Spherical Harmonics and Ambisonics

Gordon Monro

School of Mathematics and Statistics
University of Sydney F07
NSW 2006
Australia

G.Monro@maths.usyd.edu.au

Version 0.2: 10th January 2000

Introduction

Ambisonics[1] is a surround-sound technique based on spherical harmonics, which has the great advantage that the encoding and decoding are separated in such a way that the same encoding can be used for many different speaker layouts. (For a general discussion of approaches to spatialisation of sound, see [Malham 98].)

Ambisonics in its "ideal" form involves the creation of a single sound image from the output of many cooperating speakers, some of which are out of phase with respect to the others. While this gives excellent results at a single listening point, the fact that some speakers are out of phase produces undesirable results for listeners spread over an area, as in a concert hall. Malham [Malham 92] has thus proposed an *in-phase correction* which ensures that all the speakers contributing to a single sound image are in phase. This note briefly reviews the encoding and decoding for first and second order Ambisonics and then discusses the in-phase corrections, in particular for the second-order case, finding a number of interesting solutions for both two-dimensional and three-dimensional decodings.

We generally follow the treatment of [Daniel *et al.*]; we consider only speakers symmetrically placed about a central point, at a reasonable distance from that point, so that wavefronts can be treated as planar and only the directions of the speakers are important. We ignore any contribution from the room containing the speakers.

Acknowledgements

[1] "Ambisonics" is a registered trademark of Nimbus Communications International.

The work described here was done while the author was a visitor to the Music Department, University of York, U.K. I wish to thank Tony Myatt for facilitating the visit, and Dave Malham for numerous enlightening conversations.

Spherical harmonics

We consider both two-dimensional spatialisation using "circular harmonics" (i.e. Fourier series) and three-dimensional spatialisation using spherical harmonics. We use real-valued, as opposed to complex-valued, functions throughout. A general reference for spherical harmonics is [Hobson].

Unfortunately basis functions for these harmonics are given in the literature in several forms, differing in the scaling factors used. For theoretical purposes an *orthonormal* basis is best, where the integral of the square of a basis function over the whole circle (or the whole sphere) has value 1. Furse and Malham [Furse] have introduced a "Furse-Malham" set of harmonics designed for practical purposes.

Additionally, spherical harmonics can be given in a Cartesian form or an angular form. We use the Cartesian form for conciseness. The angular form uses the *azimuth* angle A (measured anti-clockwise from straight ahead) and the *elevation* angle E; the conversion is:

two-dimensional: $x = \cos A; \quad y = \sin A;$

three-dimensional: $x = \cos A \cos E; \quad y = \sin A \cos E; \quad z = \sin E.$

Note that when using Cartesian coordinates we are assuming implicitly that in 2D the point (x, y) lies on the circle $x^2 + y^2 = 1$ and in 3D the point (x, y, z) lies on the sphere $x^2 + y^2 + z^2 = 1$. We thus refer to (x, y) or (x, y, z) as a *direction*. Note also that in work with Ambisonics the x axis points to the front and the y axis points to the left.

The Furse-Malham and orthonormal harmonics are listed in Table 1 (next page). We use the suffixes FM ("Furse-Malham"), P ("planar") and S ("spherical") to differentiate the three sets. The Furse-Malham set is the same in both 2D and 3D, except that in 2D only W, X, Y, U and V are used (in each case with $z = 0$). In this note, we work with the orthonormal harmonics, and rescale to the Furse-Malham harmonics when required.

Encoding

We discuss the first-order planar case using orthonormal harmonics. Consider a point sound source ψ concentrated at (x_0, y_0) and of unit intensity (so we are in fact dealing with a Dirac delta function). We use the notation $\sigma(W_P)$ for the signal on the W

Furse-Malham	**Orthonormal 2D**	**Orthonormal 3D**
$W_{FM} = \frac{1}{\sqrt{2}}$	$W_P = \frac{1}{\sqrt{2\pi}}$	$W_S = \sqrt{\frac{1}{4\pi}}$
$X_{FM} = x$	$X_P = \frac{1}{\sqrt{\pi}}\,x$	$X_S = \sqrt{\frac{3}{4\pi}}\,x$
$Y_{FM} = y$	$Y_P = \frac{1}{\sqrt{\pi}}\,y$	$Y_S = \sqrt{\frac{3}{4\pi}}\,y$
$Z_{FM} = z$	—	$Z_S = \sqrt{\frac{3}{4\pi}}\,z$
$R_{FM} = \frac{1}{2}(3z^2 - 1)$	—	$R_S = \sqrt{\frac{15}{16\pi}} \cdot \frac{1}{\sqrt{3}}\,(3z^2 - 1)$
$S_{FM} = 2xz$	—	$S_S = \sqrt{\frac{15}{16\pi}}\,2xz$
$T_{FM} = 2yz$	—	$T_S = \sqrt{\frac{15}{16\pi}}\,2yz$
$U_{FM} = x^2 - y^2$	$U_P = \frac{1}{\sqrt{\pi}}\,(x^2 - y^2)$	$U_S = \sqrt{\frac{15}{16\pi}}\,(x^2 - y^2)$
$V_{FM} = 2xy$	$V_P = \frac{1}{\sqrt{\pi}}\,2xy$	$V_S = \sqrt{\frac{15}{16\pi}}\,2xy$

