Audio Engineering Society

Convention Paper 7297

Presented at the 123rd Convention
2007 October 5–8 New York, NY, USA

Ambisonic Panning

Martin Neukom

Institute for Computer Music and Sound Technology ICST, Zurich University of the Arts, Baslerstrasse 30,
8048 Zurich, Switzerland,
martin.neukom@zhdk.ch

ABSTRACT

Ambisonics is a surround-system for encoding and rendering a 3D sound field. Sound is encoded and stored in multi-channel sound files and is decoded for playback. In this paper a panning function equivalent to the result of ambisonic encoding and so-called in-phase decoding is presented. In this function the order of ambisonic resolution is just a variable that can be an arbitrary positive number not restricted to integers and that can be changed during playback. The equivalence is shown, limitations and advantages of the technique are mentioned and real time applications are described.

1. AMBISONICS

Ambisonics is a surround-system for encoding and rendering a 3D sound field. In Ambisonics the room information of the recorded or synthesized sound is encoded together with the sound itself in a certain number of channels independent of the speaker set-up. The encoding can be carried out in an arbitrary degree of accuracy. The accuracy is given by the so-called order of Ambisonics. The zeroth order corresponds to the mono signal and needs one channel. In first order Ambisonics the portions of the sound field in the directions x, y and z are encoded in three more channels. The interpretation of higher orders is not as simple as that of zeroth and first order. If one calculates the sum of sound waves of several speakers at an arbitrary point of an auditorium, complicated formulas arise. In Ambisonics the situation is simplified by the assumption that the sound waves are plane and that the listener is located at the origin of the coordinate system. Sound waves in the horizontal plane, represented in polar coordinates, and sound waves in the room, represented in spherical coordinates, can be calculated as sums of cylindrical or spherical harmonics respectively. This decomposition can be interpreted as the multiplication of the wave function and the directivity of the sound source [1][2].

1.1. Encoding

The formulas for ambisonic encoding are derived from the solution of the three-dimensional wave equation

$$(\Delta + \boldsymbol{k}^2)p = 0 \tag{1.}$$

in the spherical coordinate system (see figure 1) where a point P is described by radius r, azimuth θ and elevation δ.

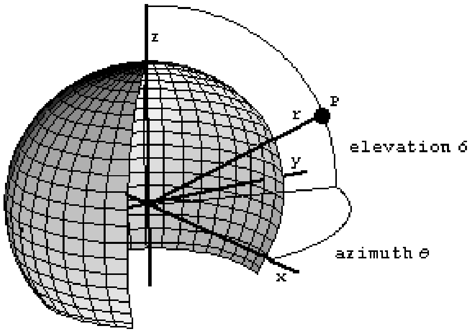


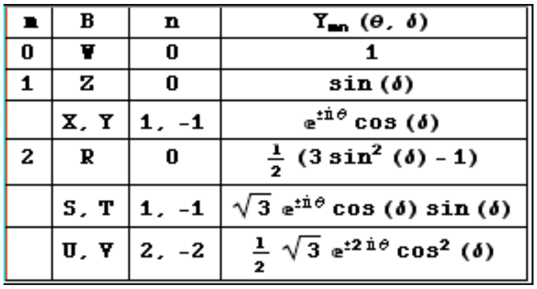
Figure 1: Spherical coordinate system

In these coordinates the pressure field can be written as the Fourier-Bessel series (2). Its terms are the weighted products of directional functions Y(θ, δ) called spherical harmonics and radial functions [3].

$$p(\boldsymbol{r}) = \sum_{m=0}^{\infty} j^m j_m(kr) \sum_{0 \le n \le m, \sigma = \pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta, \delta) \qquad (2.)$$

(with vector ***r*** of length *r*, wave vector ***k***, wave number $k = 2\pi f/c$, frequency *f* and speed of sound *c*).

