
\tilde{\chi}_N(t) = \frac{\int_{S_d} y_N(\vec{\theta}) g(\theta) y^T_N(\theta) d\vec{\theta} \chi_N(t)}{\int_{S_d} y_N(\vec{\theta}) y^T_N(\vec{\theta}) d\vec{\theta}} = T \chi_N(t)
\]

(5.4)

to find the transformed signals being just re-mixed Ambisonic input signals by the matrix \( T \) (note that it might require an increased Ambisonic order \( \tilde{n} \)). Numerical evaluation of the matrix \( T = \int_{S_d} y_N(\vec{\theta}) g(\theta) y^T_N(\theta) d\vec{\theta} \) is best done by using a high-enough \( t \)-design \( \Theta = [\theta_l] \) to discretize the integration variable \( \vec{\theta} = \tau(\theta) \). For the discretized input directions \( \theta \), an inverse mapping \( \theta = \tau^{-1}(\vec{\theta}) \) of the output direction must exist (directional re-mapping must be bijective), so that we can write

\[
T = \int_{S_d} y_N(\vec{\theta}) g(\tau^{-1}(\vec{\theta})) y^T_N(\tau^{-1}(\vec{\theta})) d\vec{\theta} = \frac{4\pi}{L} Y_{N,\Theta} \text{ diag} \{ g_{\tau^{-1}(\Theta)} \} Y^T_{N,\tau^{-1}(\Theta)}.
\]

This formalism is generic and covers simplistic and more complex tasks. It helps understanding that every frequency-independent directional weighting and/or re-mapping is just re-mixing the Ambisonic signals by a matrix, as in Fig. 5.6a.

The ambix VST plugin suite implements several effects, e.g. in the VST plugins ambix_mirror, ambix_rotate, ambix_directional_loudness, ambix_warp. The sections below explain how these and other effects work inside.

### 5.2.1 Mirror

Mirroring does not actually require the generic re-mapping and re-weighting formalism from above, yet. The spherical harmonics associated with the Ambisonic channels are shown in Fig. 4.12 and upon closer inspection one recognizes their symmetries, see Fig. 5.5. To mirror the Ambisonic sound scene with regard to planes of symmetry, it is sufficient to sign-invert channels associated with odd-symmetric spherical harmonics as in Fig. 5.6b. Formally, the transform matrix consists of a diagonal matrix \( T = \text{ diag} \{ c \} \) only, with the corresponding sign-change sequence \( c \).
Ambisonic signals associated with odd symmetric spherical harmonics are sign-inverted to mirror the sound scene. For every Cartesian axis, illustrations above show spherical harmonics up to the third-order, with the order index $n$ organized in rows and the mode index $m$ in columns. Even harmonics are blurred for visual distinction.

**Up-down:** For instance, spherical harmonics with $|m| = n$ are even symmetric with regard to $z = 0$ (up-down), and from this index on, every second harmonic in $m$ is. To flip up and down, it is therefore sufficient to invert the signs of odd-symmetric spherical harmonics with regard to $z = 0$; they are characterized by $n + m$ being an odd number, or $c_{nm} = (-1)^{n+m}$.

**Left-right:** The sin $\phi$-related spherical harmonics with $m < 0$ are odd-symmetric with regard to $y = 0$ (left-right); therefore sign-inverting the signals with the index $m < 0$ exchanges left and and right in the Ambisonic surround signal, i.e. $c_{nm} = (-1)^{m<0}$. 
Front-back: Every odd-numbered $m > 0$ is odd-symmetric with regard to $x = 0$ (front-back), and so is every even-numbered harmonic with $m < 0$. Inverting the sign of these harmonics, $c_{nm} = (-1)^{m+(m<0)}$, flips front and back in the Ambisonic surround signal.

5.2.2 3D Rotation

Rotation can be expressed by a general rotation matrix $R$ consisting of a rotation around $z$ by $\chi$, around $y$ by $\vartheta$, and again around $z$ by $\phi$, see Fig. 5.7. This rotation matrix maps every direction $\theta$ to a rotated direction $\tilde{\theta}$:

$$\tilde{\theta} = R(\varphi, \vartheta, \chi) \theta,$$

$$R = \begin{bmatrix}
\cos(\varphi) & -\sin(\varphi) & 0 \\
\sin(\varphi) & \cos(\varphi) & 0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
\cos(\vartheta) & 0 & -\sin(\vartheta) \\
0 & 1 & 0 \\
\sin(\vartheta) & 0 & \cos(\vartheta)
\end{bmatrix} \begin{bmatrix}
\cos(\chi) & -\sin(\chi) & 0 \\
\sin(\chi) & \cos(\chi) & 0 \\
0 & 0 & 1
\end{bmatrix}.$$

Using this as a transform rule $\tau(\tilde{\theta}) = R\theta$ with neutral gain $g(\theta) = 1$, we find the transform matrix by the inverse mapping $\theta = R^T \tilde{\theta}$ as

Fig. 5.7  zyz-Rotation on the plain example of great-circle navigation of a paper plane around the earth. With the original location at the zenith, a first rotation around $z$ determines the course, and the subsequent rotations around $y$ and $z$ relocate the plane in zenith and azimuth
\[ T = \frac{4\pi}{L} Y_{N,\Theta} Y_{N,R}^T \Theta. \] (5.6)

Using the \( \hat{L} \) directions of a \( t \geq 2N \)-design \( \Theta \) is sufficient to sample the harmonics accurately. With the resulting \( T \), rotation is implemented as in Fig. 5.6a.

There is plenty of potential for simplification: As only the spherical harmonics of a given order \( n \) are required to re-express a rotated spherical harmonic of the same order \( n \), \( T \) is actually block diagonal \( T = \text{blk diag}_n \{ T_n \} \), and within each spherical harmonic order, the integral could be more efficiently evaluated using a smaller \( t \geq 2n \)-design. Moreover, there are various fast and recursive ways to calculate the entries of \( T \) as in [19–25] and implemented in most plugins. And yet, in practice a naïve implementation can be fast enough and pragmatic.

