Chapter 1
XY, MS, and First-Order Ambisonics

Abstract This chapter describes first-order Ambisonic technologies starting from classical coincident audio recording and playback principles from the 1930s until the invention of first-order Ambisonics in the 1970s. Coincident recording is based on arrangements of directional microphones at the smallest-possible spacings in between. Hereby incident sound approximately arrives with equal delay at all microphones. Intensity-based coincident stereophonic recording such as XY and MS typically yields stable directional playback on a stereophonic loudspeaker pair. While the stereo width is adjustable by MS processing, the directional mapping of first-order Ambisonics is a bit more rigid: the omnidirectional and figure-of-eight recording pickup patterns are reproduced unaltered by equivalent patterns in playback. In perfect appreciation of the benefits of coincident first-order Ambisonic recording technologies in VR and field recording, the chapter gives practical examples for encoding, headphone- and loudspeaker-based decoding. It concludes with a desire for a higher-order Ambisonics format to get a larger sweet area and accommodate first-order resolution-enhancement algorithms, the embedding of alternative, channel-based recordings, etc.

Intensity-based coincident stereophonic recording such as XY uses two figure-of-eight microphones, after Blumlein’s original work [1] from the 1930s, with an angular spacing of 90°, see [2–4]). Another representative, MS, uses an omnidirectional and a lateral figure-of-eight microphone [2]. Both typically yield a stable directional playback in stereo, but signals often get too correlated, yielding a lack in depth and diffuseness of the recording space when played back [5, 6] and compared to delay-based AB stereophony or equivalence-based alternatives.

Gerzon’s work in the 1970s [7] gave us what we call first-order Ambisonic recording and playback technology today. Ambisonics preserves the directional mapping

Directionally sensitive microphones may be of the light moving strip type. [...] the strips may face directions at 45° on each side of the centre line to the sound source.

Alan Dower Blumlein [1], Patent, 1931

© The Author(s) 2019
F. Zotter and M. Frank, Ambisonics, Springer Topics in Signal Processing 19, https://doi.org/10.1007/978-3-030-17207-7_1
by recording and reproducing with spatially undistorted omnidirectional and figure-of-eight patterns on circularly (2D) or spherically (3D) surrounding loudspeaker layouts.

1.1 Blumlein Pair: XY Recording and Playback

The XY technique dates back to Blumlein’s patent from the 1930s [1] and his patents thereafter [4]. Nowadays outdated, manufacturers started producing ribbon microphones that offered means to record with figure-of-eight pickup patterns.

**Blumlein Pair using 90°-angled figure-of-eight microphones (XY).** Blumlein’s classic coincident microphone pair [3, Fig. 3] uses two figure-of-eight microphones pointing to ±45°, see Fig. 1.1. Its directional pickup pattern is described by \( \cos \phi \) when \( \phi \) is the angle enclosed by microphone aiming and sound source. Using a mathematically positive coordinate definition for X (front-right) and Y (front-left), the polar angle \( \varphi = 0 \) aiming at the front, the figure-of-eight X uses the angle \( \phi = \varphi + 45° \) and Y the angle \( \phi = \varphi - 45° \), so that the pickup pattern of the microphone pair is:

\[
g_{XY}(\varphi) = \begin{bmatrix} \cos(\varphi + 45°) \\ \cos(\varphi - 45°) \end{bmatrix}. \tag{1.1}\]

Assuming a signal \( s \) coming from the angle \( \varphi \), the signals recorded are \( [X, Y]^\top g(\varphi) \, s \). Sound sources from the left 45°, the front 0° and the right −45° will be received by the pair of gains:

![Fig. 1.1 Blumlein pair consisting of 90°-angled figure-of-eight microphones](image)
1.1 Blumlein Pair: XY Recording and Playback

right: \( g_{XY}(-45^\circ) = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \), center: \( g_{XY}(0^\circ) = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \), left: \( g_{XY}(45^\circ) = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \).

Obviously, a source moving from the right \(-45^\circ\) to the left \(45^\circ\) pans the signal from the channel X to the channel Y. This property provides a strongly perceivable later-alization of lateral sources when feeding the left and right channel of a stereophonic loudspeaker pair by Y and X, respectively.

However, ideally there should not be any dominant sounds arriving from the sides, as for the source angles between \(-135^\circ \leq \varphi \leq -45^\circ\) and \(45^\circ \leq \varphi \leq 135^\circ\) the Blumlein pair produces out-of-phase signals between X and Y. The back directions are mapped with consistent sign again, however, left-right reversed. It is only possible to avoid this by decreasing the angle between the microphone pair, which, however, would make the stereo image narrower.