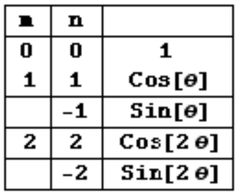
Assuming that the sound waves are plane and that the listener is located at the origin of the coordinate system the formulas can be simplified dramatically. In practice the infinite series is truncated and only a finite number of components are calculated and saved in the so-called ambisonic B-format. After all these simplifications a signal S is encoded by multiplying the signal with the first spherical harmonics in 3D and with the first harmonics in 2D. The first few components in 3D are labelled and defined as follows (semi-normalized version [1]):



m	B	n	$Y_{mn}(\theta, \delta)$
0	W	0	1
1	Z	0	$\sin(\delta)$
	X, Y	1, -1	$e^{\pm i\theta}\cos(\delta)$
2	R	0	$\frac{1}{2}(3\sin^2(\delta)-1)$
	S, T	1, -1	$\sqrt{3}\, e^{\pm i\theta}\cos(\delta)\sin(\delta)$
	U, V	2, -2	$\frac{1}{2}\sqrt{3}\, e^{\pm 2 i\theta}\cos^2(\delta)$

Table 1: Ambisonic B-format

In 2D the components are simply



m	n	
0	0	1
1	1	Cos[θ]
	-1	Sin[θ]
2	2	Cos[2 θ]
	-2	Sin[2 θ]

Table 2: B-format 2D

The order of resolution m defines the accuracy of the encoding and the number of channels in the B-format, namely 2m+1 in 2D and $(m+1)^2$ in 3D.

1.2. Decoding

In order to render a 3D sound field the speaker signals for a given speaker setup must be calculated. The signals of the single speakers can be treated in the same way as the encoded sound. If we denote the B-format of the encoded sound by **B**, the B-format of the sound from speaker i by $\mathbf{c}_i$ and the signal from speaker i by S_i

$$c_i = \begin{bmatrix} Y_{00}^{1}(\theta_i,\delta_i) \\ \ldots \\ Y_{mn}^{\sigma}(\theta_i,\delta_i) \end{bmatrix} \quad B = \begin{bmatrix} B_{00}^{1} \\ \ldots \\ B_{mn}^{\sigma} \end{bmatrix} \quad S = \begin{bmatrix} S_1 \\ \ldots \\ S_n \end{bmatrix} \qquad (3.)$$

and the matrix of all $\mathbf{c}_i$ as

$$C = [c_1, \ldots, c_n] \qquad (4.)$$

the reproduction of the encoded sound with the n speakers can be written as

$$B = C.S \qquad (5.)$$

This equation can theoretically be solved in respect to **S** for nearly any speaker setup and any number of speakers greater than the number of channels of **B**, but it turns out that solutions for asymmetrical setups often are unusable (for 5.1 surround see [4]) and that the solution of symmetric setups does not change if more speakers than necessary are used. The speaker signals for symmetrical setups of n speakers (n = 2m+1 in 2D and $(m+1)^2$ in 3D) can be calculated from the B-format and the matrix **C** as [3]

$$S = C^{-1}.B = \frac{1}{n} C^T.B \tag{6.}$$

In 2D the signal for speaker i is given by

$$\begin{aligned} S_i = \frac{1}{n}(W &+ 2X\cos\theta_i + 2Y\sin\theta_i \\ &+ 2U\cos 2\theta_i + 2V\sin 2\theta_i + ...) \end{aligned} \tag{7.}$$

1.3. Corrections

Ambisonics is based on harmonic decomposition. The truncation of the infinite series causes side effects such as signals on speakers far away from the original sound position and inverted phases (see figure 2). By windowing the decomposition i.e. weighting the ambisonic channels according to their order these side effects can be reduced at the cost of the precision of the directivity. Figure 2 shows two level functions for a speaker at position θ (sound at θ = 0, order m = 3) the first without correction (basic decoding) $f_{bas}(\theta)$, the second for so-called in-phase decoding $f_{inph}(\theta)$. The bars indicate the levels of 13 symmetrically positioned speakers.