**Rotation around \( z \).** One special case of rotation is important and particularly simple to implement. A directional encoding in azimuth always either is equal to \( \Phi_m(\varphi_s) \) in 2D, or contains it in 3D. For \( m > 0 \), the azimuth encoding \( \Phi_m(\varphi_s) \) depends on \( \cos m\varphi_s \), and its negative-sign version \( \Phi_{-m}(\varphi_s) \) depends on \( \sin(|m|\varphi_s) \). The encoding angle can be offset by the trigonometric addition theorems. They can be written as a matrix:

\[
\begin{bmatrix}
\sin m(\varphi_s + \varphi) \\
\cos m(\varphi_s + \varphi)
\end{bmatrix} =
\begin{bmatrix}
\cos m\varphi & \sin m\varphi \\
-\sin m\varphi & \cos m\varphi
\end{bmatrix}
\begin{bmatrix}
\sin m\varphi_s \\
\cos m\varphi_s
\end{bmatrix}.
\] (5.7)

By this, any Ambisonic signal, be it 2D or 3D, can be rotated around \( z \) by the matrices \( R(m\varphi) \) for the signal pairs with \( \pm m \).

\[
\begin{bmatrix}
\Phi_{-m}(\varphi_s + \varphi) \\
\Phi_m(\varphi_s + \varphi)
\end{bmatrix} = R(m\varphi) \begin{bmatrix}
\Phi_{-m}(\varphi_s) \\
\Phi_m(\varphi_s)
\end{bmatrix},
\]

\[
\begin{bmatrix}
Y_{-m}^{-m}(\varphi_s + \varphi, \vartheta_s) \\
Y_{m}^{m}(\varphi_s + \varphi, \vartheta_s)
\end{bmatrix} = R(m\varphi) \begin{bmatrix}
Y_{-m}^{-m}(\varphi_s, \vartheta_s) \\
Y_{m}^{m}(\varphi_s, \vartheta_s)
\end{bmatrix}.
\] (5.8)

Figure 5.8a shows the processing scheme implementing only the non-zero entries of the associated matrix operation \( T \). Combined with a fixed set of 90° rotations around \( y \) (read from files), it can be used to access all rotational degrees of freedom in 3D [20].

The rotation effect is one of the most important features when using head-tracked interactive VR playback for headphones. Here, rotation counteracting the head movement has the task to support the impression of a static image of the virtual outside world.

### 5.2.3 Directional Level Modification/Windowing

What might be most important when mixing is the option to treat the gains of different directions differently: it might be necessary to attenuate directions of uninteresting or disturbing content while boosting directions of a soft target signal. For such a
Fig. 5.8 Rotation around $z$ and Ambisonic widening/diffuseness apply simple $2 \times 2$ rotation matrices/filter matrices to each Ambisonic signal pair $\chi_{n,m}$, $\chi_{n,-m}$ of the same order $n$. Note that the order of the input/output channels plotted is not the typical ACN sequence to avoid crossing connections and hereby simplify the diagram

manipulation there is a neutral directional re-mapping $\tilde{\theta} = \theta$ and the transform to define the matrix $T$ that is implemented as in Fig. 5.6a remains

$$T = \frac{4\pi}{L} Y_{N, \tilde{\theta}} \text{diag}(g_{\tilde{\theta}}) Y_{N, \tilde{\theta}}^T.$$  

(5.9)

In the simplest version, as implemented in `ambix_directional_loudness`, the gain function just consists of two mutually exclusive regions, e.g. within a region of diameter $\alpha$ around the direction $\theta_g$, and a complementary region outside, with separately controlled gains $g_{\text{in}}$ and $g_{\text{out}}$:

$$g(\theta) = g_{\text{in}} u(\theta^T \theta_g - \cos \frac{\alpha}{2}) + g_{\text{out}} u(\cos \frac{\alpha}{2} - \theta^T \theta_g),$$  

(5.10)

where $u(x)$ represents the unit-step function that is 1 for $x \geq 0$ and 0 else. Note that the Ambisonic order of this effect will need to be larger to be lossless. However, with reasonably chosen sizes $\alpha$ and gain ratios $g_{\text{in}}/g_{\text{out}}$, the effect will nevertheless produce reasonable results. Figure 5.9 shows a window at azimuth and elevation at 22.5° with an aperture of 50° using $g_{\text{in}} = 1$ and $g_{\text{out}} = 0$ and the order of $N = 10$ with a grid of encoded directions to illustrate the influence of the transformation.

For reference: entries of the tensor used to analytically re-expand the product of two spherical functions $x(\theta) g(\theta)$ given by their spherical harmonic coefficients $\chi_{nm}$, $\gamma_{nm}$ are called Gaunt coefficients or Clebsh-Gordan coefficients [6, 26].
5.2 Frequency-Independent Ambisonic Effects

Fig. 5.9 Directionally windowed Ambisonic test image at every 90° in azimuth, interleaved in azimuth for the elevations ±60° and ±22.5°, using the order \( N = \tilde{N} = 10 \), a window size of \( \alpha = 50° \) around azimuth and elevation of 0° and max-\( r_E \) weighting.

5.2.4 Warping

Gerzon [27, Eq. 4a] described the effect *dominance* that is meant to warp the Ambisonic surround scene to modify how vitally the essential parts in front of the scene are presented.

_Warping wrt. a direction._ For mathematical simplicity, we describe this bilinear warping with regard to the \( z \) direction. To warp with regard to the frontal direction, one first rotates the front upwards, applies the warping operation there, and then rotates back. The bilinear warping modifies the normalized \( z \) coordinate \( \zeta = \cos \vartheta = \theta_z \) so that signals from the horizon \( \zeta = 0 \) are pulled to \( \tilde{\zeta} = \alpha \),

\[
\tilde{\zeta} = \frac{\alpha + \zeta}{1 + \alpha \zeta},
\]

while keeping for the poles \( \tilde{\zeta} = \pm 1 \) what was originally there \( \zeta = \pm 1 \). Hereby, the surround signal gets squeezed towards or stretched away from the zenith, or when rotating before and after: towards/from any direction.