Therefore, coincident XY recording pairs nowadays most often use cardioid directivities \(\frac{1}{2} + \frac{1}{2}\cos\varphi\), instead. They receive all directions without sign change and easily permit stereo width adjustments by varying the angle between the microphones.

1.2 MS Recording and Playback

Blumlein’s patent [1] considers sum and difference signals between a pair of channels/microphones, yielding M-S stereophony. In M-S [8], the sum signal represents the mid (omnidirectional, sometimes cardioid-directional to front) and the difference the side signal (figure-of-eight). MS recordings can also be taken with cardioid microphones and permit manipulation of the stereo width of the recording.

**MS recording by omnidirectional and figure-of-eight microphone (native MS).** Mid-side recording can be done by using a pair of coincident microphones with an omnidirectional (mid, W) and a side-ways oriented figure-of-eight (side, Y) directivity, Fig. 1.2. The pair of pickup patterns is described by the vector:

(a) Native MS recording (b) Picture of the recording setup

**Fig. 1.2** Native mid-side recording with the coincident arrangement of an omnidirectional microphone heading front and a figure-of-eight microphone heading left
that depends on the angle $\varphi$ of the sound source. Equation (1.2) maps a single sound $s$ from $\varphi$ to the mid $W$ and side $Y$ signals by the gains $[W, Y] = g(\varphi) s$

left: $g_{WY}(90^\circ) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, right: $g_{WY}(-90^\circ) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ center: $g_{WY}(0^\circ) = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

**MS recording with a pair of $180^\circ$-angled cardioids.** Two coincident cardioid microphones (cardioid directivity $\frac{1}{2} + \frac{1}{2} \cos \varphi$) pointing to the polar angles $90^\circ$ (left) and $-90^\circ$ (right) are also applicable to mid-side recording, Fig. 1.3. Their pickup patterns

$$g_{C^{\pm 90^\circ}}(\varphi) = \frac{1}{2} \begin{bmatrix} 1 + \cos(\varphi - 90^\circ) \\ 1 + \cos(\varphi + 90^\circ) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 + \sin(\varphi) \\ 1 - \sin(\varphi) \end{bmatrix}$$

are encoded into the MS pickup patterns $(W,Y)$ by a matrix

$$g_{WY}(\varphi) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} g_{C^{\pm 90^\circ}}(\varphi).$$

The matrix eliminates the cardioids’ figure-of-eight characteristics by their sum signal, and their omnidirectional characteristics by the difference. We obtain the MS signal pair $(W,Y)$ from the cardioid microphone signals as

$$\begin{bmatrix} W \\ Y \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} C_{90^\circ} \\ C_{-90^\circ} \end{bmatrix}.$$
Decoding of MS signals to a stereo loudspeaker pair. Decoding of the mid-side signal pair to left and right loudspeaker is done by feeding both signals to both loudspeakers, however out-of-phase for the side signal, Fig. 1.4b:

\[
\begin{bmatrix} L \\ R \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} W \\ Y \end{bmatrix}.
\]

(1.6)

An interesting aspect about the 180°-angled cardioid microphone MS is that after inserting the XY-to-MS encoder Eq. (1.5) into the decoder Eq. (1.6), a brief calculation shows that matrices invert each other. In this case, the cardioid signals are directly fed to the loudspeakers \([L, R] = [C_{90°}, C_{-90°}]\).

Stereo width. Modifying the mid versus side signal balance before stereo playback, using a blending parameter \(\alpha\), allows to change the width of the stereo image from \(\alpha = 0\) (narrow) to \(\alpha = 1\) (full), Fig. 1.4a, see also [9]:

\[
\begin{bmatrix} L \\ R \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 2 - \alpha & 0 \\ 0 & \alpha \end{bmatrix} \begin{bmatrix} W \\ Y \end{bmatrix}.
\]

(1.7)

In stereophonic MS playback, the playback loudspeaker directions at ±30° are not identical to the peaks of the recording pickup pattern of the side channel (Y) at ±90°. Ambisonics assumes a more strict correspondence between directional patterns of recording and patterns mapped on the playback system.

1.3 First-Order Ambisonics (FOA)

After Cooper and Shiga [10] worked on expressing panning strategies for arbitrary surround loudspeaker setups in terms of a directional Fourier series, the notion and technology of Ambisonics was developed by Felgett [11], Gerzon [7], and Craven [12]. In particular, they were also considering a suitable recording technology.