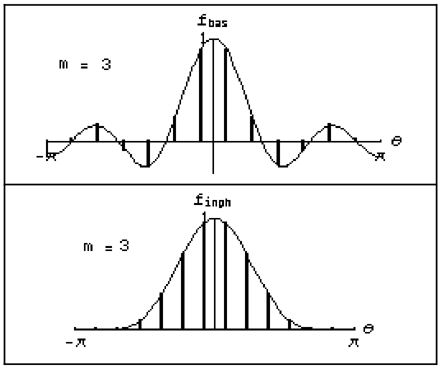


Figure 2: Level functions for basic and in-phase decoding

Putting the correcting gains into equation (6.) yields

$$S = \frac{1}{n} C^T Diag[...g_m...].B \tag{8.}$$

2. PANNING

Panning is the technique of the positioning of a single (monophonic) source within a stereophonic image. The technique has been enhanced from two to more than two speakers and from two to three dimensions. Vector Base Panning (VBP) was introduced by Ville Pulkki [5] for two dimensions. In VBP loudspeaker arrays are treated as arrangements of subsequent stereo pairs or, extended to three dimensions, as triples of loudspeakers. Panning normally uses only level differences and feeds only the loudspeakers nearest to the virtual sound source.

Amplitude panning is the most frequently used technique to position virtual sources. The sound signal x(t) is fed to speaker i with the gain factor f_i

$$x_i(t) = f_i x(t), \qquad i = 1, \dots, N \tag{9.}$$

which is a function of the difference $\Delta\varphi = \varphi - \varphi_i$ of the angles between the sound source at position φ and speaker i at position φ_i.

A typical panning function is the cosine function (see figure 3).

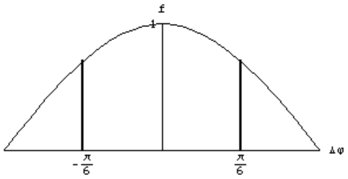


Figure 3: stereo panning with the cosine function

Panning functions fulfil the condition [2]:

$$\sqrt[p]{g_l{}^p + g_r{}^p} = 1 \tag{10.}$$

In contrast to other panning techniques ambisonic panning functions normally produce signals for all speakers at the same time. The functions are defined on the whole horizontal circle or the whole sphere. The sum of all speaker gains equals 1 (i.e. p = 1 in equation (10.)). For valid gains g equation (8.) yields the ambisonic panning functions

$$f(\theta,m)=\frac{1}{n}(g_0+2\sum_{k=1}^{m}g_k\cos k\theta) \qquad (11.)$$

for 2D and

$$f(\theta,m)=\frac{1}{n}\sum_{k=1}^{m}(2m+1)g_kP_k(\cos\theta) \qquad (12.)$$

for 3D [3].

3. SIMPLIFIED AMBISONIC PANNING FUNCTIONS

For basic and in-phase decoding the formulas (11.) and (12.) can be simplified. Basic decoding

For basic decoding (i.e. without correcting gains) the panning function (11.) is

$$f(\theta,m)=\frac{1}{n}(1+2\sum_{k=1}^{m}\cos k\theta) \qquad (13.)$$

Since the sum can be written in a simplified form we get

$$f(\theta,m)=\frac{\sin(\frac{2m+1}{2}\theta)}{n\sin(\frac{1}{2}\theta)} \qquad (14.)$$

This function is exactly equivalent to ambisonic en- and decoding in 2D. For higher orders of m the number of calculations is reduced dramatically and its implementation is straightforward. Because the function depends only on the angle between speaker and sound source it can also be used in 3D as a panning function. For tests it can be used as an approximation for basic decoding in 3D (in 3D θ denotes the angle between sound source and speaker). In figure 4 the panning functions 3D and 2D (dashed line) are compared.

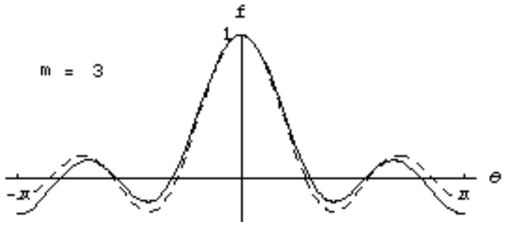


Figure 4: Basic decoding panning function 3D and 2D

3.1. In-phase decoding

With the so-called in-phase decoding negative amplitudes and side lobes in the panning functions are avoided with the following correcting gains [1].

$$g_k = g_0\frac{m!^2}{(m+k)!(m-k)!} \qquad \text{for 2D} \qquad (15.)$$

$$g_k = g_0\frac{m!(m-1)!}{(m+k+1)!(m-k)!} \qquad \text{for 3D} \qquad (16.)$$

with normalizing factors g_0 [1].