The integral can be discretized and solved by a suitable \( t \)-design as before, only that for lossless operation, the output order \( \tilde{N} \) must be higher than the input order \( N \). We get a matrix \( T \) that is implemented as in Fig. 5.6a is computed by

\[
T = \frac{4\pi}{L} Y_{\tilde{N},\Theta} \text{diag}(g_{\tau^{-1}(\Theta)}) \tilde{Y}_{\tilde{N},\tau^{-1}(\Theta)}^T,
\]

(5.12)

The inverse mapping yields

\[
\zeta = \tau^{-1}(\tilde{\zeta}) = \frac{\tilde{\zeta} - \alpha}{1 - \alpha \zeta},
\]

(5.13)

and it modifies the coordinates of the \( t \)-design inserted for \( \tilde{\theta}_l = \theta_l = [\theta_{x,l}, \theta_{y,l}, \theta_{z,l}]^T \) with \( \tilde{\zeta}_l = \theta_{z,l} \) accordingly.
Fig. 5.10 Warping of the horizontal plane by 22.5° downwards; original Ambisonic test image contains points at every 90° in azimuth, interleaved in azimuth for the elevations ±60° and ±22.5°; orders are $N = \tilde{N} = 10$; max-$r_E$ weighted

\[
\tau^{-1}\{\Theta_l\} = \begin{bmatrix}
\theta_{x,l} \sqrt{1 - (\tau^{-1}\{\zeta_l\})^2} \\
\theta_{y,l} \sqrt{1 - (\tau^{-1}\{\zeta_l\})^2} \\
\tau^{-1}\{\zeta_l\}
\end{bmatrix} = \begin{bmatrix}
\theta_{x,l} \sqrt{1 - \left(\frac{\theta_{x,l} - \alpha}{1 - \alpha \theta_{x,l}}\right)^2} \\
\theta_{y,l} \sqrt{1 - \left(\frac{\theta_{x,l} - \alpha}{1 - \alpha \theta_{x,l}}\right)^2} \\
\frac{\theta_{x,l} - \alpha}{1 - \alpha \theta_{x,l}}
\end{bmatrix}.
\] (5.14)

The gain $g(\hat{\zeta})$ of the generic transformation is useful to preserve the loudness of what becomes wider and therefore louder in terms of the $E$ measure after re-mapping. To preserve loudness, the resulting surround signal is divided by the square root of the stretch applied, which is related to the slope of the mapping by $\frac{1}{g} = \sqrt{\frac{d\zeta}{d\hat{\zeta}}}$. Expressed as de-emphasis gain, we get

\[
g(\hat{\zeta}_l) = \frac{1 - \alpha \hat{\zeta}_l}{\sqrt{1 - \alpha^2}} = \frac{1 - \alpha \theta_{l,z}}{\sqrt{1 - \alpha^2}}.
\] (5.15)

Figure 5.10 shows warping of the horizontal plane by 20° downwards, using the test image parameters as with windowing; de-emphasis attenuates widened areas.

In the same fashion, Kronlachner [17] describes another warping curve that warps with regard to fixed horizontal plane and pole, either squeezing or stretching the content towards or away from the horizon, symmetrically for both the upper and lower hemispheres (second option of the ambient_warp plugin).

### 5.3 Parametric Equalization

There are two ways of employing parametric equalizers to Ambisonic channels. Either a single-/multi-channel input of a mono-encoder or a multiple-input encoder is filtered by parametric equalizers. Or each of the Ambisonic signal’s channels is filtered by the same parametric equalizer, see Fig. 5.11a.
5.3 Parametric Equalization

Fig. 5.11 Block diagram of processing that commonly and equally affects all Ambisonic signals, such as parametric equalization and dynamic processing (compression), without recombining the signals

Bass management is often important to not overdrive smaller loudspeaker systems of, e.g., a 5th-order hemispherical playback system with subwoofer signals: All 36 channels from the Ambisonic bus can be sent to a decoder section, in which frequencies below 70–100 Hz are high-cut by a 4th-order filter before running through the Ambisonic decoder, while the first channel from the Ambisonics bus alone, the omnidirectional channel, is being sent to a subwoofer section, in which a 4th-order filter high-cut removes the high frequencies above 70–100 Hz before the signal is sent to the subwoofers. If the playback system is time-aligned between subwoofer and higher frequencies, the 4th-order crossovers should be Linkwitz–Riley filters (either squared Butterworth high-pass or low-pass filters) to preserve phase equality [28].

For more information on parametric equalizers, the reader is referred to Udo Zölzer’s book on Digital audio effects [29].

5.4 Dynamic Processing/Compression

Individual compression of different Ambisonic channels would destroy the directional consistency of the Ambisonics signal. Consequently, dynamic processing should rather affect the levels of all Ambisonic channels in the same way. As it typically contains all the audio signals, it is useful to have the first, omnidirectional Ambisonic channel control the dynamic processor as side-chain input, see Fig. 5.11b. For more information on dynamic processing, the reader is referred to Udo Zölzer’s book on Digital audio effects [29].

Moreover, it is sometimes useful to compress the vocals of a singer separately. To this end, the directional compression would first extract a part of the Ambisonic signals by a directional window, creating one set of Ambisonic signals without the directional region of the window, and another one exclusively containing it. The compression is applied on the resulting window signal before re-combining it with the rest signals.
5.5 Widening (Distance/Diffuseness/Early Lateral Reflections)

Basic widening and diffuseness effects can be regarded as being inspired by Gerzon [30] and Laitinen [31] who proposed to apply frequency-dependent panning filters, mapping different frequencies to directions dispersed around the panning direction. The resulting effect is fundamentally different from and superior to increasing the spread in frequency-independent MDAP with enlarged spread or Ambisonics with reduced order, which could yield audible comb filtering.