Essentially based on similar considerations as MS, one can define first-order Ambisonic recording. For 2D recordings, a Double-MS microphone arrangement is suitable and only requires one more microphone than MS recording: a front-back
oriented figure-of-eight microphone. The scheme is extended to 3D first-order Ambisonics by a third figure-of-eight microphone of up-down aiming. Oftentimes, first-order Ambisonics still is the basis of nowadays’ virtual reality applications and 360° audio streams on the internet. In addition to potential loudspeaker playback, it permits interactive playback on head-tracked headphones to render the acoustic sound scene static to the listener.

First-order Ambisonic recording has the advantage that it can be done with only a few high-quality microphones. However, the sole distribution of first-order Ambisonic recordings to playback loudspeakers is typically not convincing without going to higher orders and directional enhancements (Sect. 5.8).

### 1.3.1 2D First-Order Ambisonic Recording and Playback

The first-order Ambisonic format in 2D consists of one signal corresponding to an omnidirectional pickup pattern (called W), and two signals corresponding to the figure-of-eight pickup patterns aligned with the Cartesian axes (X and Y).

**Native 2D Ambisonic recording (Double-MS).** To record the Ambisonic channels W, X, Y, one can use a Double-MS arrangement as shown in Fig. 1.5.

**2D Ambisonic recording with four 90°-angled cardioids.** Extending the MS scheme for recording with cardioid microphones, Fig. 1.3, cardioid microphones could be used to obtain the front-back and left-right figure-of-eight pickup patterns by corresponding pair-wise differences, and one omnidirectional pattern as their sum, Fig. 1.6. However, the use of 4 microphones for only 3 output signals is inefficient.

![Fig. 1.5](image.png) Native 2D first-order Ambisonic recording with an omnidirectional and a figure-of-eight microphone heading front, and a figure-of-eight microphone heading left; photo shown on the right
1.3 First-Order Ambisonics (FOA)

Fig. 1.6 2D first-order Ambisonic recording with four 90°-angled cardioid microphones, sums and differences between them (front±back, left±right)

2D Ambisonic recording with three 120°-angled cardioids. Assuming 3 coincident cardioid microphones aiming at the angles 0°, ±120° in the horizontal plane, cf. Fig. 1.7, we obtain as the pickup pattern for the incoming sound

\[
g(\theta) = \frac{1}{2} + \frac{1}{2} \begin{bmatrix} \cos(\phi) \\ \cos(\phi + 120^\circ) \\ \cos(\phi - 120^\circ) \end{bmatrix}
\]

Combining all the three microphone signals yields an omnidirectional pickup pattern as \(\sum_{k=0}^{N-1} \cos(\phi + \frac{2\pi}{N} k) = 0\). Moreover introducing the differences between the front and two back microphone signals and between the left and right microphone signals yields an encoding matrix to obtain the omnidirectional W and the two X and Y figure-of-eight characteristics

\[
\frac{2}{3} \begin{bmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ 0 & \sqrt{3} & -\sqrt{3} \end{bmatrix} \begin{bmatrix} \cos(\phi) \\ \sin(\phi) \end{bmatrix}.
\]
**2D Ambisonic decoding to loudspeakers.** The W, X, and Y channel of 2D first-order Ambisonics (Double-MS) can easily be played on an arrangement of four loudspeakers, front, back, left, right. While the omnidirectional signal contribution is played by all of the loudspeakers, the figure-of-eight contributions are played out-of-phase by the corresponding front-back or left-right pair of loudspeakers, Fig. 1.8.

\[
\begin{bmatrix}
F \\
L \\
B \\
R
\end{bmatrix} = \begin{bmatrix}
1 & 1 & 0 \\
1 & 0 & 1 \\
1 & -1 & 0 \\
1 & 0 & -1
\end{bmatrix}
\begin{bmatrix}
W \\
X \\
Y
\end{bmatrix}.
\]

(1.9)

The decoding weights obviously discretizes the directional pickup characteristics of the Ambisonic channels at the directions of the loudspeaker layout. Consequently, if the loudspeaker layout is more arbitrary and described by the set of its angles \( \{ \varphi_l \} \), the *sampling decoder* can be given as

\[
\begin{bmatrix}
S_{\varphi_1} \\
\vdots \\
S_{\varphi_L}
\end{bmatrix} = \frac{1}{2}
\begin{bmatrix}
1 & \cos(\varphi_1) & \sin(\varphi_1) \\
\vdots & \vdots & \vdots \\
1 & \cos(\varphi_L) & \sin(\varphi_L)
\end{bmatrix}
\begin{bmatrix}
W \\
X \\
Y
\end{bmatrix}.
\]

(1.10)

To achieve a panning-invariant and balanced mapping by this decoder, loudspeakers should be evenly arranged. Moreover, it can be favorable to sharpen the spatial image by attenuating \( W \) by \( \frac{1}{\sqrt{3}} \) to map a sound by a supercardioid playback pattern.