The panning functions (11.) and (12.) with the gains (15.) and (16.) are equivalent to the simple function

$$f_{inph}(\theta,p)=(\tfrac{1}{2}+\tfrac{1}{2}\cos\theta)^p=(\cos\frac{\theta}{2})^{2p} \qquad (17.)$$

where θ is the angle between the speaker and the position of the sound source and p corresponds to the ambisonic order. The following diagrams show the function $f_{inph}(\theta,p)$ for p = 2 and 6 and seven symmetrically positioned loudspeakers.

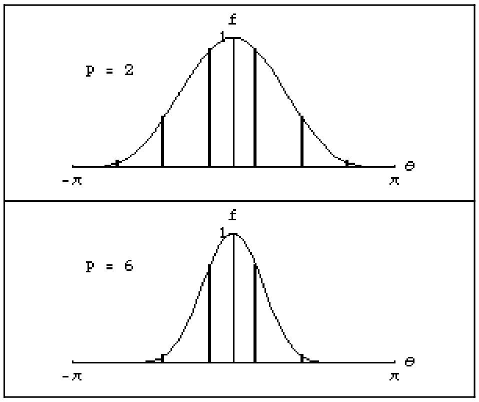


Figure 5: in-phase panning function

In order to show the equivalence of this function and the function we get by ambisonic encoding, decoding and multiplying with the standard gains for in-phase decoding, the function $f_{inph}(\theta,p)$ is first expanded

$$f_{inph}(\theta,p)=\frac{1}{2^p}\sum_{i=0}^{p}\binom{p}{i}\cos^p\theta \qquad (18.)$$

Reducing the powers of the cosine functions we get

$$f_{inph}(\theta,p)=\frac{1}{2^{2p-1}}\left(\frac{1}{2}\binom{2n}{n}+\sum_{i=1}^{p}\binom{2p}{p+i}\cos i\theta\right) \tag{19.}$$

The coefficients of the cosines of multiples of the angle θ correspond to the gains for in-phase decoding (15.) and (16.).

Since the panning function is equivalent to ambisonic en- and decoding, for integer exponents the results of ambisonic theory hold. In ambisonic theory there must be at least as many speakers as ambisonic channels (n = 2m+1 in 2D and n = $(m+1)^2$ in 3D for order m) to calculate the decoding formulas. Nevertheless n = m-1 speakers suffice to produce speaker signals that sum up to 1. For small orders (ca. up to 5th order) this is shown by simplifying the sum of the panning functions f_{inph} for n symmetrically positioned speakers. For large orders numerical calculations give the same result. In figure 6 the sum of 9 speaker signals as a function of the angle of the sound source are compared for orders p = 8 and p = 9.

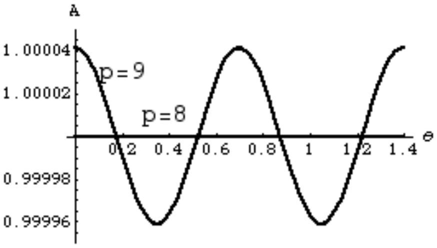


Figure 6: Sum of 9 speaker signals as a function of the position of a sound source for orders p = 8 and p = 9

3.2.1 Fractional orders

While ambisonic encoding is only possible with integer orders the exponent in the panning function $f_{inph}(\theta,p)$ (17.) can be an arbitrary positive number. For fractional order the sum of the speaker signals does not exactly equal one but the deviation is very small, so that it is possible to change the exponent continuously without perceivable inaccuracies. Figure 7 shows the function $f_{inph}(\theta,p)$ for n = 8 speakers. It is nearly constant between p = 2 and p = n-1.