To apply this technique to Ambisonics, Zotter et al. [32] proposed to employ a dispersive, i.e. frequency-dependent, rotation of the Ambisonic scene around the $z$-axis as in Eq. (5.8) by the matrix $R$ as described above and in Fig. 5.8b, using $2 \times 2$ matrices of filters to implement the frequency-dependent argument $m\hat{\phi}\cos \omega \tau$

$$R(m\hat{\phi}\cos \omega \tau) = \begin{bmatrix} \cos(m\hat{\phi}\cos \omega \tau) & \sin(m\hat{\phi}\cos \omega \tau) \\ -\sin(m\hat{\phi}\cos \omega \tau) & \cos(m\hat{\phi}\cos \omega \tau) \end{bmatrix}, \quad (5.16)$$

whose parameters $\hat{\phi}$ and $\tau$ allow to control the magnitude and change rate of the rotation with increasing frequency. How this filter matrix is implemented efficiently was described in [33], where a sinusoidally frequency-varying pair of functions

$$g_1(\omega) = \cos [\alpha \cos (\omega \tau)], \quad g_2(\omega) = \sin [\alpha \cos (\omega \tau)], \quad (5.17)$$

was found to correspond to the sparse impulse responses in the time domain

$$g_1(t) = \sum_{q=-\infty}^{\infty} J_{|q|}(\alpha) \cos(\frac{\pi}{2} |q|) \delta(t - q \tau) \quad (5.18)$$

$$g_2(t) = \sum_{q=-\infty}^{\infty} J_{|q|}(\alpha) \sin(\frac{\pi}{2} |q|) \delta(t - q \tau),$$

allowing for truncation to just a few terms in $q$, typically 11 taps between $-5 \leq q \leq 5$ or fewer, and hereby an efficient implementation. For the implementation of the filter matrix, for each degree $m$, the value $\alpha = m\hat{\phi}$. (It might be helpful to be reminded of a phase-modulated cosine and sine from radio communication, whose spectra are the same functions as this impulse response pair.)

As the algorithm places successive frequencies at slightly displaced directions, the auditory source width increases. Moreover, the frequency-dependent part causes a smearing of the temporal fine-structure in the signal. In [34], it was found that implementations discarding the negative values of $q$, i.e. keeping $q \geq 0$ sound more natural and still exhibit a sufficiently strong effect. Time constants $\tau$ around 1.5 ms yield a widening effect, and a diffuseness and distance impression is obtained with $\tau$ around 15 ms. The parameter $\hat{\phi}$ is adjustable between 0 (no effect) and larger values.
Beyond 80° the audio quality starts to degrade. The use as diffusing effect has turned out to be useful as simple simulation of early lateral reflections, because most parts of the spectrum are played back near the reversal points ±\(\phi\) of the dispersion contour. For naturally sounding early reflections, additional shelving filters introducing attenuation of high frequencies prove useful.

Figures 5.12 and 5.13 show experimental ratings of the perceived effect strength (width or distance) of the above algorithm in [34], which was implemented as frequency-dependent (dispersive) panning on just a few loudspeakers \(L = 3, 4, 5, 7\) evenly arranged from −90° to 90° on the horizon at 2.5 m distance from the central listening position. Loudspeakers were controlled by a sampling decoder of the orders \(N = 1, 2, 3, 5\) with the center of the max-\(r_E\)-weighted panning direction at 0° in front. The signal was speech and as a reference it used the frontal loudspeaker with the unprocessed signal “REF”. The experiment tested the algorithm with both the symmetric impulse responses suggested by Eq. (5.18), and such truncated to their causal \(q \geq 0\)-side, for a listening position at the center of the arrangement (bullet marker) and at 1.25 m shifted to the right, off-center (square marker). Figure 5.12 indicates for the widening algorithm with \(\tau = 1.5\) ms that the perceived width saturates above \(N > 2\) at both listening positions. Despite the effect of the causal-sided
implementation is weaker in effect strength, it highly outperforms the symmetric FIR implementation in terms audio quality (right diagram), while still producing a clearly noticeable effect when compared to the unprocessed reference (left diagram).

A more pronounced preference of the causal-sided implementation in terms of audio quality is found in Fig. 5.13 for the setting \( \tau = 15 \text{ ms} \), where the algorithm is increasing the diffuseness or perceived distance for orders \( N > 2 \) at both listening positions.

### 5.6 Feedback Delay Networks for Diffuse Reverberation

Feedback delay networks (FDN, cf. [35, 36]) can directly be employed to create diffuse Ambisonic reverberation. A dense response and an individual reverberation for every encoded source can be expected when feeding the Ambisonic signals directly into the inputs of the FDN.

As in Fig. 5.14, FDNs consists of a matrix \( A \) that is orthogonal \( A^T A = I \) and should mix the signals of the feedback loop well enough to distribute them across all different channels to couple the resonators associated with the different delays \( \tau_i \). These delays should not have common divisors to avoid pronounced resonance frequencies, and are therefore typically chosen to be related to prime numbers. Small delays are typically selected to be more closely spaced \( \{2, 3, 5, \ldots\} \) ms to simulate a diffuse part with densely spaced response at the beginning, and long delays further apart often make the reverberation more interesting. Using unity factors as channel gains \( g_{\tau_i}^{lo}, g_{\tau_i}^{mi}, g_{\tau_i}^{hi} = 1 \) and any orthogonal matrix \( A \), the reverberation time becomes infinite. For smaller channel gains, the FDN produces decaying output.

Reverberation is characterized by the exponentially decaying envelope \( 10^{-3 \frac{\tau_i}{T_{60}}} \). For a single delay of the length \( \tau_i \), the corresponding gain is \( g^{\tau_i} \) with \( g = 10^{-\frac{\tau_i}{T_{60}}} \). This factor with the corresponding exponent provides equal reverberation decay rate in every channel, and hereby exact control of the reverberation time. To make the effect sound natural, it is typical to adjust the gains within a high-mid-low filter set to decrease the reverberation towards higher frequency bands by the gains \( g_{hi}^{\tau_i} \leq g_{mi}^{\tau_i} \leq g_{lo}^{\tau_i} \).

The vector gathering the current sample for every feedback path is multiplied by the matrix \( A \). For calculation in real-time, Rocchesso proposed to use a scaled Hadamard matrix \( A = \frac{1}{M} H \) of the dimensions \( M = 2^k \) in [37]. It consists of \( \pm 1 \) entries only and hereby perfectly mixes the signal across the different feedbacks to create a diffuse set of resonances. What is more, this not only replaces the \( M \times M \) multiply and adds of matrix multiplication multiplies by sums and differences, it is
5.6 Feedback Delay Networks for Diffuse Reverberation

Fig. 5.14 Feedback delay network (FDN) for Ambisonic reverb. The matrix $A$ is unitary and the gain $g = 10^{-\frac{1}{60}}$ to the power of the delay $\tau_i$ allows to adjust a spatially and temporally diffuse reverberation effect in different bands (lo, mi, hi) moreover equivalent to the efficient implementation as Fast Walsh-Hadamard Transform (FWHT), a butterfly algorithm. Figure 5.15 shows a graphical implementation example of a 16-channel FWHT in the real-time signal processing environment Pure Data.