**Playback to head-tracked headphones and interactive rotation.** In headphone playback, the headphone signals are generated by convolution with the head-related impulses responses of all four loudspeakers contributing to the left and the right ear signals

\[
\begin{bmatrix}
L_{\text{ear}} \\
R_{\text{ear}}
\end{bmatrix} = \begin{bmatrix}
h^0_{L}(t) & h^{90\circ}_{L}(t) & h^{180\circ}_{L}(t) & h^{-90\circ}_{L}(t) \\
h^0_{R}(t) & h^{90\circ}_{R}(t) & h^{180\circ}_{R}(t) & h^{-90\circ}_{R}(t)
\end{bmatrix}
\begin{bmatrix}
F \\
L \\
B \\
R
\end{bmatrix}.
\]

(1.11)
To rotate the Ambisonic input scene of the decoder, it is sufficient to obtain a new set of figure-of-eight signals by mixing the X, Y channels with the following matrix depending on the rotation angle $\rho$, keeping W unaltered

\[
\begin{bmatrix}
W \\
\tilde{X} \\
\tilde{Y}
\end{bmatrix} =
\begin{bmatrix}
1 & 0 & 0 \\
0 & \cos \rho & -\sin \rho \\
0 & \sin \rho & \cos \rho
\end{bmatrix}
\begin{bmatrix}
W \\
X \\
Y
\end{bmatrix}.
\] (1.12)

This effect is important for head-tracked headphone playback to render the VR/360° audio scene static around the listener. A complete playback system is shown in Fig. 1.9. The big advantage of such a system is that rotational updates can be done at high control rates and the HRIRs of the convolver are constant.

### 1.3.2 3D First-Order Ambisonic Recording and Playback

The first-order Ambisonic format in 3D consists of a signal W corresponding to an omnidirectional pickup pattern, and three signals (X, Y, and Z) corresponding to figure-of-eight pickup patterns aligned with the Cartesian coordinate axes.

In three dimensions, we cannot work with figure-of-eight patterns described by $\sin \phi$ or $\cos \phi$ of the azimuth angle only, anymore. It is more convenient to describe the arbitrarily oriented figure-of-eight characteristics $\cos(\phi)$ using the inner product between a variable direction vector (direction of arriving sound) and a fixed direction vector (microphone direction). Direction vectors are of unit length $\|\theta\| = 1$ and their inner product corresponds to $\theta^T \phi = \cos(\phi)$, where $\phi$ is the angle enclosed by the direction of arrival $\theta$ and the microphone direction $\theta_1$. Consequently, a cardioid pickup pattern aiming at $\theta_1$ is described by

\[
\frac{1}{2} + \frac{1}{2} \theta^T \theta.
\]

**Native 3D Ambisonic recording (Triple-MS).** To record the Ambisonic channels W, X, Y, Z, one can use a Triple-MS scheme as shown in Fig. 1.10. With the transposed unit direction vectors representing the aiming of the figure-of-eight channels $\theta^T_1 = [1, 0, 0], \theta^T_Y = [0, 1, 0], \theta^T_Z = [0, 0, 1]$, to produce the direction dipoles $\theta^T_1 \theta$, ...
\( \theta^T_X, \theta^T_Y, \text{and} \theta^T_Z \), we can mathematically describe the pickup patterns of native 3D first-order Ambisonics as

\[
g_{WXYZ}(\theta) = \begin{bmatrix} \theta^T_X \\ \theta^T_Y \\ \theta^T_Z \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \theta \\ \theta \\ \theta \end{bmatrix} = \begin{bmatrix} 1 \end{bmatrix}. \quad (1.13)
\]

**3D Ambisonic recording with a tetrahedral arrangement of cardioids.** The principle that worked for three cardioid microphones on the horizon also works for a coincident tetrahedron microphone array of cardioids with the aiming directions FLU-FRD-BLD-BRU, see Fig. 1.11, and [12],

\[
g(\theta) = \frac{1}{2} + \frac{1}{2} \begin{bmatrix} \theta^T_{FLU} \\ \theta^T_{FRD} \\ \theta^T_{BLD} \\ \theta^T_{BRU} \end{bmatrix} \theta = \frac{1}{2} + \frac{1}{2} \sqrt{3} \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} \theta. \quad (1.14)
\]