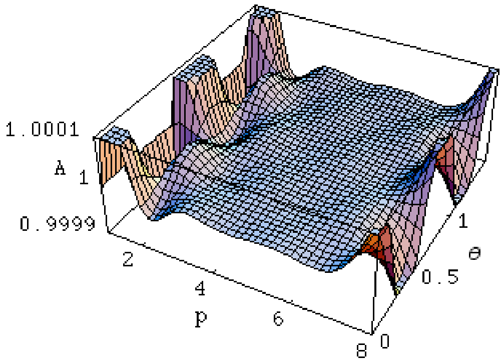


Figure 7: $f_{inph}(\theta,p)$ for 8 speakers

Since with increasing order p the function $f_{inph}(\theta,p)$ narrows more and more slowly, fewer speakers per order are necessary. Figure 8 shows that with as few as n = 20 speakers it is possible to use orders up to p = 60.

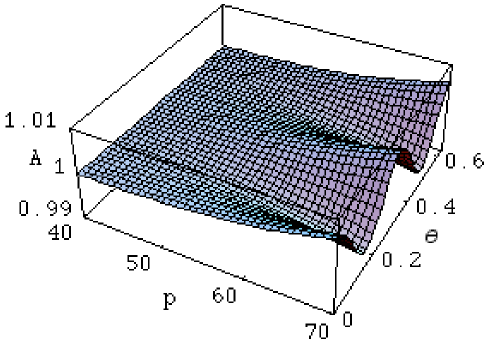


Figure 8: $f_{inph}(\theta,p)$ for 20 speakers

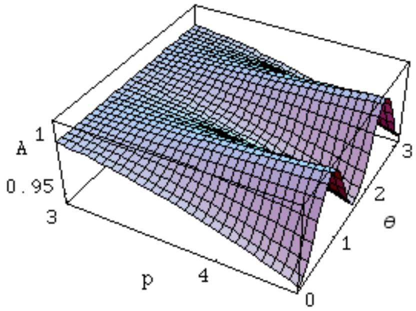
3.2.2 3D-panning

The same panning function can be used in 3D. The only five symmetrical speaker setups correspond to the five platonic solids. For these setups it can be shown that the sum of the speaker signals is independent of the position of the sound source for small integer orders (see table 4). The sum can be normalized by the factor (p+1)/n.

$$\frac{p+1}{n}\sum_{i=1}^{n}f_{inph}(\theta_i,p)=1 \tag{20.}$$

where θ_i is the angle between the sound source and speaker i, p the order and n the number of speakers.

Figure 9 shows the sum of the speaker signals as a function of the order p and the horizontal angle θ (δ = 0) for a setup with 8 speakers in a cube.



Figur 9: Sum of the speaker signals as a function of order p and angle θ

For fractional orders p < 3 small deviations occur (see figure 10).

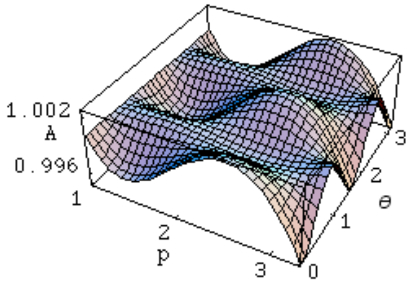


Figure 10: Sum of the speaker signals as a function of order p and angle θ

The 3D B-format contains n = $(m+1)^2$ channels for order m. Thus at least n equations (and therefore n speakers) are required to calculate the decoding formulas for orders up to m = √n -1. Table 3 shows that the order chosen can be nearly twice that number.

speakers n	$\sqrt{n} - 1$	exact up to p (integer)	error < .01 for p <
4	1	2	2
6	1.45	3	3
8	1.83	3	3.5
12	2.46	5	6.7
20	3.47	5	7.5

Table 3: relation between ambisonic order and number of speakers in 3D

4. APPLICATIONS AND PERSPECTIVES

4.1. Implementations

The implementation of the panning functions f_{inph} is straightforward. In order to produce the signal for a certain speaker at position $P_s = (x_s, y_s, z_s)$ a sound at position P = (x, y, z) is multiplied by f(θ,p) where θ denotes the angle between the sound source and the speaker. The cosine of this angle is calculated as the scalar product $(x, y, z).(x_s, y_s, z_s)$. For a speaker setup on a sphere or a circle with radius 1 and a sound sources at distance r we get

$$f_{inph}(P,P_s,p) = \left(\frac{xx_s + yy_s + zz_s + r}{2r}\right)^p \qquad (21.)$$
$$where\ r = \sqrt{x^2 + y^2 + z^2}$$

If the sounds are positioned on the unit sphere we get

$$f_{inph}(P,P_s,p) = \left(\frac{xx_s + yy_s + zz_s + 1}{2}\right)^p \qquad (22.)$$

in Cartesian coordinates and

$$f_{inph}(P,P_s,p) = \left(\frac{1+\cos(\theta-\theta_s)\cos(\delta)\cos(\delta_s)+\sin(\delta)\sin(\delta_s)}{2}\right)^p \qquad (23.)$$

in spherical coordinates.