Fig. 5.15 The fast Walsh-Hadamard transform (FWHT) variant implemented in the 16-channel feedback delay network reverberator [rev3~] in Pure Data requires only $4 \times 16$ sums/differences to replace the $16 \times 16$ multiplies of matrix multiplication by $A$. 
5.7 Reverberation by Measured Room Impulse Responses and Spatial Decomposition Method in Ambisonics

The first-order spatial impulse response of a room at the listener can be improved by resolution enhancements of the spatial decomposition method (SDM) by Tervo [38], which is a broad-band version of spatial impulse response rendering (SIRR) by Merimaa and Pulkki [39, 40]. For reliable measurements, typically loudspeakers are employed, and the typical measurement signals aren’t impulses, but swept-sine signals that can are reverted to impulses by deconvolution. A room impulse response is typically sparse in its beginning whenever direct sound and early reflections arrive at the measurement location. Generally, it is likely that those arrival times in the early part do not coincide and are well separated from each other, so that one can assume their temporal disjointness at the receiver.

From a room impulse response $h(t)$ that complies with this assumption, for which there consequently is a direction of arrival (DOA) $\theta_{\text{DOA}}(t)$ for every time instant, one could construct an Ambisonic receiver-directional room impulse response as in [41]

$$h(\theta_R, t) = h(t) \delta[1 - \theta_R^T \theta_{\text{DOA}}(t)],$$

depending on the direction $\theta_R$ at the receiver. This response can be transformed into the spherical harmonic domain by integrating it over $y_N(\theta_R) d\theta_R$, to get the set of $N$th-order Ambisonic room impulse responses

$$\tilde{h}_N(t) = h(t) y_N[\theta_{\text{DOA}}(t)].$$

A signal $s(t)$ convolved by this vector of impulse responses theoretically generates a 3D Ambisonic image of the mono sound in the room of the measurement. This can be done, e.g., by the plug-in mcfx_convolver. Now there are two problems to be solved: (i) how to estimate $\theta_{\text{DOA}}(t)$, (ii) how to deal with the diffuse part of $h(t)$, when there are more sound arrivals at a time than one.

Estimation of the DOA. One could now just detect the temporal peaks of the room impulse response and assign the guessed evolution of the direction of arrival as suggested in [42], and hereby span the envelopment of the room impulse response. Alternatively, if the room impulse response was recorded by a microphone array as in [38], array processing can be used to estimate the direction of arrival $\theta_{\text{DOA}}(t)$. For first-order Ambisonic microphone arrays, when suitably band-limited to the frequency range in which the directional mapping is correct, e.g. between 200 Hz and 4 kHz, the vector $r_{\text{DOA}}$ of Eq. (A.83) in Appendix A.6.2 yields a suitable estimate

$$\tilde{r}_{\text{DOA}}(t) = W(t) \begin{bmatrix} X(t) \\ Y(t) \\ Z(t) \end{bmatrix} = -\rho c h(t) v(t), \quad \theta_{\text{DOA}}(t) = \frac{\tilde{r}_{\text{DOA}}(t)}{\|\tilde{r}_{\text{DOA}}(t)\|}. \quad (5.19)$$

Figure 5.16 shows the directional analysis of the first 100 ms of a first-order directional impulse response taken from the openair lib\(^1\) This response was measured in St. Andrew’s Church Lyddington, UK (2600 m\(^3\) volume, 11.5 m source-receiver distance) with a Soundfield SPS422B microphone.

\(^1\)http://www.openairlib.net.
5.7 Reverberation by Measured Room Impulse Responses

The direct sound from the front is clearly visible, as well as strong early reflections from front and back, and equally distributed weak directions from the diffuse reverb.

**Spectral decay recovery for higher-order RIRs.** The second task mentioned above is that the multiplication of $h(t)$ by $y_N[\theta_{DOA}(t)]$ to obtain $\hat{h}(t)$ degrades the spectral decay at higher orders. If there is no further processing, the resulting response typically exhibits a noticeable increased spectral brightness [38, 41, 43]. This unnatural brightness mainly affects the diffuse reverberation tail, where *temporal disjointness* is a poor assumption. There, the corresponding rapid changes of $\theta_{DOA}(t)$ cause a strong amplitude modulation in the pre-processing of the late room impulse response at high Ambisonic orders. Typically, long decays of low frequencies leak into high frequencies, and hereby result in an erroneous spectral brightening of the diffuse tail. Figure 5.17 analyses the behavior in terms of an erroneous increase of reverberation time at high frequencies, especially when using high orders.

In order to equalize the spectral decay and hereby the reverberation time of the SDM-enhanced impulse response, there is a helpful pseudo-allpass property of the spherical harmonics for direct and diffuse fields, as described in Eqs. (A.52) and (A.55) of Appendix A.3.7. The signals in the vector $\hat{h}(t) = [\hat{h}_n^m(t)]_{in}$ are first decomposed into frequency bands, yielding the sub-band responses $\hat{h}_n^m(t, b)$. We can equalize the spectral sub-band decay for every band $b$ and order $n$ by targeting fulfillment of the pseudo-allpass property
\[ \sum_{m=-n}^{n} E\{ |h_n^m(t,b)|^2 \} = (2n + 1)E\{ |h_0^0(t,b)|^2 \}. \quad (5.20) \]

The formulation above relies on the correct spectral decay of the omnidirectional signal \( h_0^0(t,b) = \tilde{h}_0^0(t,b) \), which is unaffected by modulation. Correction is achieved by

\[ h_n^m(t,b) = \tilde{h}_n^m(t,b) \frac{(2n + 1) E\{ |\tilde{h}_0^0(t,b)|^2 \}}{\sum_{m=-n}^{n} E\{ |\tilde{h}_n^m(t,b)|^2 \}}, \quad (5.21) \]

here, the expression \( E\{ |\cdot|^2 \} \) refers to estimation of the squared signal envelope.