Encoding is achieved there by the matrix that adds all microphone signals in the first line (W omnidirectional), subtracts back from front microphone signals in the second line (X figure-of-eight), subtracts right from left microphone signals in the third line (Y figure-of-eight), and subtracts down from up microphone signals in the last line (Z figure-of-eight), see also Fig. 1.11,

\[
g_{WXYZ}(\theta) = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & -1 & -1 \end{bmatrix} \sqrt{3} \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \end{bmatrix} g(\theta). \quad (1.15)
\]
1.3 First-Order Ambisonics (FOA)

(a) Tetrahedral array with four cardioids

(b) Encoder of microphone signals

**Fig. 1.11** Tetrahedral arrangement of cardioid microphones for 3D first-order Ambisonics and their encoding; microphone capsules point inwards to minimize their spacing.

(a) Tetrahedral array of 4 cardioids

(b) SPS200

(c) MK4012

(d) ST450

**Fig. 1.12** Practical tetrahedral recording setups with cardioid microphones; the Soundfield SPS200, Oktava MK4012, and Soundfield ST450 offer a fixed 4-capsule geometry. Equally important: Zoom’s H3-VR, Røde’s NT-SF1, Sennheiser’s AMBEO VR Mic.

As Fig. 1.12 shows, practical microphone layouts should be as closely spaced as possible. Nevertheless for high frequencies, the microphones cannot be considered coincident anymore, and besides a directional error, there will be a loss of presence in the diffuse field. Typically a shelving filter is used to slightly boost high frequencies. Roughly, a high-shelf filter with a 3 dB boost is sufficient to correct timbral defects at frequencies above which the microphone spacing exceeds half a wavelength, e.g., 5 kHz for a 3.4 cm spacing of the microphones. More advanced strategies are found, e.g., in [7, 13–15].

**3D Ambisonic decoding to loudspeakers.** As before in the 2D case, a *sampling decoder* can be defined that represents the continuous directivity patterns associated with the channels W, X, Y, Z to map the signals to the discrete directions of the
loudspeakers. Given the set of loudspeaker directions \( \{ \theta_l \} \) and the unit-vectors to \( X \), \( Y \), \( Z \), the loudspeaker signals of the sampling decoder become

\[
\begin{bmatrix}
S_1 \\
\vdots \\
S_L
\end{bmatrix} = \frac{1}{2}
\begin{bmatrix}
1 & \theta_1^T \theta_X & \theta_1^T \theta_Y & \theta_1^T \theta_Z \\
\vdots & \vdots & \vdots & \vdots \\
1 & \theta_L^T \theta_X & \theta_L^T \theta_Y & \theta_L^T \theta_Z
\end{bmatrix}
\begin{bmatrix}
W \\
X \\
Y \\
Z
\end{bmatrix} = \frac{1}{2}
\begin{bmatrix}
1 & \theta_1^T \\
\vdots & \vdots \\
1 & \theta_L^T
\end{bmatrix}
\begin{bmatrix}
W \\
X \\
Y \\
Z
\end{bmatrix}.
\] (1.16)

**Equivalent panning function/virtual microphone.** The sampling decoder together with the native Ambisonic directivity patterns \( g_{WXYZ}^T(\theta) = [1, \theta] \) yields the mapping of a signal \( s \) from the direction \( \theta \) to the loudspeakers to be

\[
\begin{bmatrix}
S_1 \\
\vdots \\
S_L
\end{bmatrix} = \frac{1}{2}
\begin{bmatrix}
1 & \theta_1^T \\
\vdots & \vdots \\
1 & \theta_L^T
\end{bmatrix}
\begin{bmatrix}
1 \\
\theta
\end{bmatrix}s = \frac{1}{2}
\begin{bmatrix}
1 + \theta_1^T \theta \\
\vdots \\
1 + \theta_L^T \theta
\end{bmatrix}s.
\] (1.17)

This result means that the gain of a source from \( \theta \) at each loudspeaker \( \theta_l \) corresponds to evaluating a cardioid pattern aligned with \( \theta \). Consequently, the Ambisonic mapping corresponds to a signal distribution to the loudspeakers using weights obtained by discretization of an Ambisonics-equivalent first-order panning function.

Equivalently, Ambisonic playback using a sampling decoder is comparable to recording each loudspeaker signal with a virtual first-order cardioid microphone aligned with the loudspeaker’s direction \( \theta_l \).