Various Max/MSP tools for sound spatialization, ambisonic encoding and decoding and ambisonic panning can be downloaded from the website of Institute for Computer Music and Sound Technology ICST [8].

4.2. Computational costs

The complexity of the encoding formulas increases rapidly with the order. (Table 4 shows the encoding formula for $Y^1_{4,3}$ in spherical and Cartesian coordinates.)

$Y^1_{4,3}$	$\frac{1}{2}\sqrt{\frac{35}{2}}$ Cos[δ]^3 Cos[3 θ] Sin[δ]	$\frac{1}{4}\sqrt{70}$ (x^3 z - 3 x y^2 z)

Table 4: Encoding formula for $Y^1_{4,3}$

It is quite difficult to estimate the computational costs for the various techniques. They not only depend on the hardware used but also on the implementation of the formulas. The following is an assessment in 3D using Cartesian coordinates. Given that the coefficients are calculated already and the powers of x, y and z are stored during the calculations, there are about 0, 3, 16, 45, 96, 177, 300, …(ca. $(m+.5)^3$) multiplications for the orders 0, 1, 2, … for the encoding of each signal.

In the decoding process the matrix $\mathbf{C}^\mathbf{T}$ (which is calculated just once) is multiplied with the B-format. This needs $n*(m+1)^2$ multiplications.

The panning function (22.) needs 4 multiplications and 1 function call for every sound and speaker.

4.3. Conclusions

There are different ways to use ambisonics for rendering surround sound: 1) Calculate and save sounds in the B-format and decode the B-format signal for the required speaker setup, 2) store sounds and information about the sounds' positions separately and calculate the speaker signals with a panning function, 3) en- and decode the B-format signal in real-time, 4) produce the sounds and pan them in real time. The great advantage of the B-format is the possibility of adding up an arbitrary number of independent sounds, each with its own position or movement in a number of channels depending only on the chosen order. The disadvantages are the large number of channels for good spatial resolution and the considerable computational costs for high orders. Thus using the B-format (first and third approach) is reasonable only if many independent sound sources are treated. In most cases the second and forth approach perform better. With the second approach the information for the sounds and their position are stored without loss and further sound processing is still possible, with the fourth approach high orders can be used and the order can be changed continuously during performance.

The mathematics used in ambisonics theory is beyond the skills of non-scientists or non-engineers. Since panning functions are familiar and easy to visualize they provide a good didactical means for explaining ambisonics to laymen and for deriving encoding formulas and gains for in-phase decoding.

5. REFERENCES

[1] J. Daniel, Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia, Ph.D. Thesis, University of Paris VI, France, 2000, http://gyronymo.free.fr

[2] A. Sontacchi, R. Höldrich, Konzepte zur Schallfeldsynthese und Schallfeldreproduktion, Jahrestagung der ÖPG FA-Akustik, 2000, http://iem.at/projekte/publications/paper

[3] J. Daniel, R. Nicol, S. Moreau, Further Investigations of Higher Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging, in AES 114th Convention, Amsterdam, The Netherlands, 2003

[4] M. Neukom, Decoding Second Order Ambisonics to 5.1 Surround Systems, in AES 121st Convention, San Francisco, CA, USA, 2006

[5] Ville Pulkki, Virtual sound source positioning using Vector Base Amplitude Panning, J. Audio Eng. Soc., 45, June 1997.

[6] J.C. Bennet, K. Barker and F. O. Edeko, A new approach to the assessment of stereophonic sound system performance, J. Audio Eng. Soc., 33, May 1985.

[7] Institute for Computer Music and Sound Technology ICST, Zurich University of the Arts, http://www.icst.net