**Perceptual evaluation.** Frank’s 2016 experiments \cite{44} measuring the area of the sweet spot also investigated the plausibility of reverberation created by their Ambisonically SDM-processed measurements at different order settings, \( N = 1, 3, 5 \). For Fig. 5.18b listeners indicated at which distance from the room’s center they heard that envelopment began to collapse to the nearest loudspeakers. One can observe that rendering diffuse reverberation for a large audience benefits from a high Ambisonic order. Moreover, experiments in \cite{43} revealed an improvement of the perceived spatial depth mapping, i.e. a clearer separation between foreground and background sound for the SDM-processed higher-order reverberation, cf. Fig. 1.21b.

![Fig. 5.18](image-url) The perceptual sweet spot size as investigated by Frank \cite{44} for SDM processed RIRs cover an area in IEM CUBE that increases with the SDM order \( N \) chosen (black = 5th, gray = 3rd, light gray = 1st order Ambisonics). In comparison to panned direct sound, one should keep some distance to the loudspeakers to avoid breakdown of envelopment.
5.8 Resolution Enhancement: DirAC, HARPEX, COMPASS

The concept of parametric audio processing [5] describes ways to obtain resolution-enhanced first-order Ambisonic recordings by parametric decomposition and rendering. One main idea is to decompose short-term stationary signals of a sound scene into a directional and a less directional diffuse stream.

For synthesis of the directional part based on mono signals, it is clear how to obtain the most narrow presentations by amplitude panning or higher-order Ambisonic panning of consistent $r_E$ vector predictions as in Chap. 2.

The synthesis of diffuse and enveloping parts based on a mono signal can require extra processing such as either widening/diffuseness effects or reverberation as in Sects. 5.5 and 5.6, which both also provide a directionally wide distribution of sound. Or more practically, the recording itself could deliver sufficiently many uncorrelated instances of the diffuse sound to be played back by surrounding virtual sources. Envelopment and diffuseness is based on providing a consistently low interaural covariance or cross correlation of sufficiently high decorrelation.

**DirAC.** A main goal of DirAC (Directional Audio Coding [5]) is finding signals and parameters for sound rendering by analyzing first-order Ambisonic recordings. One variant is to use the intensity-vector-based analysis in the short-term Fourier transform (STFT), see also Appendix A.6.2:

$$r_{DOA}(t, \omega) = -\rho c \Re\{p(t, \omega)\ast v(t, \omega)\} = \frac{\Re\{W(t, \omega)\ast[X(t, \omega), Y(t, \omega), Z(t, \omega)]^T\}}{\sqrt{2}|W(t, \omega)|^2},$$

(5.22)

which can be treated similarly as the $r_E$ vector, regarding direction and diffuseness $\psi = 1 - ||r_{DOA}||^2$.

**Single-channel DirAC** is Ville Pulkki’s original way to decompose the $W(t, \omega)$ signal in the STFT domain into a directional signal $\sqrt{1 - \psi} W(t, \omega)$ that is synthesized by amplitude panning and a diffuse signal $\sqrt{\psi} W(t, \omega)$ to be synthesized diffusely [45].

**Virtual-microphone DirAC** uses a first-order Ambisonic decoder to the given loudspeaker layout and time-frequency-adaptive sharpening masks increasing the focus of direct sounds, see Vilkamo [46] and [5, Ch. 6], or order e.g. Sect. 5.2.3. Playback of diffuse sounds benefits from an optional diffuseness effect.

**HARPEX** (high angular-resolution plane-wave expansion [47]) is Svein Berge’s patented solution to optimally decode sub-band signals. It is based on the observation he made with Natasha Barrett that decoding to a tetrahedral loudspeaker layout is perceptually outperforming if the tetrahedron nodes are rotationally aligned with the sources of the recording. HARPEX accomplishes convincing diffuse and direct sound reproduction by decoding to a variably adapted virtual loudspeaker layout in every sub band. The layout is adaptively rotation-aligned with sources detected in the band. HARPEX is typically described using an estimator for direction pairs.
**COMPASS** (COding and Multidirectional Parameterization of Ambisonic Sound Scenes [48]) by Archontis Politis can be seen as an extension of DirAC. In contrast to DirAC, it tries to detect and separate multiple direct sound sources from the ambient or background sound. This is done by applying two different kinds of beamformers: one that contains only the direct sound for each sound source (source signals) and one that contains everything but the direct sound (ambient signal). Similar as before, the source signals are reproduced using amplitude panning and the ambient signal is sent to the decorrelator. In contrast to DirAC, COMPASS is not limited to first-order input but can also enhance the spatial resolution of higher-order inputs.

### 5.9 Practical Free-Software Examples

#### 5.9.1 IEM, ambix, and mcfx Plug-In Suites

The ambix_converter is an important tool when adapting between the different Ambisonic scaling conventions, e.g. the standard SN3D normalization that uses only \( \sqrt{\frac{(n-|m|)!}{(n+|m|)!}} \cdot \frac{2-\delta_m}{4\pi} \) for normalization instead of the full \( \sqrt{\frac{(n-|m|)!}{(n+|m|)!}} \cdot \frac{(2-\delta_m)(2n+1)}{4\pi} \) that is called N3D, see Fig. 5.19. This alternating definition is because of a practical choice of the ambix format [49] to avoid high-order channels becoming louder than the zeroth-order channel. Also it permits to adapt between channel sequences such as ACN’s \( i = n^2 + n + m \) or SID’s \( i = n^2 + 2(n - |m|) + (m < 0) \). It is advisable to use test recordings with the main directions, e.g. front, left, top, and to check that the channel separation for decoded material is roughly exceeding 20 dB for 5th-order material. Moreover, it contains inversion of the Condon-Shortley phase that typically causes a 180° rotation around the \( z \) axis, and it contains the left-right, front-back, and top-bottom flips discussed in the mirroring operations above.
The ambix_warping plugin, see Fig. 5.20, implements the above-mentioned warping operations shifting horizontal sounds towards one of the poles, or into both polar directions. Warping can be applied to any other direction than zenith and nadir when placing it between two mutually inverting ambix_rotation or IEM SceneRotator objects that intermediate rotate zenith to another direction.

The IEM SceneRotator as the ambix_rotation plugin can be controlled by head tracking and is essential for an immersive headphone-based experience, see Fig. 5.21. Its processing is done as described above.