It is decisive for a panning-independent loudness mapping and balanced performance that the directions of the loudspeaker layout are well chosen. Also, it can be preferred to reduce the level of the omnidirectional channel \( W \) by \( \frac{1}{\sqrt{3}} \) to map a sound by the narrower supercardioid playback pattern instead of a cardioid pattern, which is rather broad.

Decoder design problems were early addressed by Gerzon [16], Malham [17], and Daniel [18]. A current solution for higher-order decoding is given in Sect. 4.9.6 on All-round Ambisonic decoding.

**3D Ambisonic decoding to headphones.** 3D Ambisonic decoding to headphones uses the same approach as for 2D above, except that additional rotational degrees are implemented to compensate for any change in head orientation. Rotation concerns the three directional components \( X, Y, Z \)

\[
\begin{bmatrix}
\tilde{X} \\
\tilde{Y} \\
\tilde{Z}
\end{bmatrix} = R(\alpha, \beta, \gamma)
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix}.
\] (1.18)

For the definition of the rotation matrix \( R(\alpha, \beta, \gamma) \) and the meaning of its angles refer to Eq. 5.5 of Sect. 5.2.2. The selection of a suitable set of HRIRs is a question
of directional discretization of the 3D directions, as addressed in the decoder above. Signals obtained for virtual loudspeakers are again to be convolved with the corresponding HRIRs for the left and the right ear.

1.4 Practical Free-Software Examples

The practical examples below show first-order Ambisonic panning a mono sound, decoded to simple loudspeaker layouts. These are either a square layout with 4 loudspeakers at the azimuth angles [0°, 90°, 180°, −90°] or an octahedral layout with 6 loudspeakers at azimuth [0°, 90°, 180°, −90°, 0°, 0°] and elevation [0°, 0°, 0°, 0°, 90°, −90°].

1.4.1 Pd with Iemmatrix, Iemlib, and Zexy

Pd is free and it can load and install its extensions from the internet. Required software components are:

- pure-data (free, http://puredata.info/downloads/pure-data)
- iemmatrix (free download within pure-data)
- zexy (free download within pure-data)
- iemlib (free download within pure-data)

Figure 1.13 gives an example for horizontal (2D) first-order Ambisonic panning, decoded to 4 loudspeaker and 2 headphone signals.

Figure 1.14 shows the processing inside the Pd abstraction [FOA_binaural_decoder] contained in the Fig. 1.13 example, which uses SADIE database\(^1\) subject 1 (KU100 dummy head) HRIRs to render headphone signals.

Figure 1.15 sketches a first-order Ambisonic panning in 3D with decoding to an octahedral loudspeaker layout; master level [multiline~] and hardware outlets [dac~] were omitted for easier readability.

1.4.2 Ambix VST Plugins

This example uses a DAW and ready-to-use VST plug-ins to render first-order Ambisonics. As DAW, we recommend Reaper (reaper.fm) because it nicely facilitates higher-order Ambisonics by allowing tracks of up to 64 channels. Moreover, it is relatively low-priced and there is a fully functional free evaluation version available. You can also use any other DAW that supports VST and sufficiently many multi-track

\(^1\)https://www.york.ac.uk/sadie-project/Resources/SADIEIIIDatabase/D1/D1_HRIR_WAV.zip.
Fig. 1.13  First-order 2D encoding and decoding in pure data (Pd) for a square layout

Fig. 1.14  Binaural rendering to headphones on pure data (Pd) by convolution with SADIE KU100 HRIRs

channels. The example employs the freely available ambiX plug-in suite (http://www.matthiaskronlachner.com/?p=2015), although there exist other Ambisonics plug-ins, especially for first-order.

| Track Name   | Ins | Outs | FX          |
|--------------|-----|------|-------------|
| Virtual source | 1   | 4    | ambix_encoder_o1 |
| MASTER        | 4   | 6    | ambix_decoder_o1 |
After creating the new track for the virtual source and importing a mono/stereo audio file (per drag-and-drop), the next step is the setup of the track channels. As shown in the table, the virtual source has a single-channel (mono) input and 4 output channels to send the 4 channels of first-order Ambisonics to the Master. The option to send to the Master is activated by default, cf. left in Fig. 1.16. The Master track itself requires 4 input channels and 6 output channels to feed the 6 loudspeakers (right). In Reaper, there is no separate adjustment for input and output channels, thus the Master track has to be set to 6 channels.

In the source track FX, the `ambix_encoder_o1` can be used to encode the virtual source signal at an arbitrary location on a sphere by inserting the plug-in into the track of the virtual source, cf. its panning GUI in Fig. 1.17. For adding more sources, the track of the virtual source can simply be copied or duplicated. All effects and routing options are maintained for the new tracks.