The ambix_directional_loudness plugin in Fig. 5.22 implements the above-mentioned directional amplitude window in either circular or equi-rectangular spherical shape. Several of these windows can be made, soloed, and remote controlled, each one of which allowing to set a gain for the inside and outside region. This is often useful in practice, when, e.g., reinforcing or attenuating desired or undesired signal parts within an Ambisonic scene.

Fig. 5.20 ambix_warping plug-in in Reaper

Fig. 5.21 IEM SceneRotator and ambix_rotator plug-ins
To observe the changes made to the Ambisonic scene, the IEM Energy Visualizer can be helpful, see Fig. 5.23.

If, for instance, the Ambisonic scene requires dynamic compression, as outlined in the section above, the IEM OmniCompressor is a helpful tool. It uses the omnidirectional Ambisonic channel to derive the compression gains (as a side-chain for all other Ambisonic channels). Similarly as the directional_loudness plug-in, the IEM DirectionalCompressor allows to select a window, but this time for setting different dynamic compression within and outside the selected window, see Fig. 5.24.

The multichannel mcfx_filter plugin in Fig. 5.25 does not only implement a set of parametric equalizers, a low- and high cut that can be toggled between filter skirts
of either 2nd or 4th order, but it also features a real-time spectrum analyzer to observe the changes done to the signal. It is not only practical for Ambisonic purposes, it’s just a set of parametric filters that is equally applied to all channels and controlled from one interface.

The mcfx_convolver plug-in in Fig. 5.26 is useful for many purposes, also scientific ones, e.g., when testing binaural filters or driving multi-channel arrays with filters, etc. Its configuration files use the jconvolver format that specifies which filter file (typically stored in multi-channel wav files) connects which of its multiple inlets to which of its multiple outlets. It is also used to implement the SDM-based reverberation described in the above sections.

For a cheaper reverberation network, the IEM FDNReverb network described above can be used, see Fig. 5.27. It is not in particular an Ambisonic tool, but can
be used in any multi-channel environment. The particularity of the implementation in the IEM suite is that a slow onset can be adjusted.

The ambix_widening plug-in in Fig. 5.28 implements the widening by frequency-dependent, dispersive rotation of the Ambisonic scene around the z axis as described above. It can also be used to cheaply stylize lateral reflections instead of the IEM RoomEncoder (Fig. 4.36) with time constant settings exceeding 5 ms, or just as a
widening effect. The setting single-sided permits to suppress the slow attack of the Bessel sequence.

Another tool is quite helpful, the mcfx_gain_delay plug-in in Fig. 5.29. It permits to solo or mute individual channels, as well as delay and attenuate them differently. What is more and often even more useful: It is invaluably helpful for testing the signal chain, as one can step through the channels with different signals.

5.9.2 Aalto SPARTA

The SPARTA plug-in suite by Aalto University provides Ambisonic tools for encoding, decoding on loudspeakers and headphones, as well as visualization. A special feature is the COMPASS decoder plug-in Fig. 5.30 that can increase the spatial resolution of first-, second-, and third-order recordings. Playback can be done either
on arbitrary loudspeaker arrangements or their virtualization on headphones. The signal-dependent parametric processing allows to adjust the balance between direct and diffuse sound in each frequency band. In order to suppress artifacts due to the processing, the parametric playback (Par) can be mixed with the static decoding (Lin) of the original recording. While it is advisable to keep the parametric contribution below \( \frac{2}{3} \) for noticeable directional improvements and low artifacts, in general, in recordings with cymbals or hi-hats it is advisable to fade towards Lin starting at around 4 kHz.

### 5.9.3 Røde

The Soundfield plug-in by Røde in Fig. 5.31 was originally designed to process the signals from the four cardioid microphone capsules of their Soundfield microphone. However, it also supports first-order Ambisonics as input format. It can decode to various loudspeaker arrangements by placing virtual microphones into the directions of the loudspeakers. The directivity of each virtual microphone can be adjusted between first-order cardioid and hyper-cardioid. Moreover, higher-order directivity patterns are possible using a parametric signal-dependent processing, resulting in an increase of the spatial resolution.
Fig. 5.31 Soundfield by Røde plug-in

References

1. D.G. Malham, Higher-order ambisonic systems for the spatialisation of sound, in *Proceedings of the ICMC* (Beijing, 1999)
2. M. Frank, F. Zotter, A. Sontacchi, Producing 3D audio in ambisonics, in *57th International AES Conference, LA, The Future of Audio Entertainment Technology? Cinema, Television and the Internet* (2015)
3. M. Frank, A. Sontacchi, Case study on ambisonics for multi-venue and multi-target concerts and broadcasts. J. Audio Eng. Soc. 65(9) (2017)
4. D. Rudrich, F. Zotter, M. Frank, Efficient spatial ambisonic effects for live audio, in *29th TMT*, Cologne, November 2016
5. V. Pulkki, S. Delikaris-Manias, A. Politis (eds.), *Spatial Decomposition by Spherical Array Processing* (Wiley, New Jersey, 2017)
6. A. Politis, Microphone array processing for parametric spatial audio techniques, Ph.D. dissertation, Aalto University, Helsinki (2016)
7. B. Bernschütz, Microphone arrays and sound field decomposition for dynamic binaural recording, Ph.D. Thesis, TU Berlin (TH Köln) (2016), http://dx.doi.org/10.14279/depositonce-5082
8. B. Bernschütz, A.V. Giner, C. Pörschmann, J. Arend, Binaural reproduction of plane waves with reduced modal order. Acta Acustica u. Acustica 100(5) (2014)
9. 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)
10. C. Andersson, Headphone auralization of acoustic spaces recorded with microphone arrays, M. Thesis, Chalmers TU, Gothenburg (2017)
11. 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)
12. E. Kurz, F. Pfahler, M. Frank, Comparison of first-order ambisonic microphone arrays, in International Conference, Spatial Audio (ICSA) (2015)
13. E. Hendrickx, P. Stitt, J.-C. Messonier, J.-M. Lyzwa, B.F. Katz, C. Boishéraud, Influence of head tracking on the externalization of speech stimuli for non-individualized binaural synthesis. J. Ac. Soc. Am. 141(3), 2011–2023 (2017)
14. H. Lee, C. Gribben, Effect of vertical microphone layer spacing for a 3d microphone array. J. Audio Eng. Soc. 62(12), 870–884 (2014)
15. R. Wallis, H. Lee, The reduction of vertical interchannel cross talk: The analysis of localisation thresholds for natural sound sources, MDPI Sci. 7(3) (2017)
16. H. Lee, M. Frank, F. Zotter, Spatial and timbral fidelities of binaural ambisonics decoders for main microphone array recordings, in AES IIA Conference (York, 2019)
17. M. Kronlachner, F. Zotter, Spatial transformations for the enhancement of ambisonic recordings, in Proceedings of the ICSA, Erlangen (2014)
18. M. Hafsati, N. Epain, J. Daniel, Editing ambisonic sound scenes, in 4th International Conference, Spatial Audio (ICSA) (Graz, 2017)
19. G. Aubert, An alternative to wigner d-matrices for rotating real spherical harmonics. AIP Adv. 3(6), 062121 (2013). https://doi.org/10.1063/1.4811853
20. F. Zotter, Analysis and synthesis of sound-radiation with spherical arrays, PhD. Thesis, University of Music and Performing Arts, Graz (2009)
21. D. Pinchon, P.E. Hoggan, Rotation matrices for real spherical harmonics: general rotations of atomic orbitals in space-fixed axes. J. Phys. A 40, 1751–8113 (2007)
22. J. Ivanic, K. Ruedenberg, Rotation matrices for real spherical harmonics. Direct determination by recursion. J. Phys. Chem. A 100, 6342–6347 (1996)
23. J. Ivanic, K. Ruedenberg, Additions and corrections. J. Phys. Chem. A 102, 9099–9100 (1998)
24. C.H. Choi, J. Ivanic, M.S. Gordon, K. Ruedenberg, Rapid and stable determination of rotation matrices between spherical harmonics by direct recursion. J. Chem. Phys. 111, 8825–8831 (1999)
25. N.A. Gumerov, R. Duraiswami, Recursions for the computation of multipole translation and rotation coefficients for the 3-d helmholtz equation. SIAM 25, 1344–1381 (2003)
26. J.R. Driscoll, D.M.J. Healy, Computing fourier transforms and convolutions on the 2-sphere. Adv. Appl. Math. 15, 202–250 (1994)
27. M.A. Gerzon, G.J. Barton, Ambisonic decoders for HDTV, in prepr. 3345, 92nd AES Convention (Vienna, 1992)
28. S.P. Lipshitz, J. Vanderkoy, In-phase crossover network design. J. Audio Eng. Soc. 34(11), 889–894 (1986)
29. U. Zölzer, Digital Audio Effects, 2nd edn. (Wiley, New Jersey, 2011)
30. M.A. Gerzon, Signal processing for simulating realistic stereo images, in prepr. 3423, Convention Audio Engineering Society (San Francisco, 1992)
31. M.-V. Laitinen, T. Pihlajamäki, C. Erkut, V. Pulkki, Parametric time-frequency representation of spatial sound in virtual worlds. ACM TAP 9(2) (2012)
32. F. Zotter, M. Frank, M. Kronlachner, J. Choi, Efficient phantom source widening and diffuseness in ambisonics, in EAA Symposium on Auralization and Ambisonics (Berlin, 2014)
33. F. Zotter, M. Frank, Efficient phantom source widening. Arch. Acoust. 38(1) (2013)
34. F. Zotter, M. Frank, Phantom source widening by filtered sound objects, in prepr. 9728, Convention, Audio Engineering Society (Berlin, 2017)
35. J. Stautner, M. Puckette, Designing multi-channel reverberators. Comput. Music J. 6(1), 52–65 (1982)
36. J.-M. Jot, A. Chaigne, Digital delay networks for designing artificial reverberators, in 90th AES Convention, prepr. 3030 (Paris, 1991)
37. D. Rocchesso, Maximally diffusive yet efficient feedback delay networks for artificial reverberation. IEEE Signal Process. Lett. 4(9) (1997)
38. S. Tervo, J. Pätyinen, A. Kuusinen, T. Lokki, Spatial decomposition method for room impulse responses. J. Audio Eng. Soc. 61(1/2), 17–28 (2013), http://www.aes.org/e-lib/browse.cfm?elib=16664
39. J. Merimaa, V. Pulkki, Spatial impulse response rendering i: Analysis and synthesis. J. Audio Eng. Soc. 53(12), 1115–1127 (2005)
40. V. Pulkki, J. Merimaa, Spatial impulse response rendering ii: Reproduction of diffuse sounds and listening tests. J. Audio Eng. Soc. 54(1/2), 3–20 (2006)
41. M. Zaunschirm, M. Frank, F. Zotter, Brir synthesis using first-order microphone arrays, in prepr. 9944, 144th AES Convention (Milano, 2018)
42. C. Pörschmann, P. Stade, J.M. Arend, Binauralization of omnidirectional room impulse responses – algorithm and technical evaluation,” in DAFx-17 (Edinburgh, 2017)
43. M. Frank, F. Zotter, Spatial impression and directional resolution in the reproduction of reverberation, in Fortschrritte der Akustik - DEGA (Aachen, 2016)
44. M. Frank, F. Zotter, Exploring the perceptual sweet area in ambisonics, in AES 142nd Convention (2017)
45. V. Pulkki, Spatial sound reproduction with directional audio coding. J. Audio Eng. Soc. 55(6), 503–516 (2007)
46. J. Vilkamo, T. Lokki, V. Pulkki, Directional audio coding: virtual microphone-based synthesis and subjective evaluation. J. Audio Eng. Soc. 57(9) (2009)
47. S. Berge, Device and method converting spatial audio signal, US Patent, no. 8,705,750 B2 (2014)
48. A. Politis, S. Tervo, V. Pulkki, Compass: Coding and multidirectional parameterization of ambisonic sound scenes, in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (IEEE, 2018), pp. 6802–6806
49. C. Nachbar, F. Zotter, A. Sontacchi, E. Deleflie, Ambix - suggesting an ambisonic format, in Proceedings of the 3rd Ambisonics Symposium (Lexington, Kentucky, 2011)

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.