In order to decode the 4 first-order Ambisonics Master channels to the loudspeakers the `ambix_decoder_o1` plug-in is added to the Master track. The plug-in requires a preset that defines the decoding matrix and its channel sequence and normalization. For the exemplary octahedral setup with 6 loudspeakers, the following text can be copied to a text file and saved as config-file, e.g., “octahedral.config”. The decoder
matrix contains $W$, $-Y$, $Z$, $X$, with $W$ as constant and $-Y$, $Z$, $X$ refer to Cartesian coordinates of the octahedron.

```
#GLOBAL
/coeff_scale n3d
/coeff_seq acn
#END

#DECODERMATRIX
1 0 0 1
1 1 0 0
1 -1 0 0
1 0 0 -1
1 0 1 0
1 0 -1 0
#END
```
After loading the preset into the decoder plug-in, the decoder can generate the loudspeaker signals as shown in Fig. 1.18. In the example, the virtual source is panned to the front, resulting in the highest level for loudspeaker 1 (front). The loudspeaker 3 (back) is 12dB quieter because of a side-lobe suppressing super cardioid weighting implied by the switch /coeff_scale n3d, as a trick to keep things simple.

As shown on the SADIE-II website, the SADIE-II head-related impulse responses can be used to rendering Ambisonics to headphones. The listing below shows a configuration file to be used with ambix_binaural, cf. Fig. 1.19, again using the trick to select n3d to keep the numbers simple and super-cardioid weighting

```
#GLOBAL
/coeff_scale n3d
/coeff_seq acn
#END

#HRTF
44K_16bit/azi_0,0_ele_0,0.wav
44K_16bit/azi_90,0_ele_0,0.wav
44K_16bit/azi_180,0_ele_0,0.wav
44K_16bit/azi_270,0_ele_0,0.wav
44K_16bit/azi_0,0_ele_90,0.wav
44K_16bit/azi_0,0_ele_-90,0.wav
#END
```

2https://www.york.ac.uk/sadie-project/ambidec.html.
For decoding to less regular loudspeaker layouts, the IEM AllRADecoder\textsuperscript{3} permits editing loudspeaker coordinates and automatically calculating a decoder within the plugin. For decoding to headphones, the IEM BinauralDecoder offers a high-quality decoder. The technology behind both plugins is explained in Chap. 4.

In addition to the virtual sources, you can also add a 4-channel recording done with a B-format microphone by placing the 4-channel file in a new track. Reaper will automatically set the number of track channels to 4 and send the channels to the Master. Note that some B-format microphones use a different order and/or weighting of the Ambisonics channels. Simple conversion to the AmbiX-format can be done by inserting the ambix_converter_o1 plug-in into the microphone track.

### 1.5 Motivation of Higher-Order Ambisonics

*Diffuseness, spaciousness, depth?* Diffuse sound fields are typically characterized by sound arriving randomly from evenly distributed directions at evenly distributed delays. It is practical knowledge that the impression of diffuseness and spaciousness

\textsuperscript{3}https://plugins.iem.at/.
requires benefits from decorrelated signals, which is typically achieved by large
distances between the microphones rather than by coincident microphones.

Due to the evenness of diffuse sound fields, one would still hope that a low spatial
resolution is sufficient to map diffuseness and spatial depth of a room, using coin-
cident microphones or first-order Ambisonics. Nevertheless, high directional corre-
lation during playback destroys this hope and in fact yields a perceptually impeded
playback of diffuseness, spaciousness, and depth.

The technical advantages in interactivity and VR as well as the known short-
comings of first-order coincident recording techniques offer enough motivation to
increase the directional resolution and go to higher-order Ambisonics, as presented
in the subsequent chapters. For professional productions, it is often not sufficient to
only rely on first-order coincident microphone recordings. By contrast, higher-order
Ambisonics is able to drastically improve the mapping of diffuseness, spaciousness,
and depth, as shown in the upcoming chapter about psychoacoustical properties of
many-loudspeaker systems.

Recording with a higher-order main microphone array increases the required
technological complexity. Nevertheless, digital signal processing and the theory pre-
sented in the later chapters is powerful nowadays to achieve this goal.

After all, it seems that delay-based stereophonic recording, such as AB, or
equivalence-based recording, such as ORTF, INA5, etc., is often required and well-
known in its mapping properties for spaciousness and diffuseness, correspondingly.
What is nice about higher-order Ambisonics: it can make use of these benefits by
embedding such recordings appropriately, see Fig. 1.20.

**Facts about higher orders**: Ambisonics extended to higher orders permits a refine-
ment of the directional resolution and hereby improves the mapping of uncorrelated
sounds in playback. Figure 1.21a shows the correlation introduced in two neighbor-
ing loudspeaker signals when using Ambisonics, given their spacing of $60^\circ$. Given
the just noticeable difference (JND) of the inter-aural cross correlation, the figure
indicates that an Ambisonic order of $\geq 3$ might be necessary to perceptually preserve
decorrelation.

---

**Fig. 1.20** How is a microphone tree represented in Ambisonics, when it consist of 6 cardioids
spaced by 60 cm and $60^\circ$ on a horizontal ring, and a ring of 4 super cardioids spaced by 40 cm and
$90^\circ$ as height layer, pointing upwards?
For this reason, the perception of spatial depth strongly improves when increasing the Ambisonic order from 1 up to 3, Fig. 1.21b. However, this is only the case when seated at the central listening position. Outside this sweet spot, higher orders than 3, e.g., 5, additionally improve the mapping of depth [19]. Therefore, higher-order Ambisonics is important for preserving spatial impressions and when supplying a large audience.

Figure 1.22 shows that the sweet area of perceptually plausible playback increases with the Ambisonic order [20]. With fifth-order Ambisonics, nearly all the area spanned by the horizontal loudspeakers at the IEM CUBE, the $12 \times 10$ m concert space at our lab, becomes a valid listening area.
References

1. A.D. Blumlein, Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems. GB patent 394325 (1933), http://www.richardbrice.net/blumlein_patent.htm
2. M.A. Gerzon, Why coincident microphones? in Reproduced from Studio Sound, by permission of IPC Media Ltd, publishers of Hi-Fi News (www.hifinews.co.uk), vol. 13 (1971), https://www.michaelgerzonphotos.org.uk/articles/Why
3. P.B. Vanderlyn, In search of Blumlein: the inventor incognito. J. Audio Eng. Soc. 26(9) (1978)
4. S.P. Lipshitz, Stereo microphone techniques: Are the purists wrong? J. Audio Eng. Soc. 34(9) (1986)
5. S. Weinzierl, Handbuch der Audiotechnik (Springer, Berlin, 2008)
6. A. Friesen, Die Audio-Enzyklopädie. G. Sazur (2007)
7. M.A. Gerzon, The design of precisely coincident microphone arrays for stereo and surround sound, prepr. L-20 of 50th Audio Eng. Soc. Conv. (1975)
8. G. Bore, S.F. Temmer, M-s stereophony and compatibility, Audio Magazine (1958), http://www.americanradiohistory.com
9. M.A. Gerzon, Application of blumlein shuffling to stereo microphone techniques. J. Audio Eng. Soc. 42(6) (1994)
10. D.H. Cooper, T. Shiga, Discrete-matrix multichannel stereo. J. Audio Eng. Soc. 20(5), 346–360 (1972)
11. P. Felgett, Ambisonic reproduction of directionality in surround-sound systems. Nature 252, 534–538 (1974)
12. P. Craven, M.A. Gerzon, Coincident microphone simulation covering three dimensional space and yielding various directional outputs, U.S. Patent, no. 4,042,779 (1977)
13. C. Faller and M. Kolundžija, “Design and limitations of non-coincidence correction filters for soundfield microphones,” in prepr. 7766, 126th AES Conv, Munich (2009)
14. J.-M. Batke, The b-format microphone revisited, in 1st Ambisonics Symposium, Graz (2009)
15. A. Heller E. Benjamin, Calibration of soundfield microphones using the diffuse-field response, in prepr. 8711, 133rd AES Conv, San Francisco (2012)
16. M. Gerzon, General metatheory of auditory localization, in prepr. 3306, Conv. Audio Eng. Soc. (1992)
17. D.G. Malham, A. Myatt, 3D Sound spatialization using ambisonic techniques. Comput. Music J. 19(4), 58–70 (1995)
18. J. Daniel, J.-B. Rault, J.-D. Polack, Acoustic properties and perceptive implications of stereophonic phenomena, in AES 6th International Conference: Spatial Sound Reproduction (1999)
19. M. Frank, F. Zotter, Spatial impression and directional resolution in the reproduction of reverberation, in Fortschritte der Akustik - DEGA, Aachen (2016)
20. M. Frank, F. Zotter, Exploring the perceptual sweet area in ambisonics, in AES 142nd Conv. (2017)
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.