Table 1

channel, and so on. The suffix "P" reminds us that this is a planar case. Then

$$\sigma(W_P) = \int_{\text{circle}} \frac{1}{\sqrt{2\pi}}\,\psi\,ds = \frac{1}{\sqrt{2\pi}};$$
$$\sigma(X_P) = \int_{\text{circle}} \frac{1}{\sqrt{\pi}}\,x\,\psi\,ds = \frac{1}{\sqrt{\pi}}\,x_0;$$
$$\sigma(Y_P) = \int_{\text{circle}} \frac{1}{\sqrt{\pi}}\,y\,\psi\,ds = \frac{1}{\sqrt{\pi}}\,y_0;$$

where ds is an element of the circumference of the circle. Similar results obtain for the other harmonics.

Basic decoding

Here we recapitulate the basic decoding theory for planar first-order Ambisonics, following the exposition in [Daniel *et al.*], but using the orthonormal harmonics

$$W_P = \frac{1}{\sqrt{2\pi}}, \quad X_P = \frac{1}{\sqrt{\pi}}\,x, \quad Y_P = \frac{1}{\sqrt{\pi}}\,y.$$

Consider n speakers, located in the directions $(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)$. Write

$$M_e = \begin{pmatrix} \frac{1}{\sqrt{2\pi}} & \frac{1}{\sqrt{2\pi}} & \cdots & \frac{1}{\sqrt{2\pi}} \\ \frac{1}{\sqrt{\pi}}\,x_1 & \frac{1}{\sqrt{\pi}}\,x_2 & \cdots & \frac{1}{\sqrt{\pi}}\,x_n \\ \frac{1}{\sqrt{\pi}}\,y_1 & \frac{1}{\sqrt{\pi}}\,y_2 & \cdots & \frac{1}{\sqrt{\pi}}\,y_n \end{pmatrix}.$$

The decoding matrix M_d is given by

$$M_d = M_e^T (M_e M_e^T)^{-1},$$

where M_e^T is the transpose of M_e. The product $M_e M_e^T$ is the following symmetric 3×3 matrix:

$$\begin{pmatrix} \frac{n}{2\pi} & \frac{1}{\pi\sqrt{2}} \Sigma_i x_i & \frac{1}{\pi\sqrt{2}} \Sigma_i y_i \\ \frac{1}{\pi\sqrt{2}} \Sigma_i x_i & \frac{1}{\pi} \Sigma_i x_i^2 & \frac{1}{\pi} \Sigma_i x_i y_i \\ \frac{1}{\pi\sqrt{2}} \Sigma_i y_i & \frac{1}{\pi} \Sigma_i x_i y_i & \frac{1}{\pi} \Sigma_i y_i^2 \end{pmatrix},$$

where Σ_i means $\Sigma_{i=1}^n$.

For speakers arranged in a regular polygon, all the off-diagonal entries in $M_e M_e^T$ are zero, and $\Sigma_i x_i^2 = \Sigma_i y_i^2 = \frac{n}{2}$ [Daniel *et al.*]. Thus for a regular polygon,

$$M_e M_e^T = \begin{pmatrix} \frac{n}{2\pi} & 0 & 0 \\ 0 & \frac{n}{2\pi} & 0 \\ 0 & 0 & \frac{n}{2\pi} \end{pmatrix}$$

and

$$\begin{aligned} M_d &= M_e^T (M_e M_e^T)^{-1} \\ &= \begin{pmatrix} \frac{1}{\sqrt{2\pi}} & \frac{1}{\sqrt{\pi}} x_1 & \frac{1}{\sqrt{\pi}} y_1 \\ \frac{1}{\sqrt{2\pi}} & \frac{1}{\sqrt{\pi}} x_2 & \frac{1}{\sqrt{\pi}} y_2 \\ \vdots & \vdots & \vdots \\ \frac{1}{\sqrt{2\pi}} & \frac{1}{\sqrt{\pi}} x_n & \frac{1}{\sqrt{\pi}} y_n \end{pmatrix} \begin{pmatrix} \frac{2\pi}{n} & 0 & 0 \\ 0 & \frac{2\pi}{n} & 0 \\ 0 & 0 & \frac{2\pi}{n} \end{pmatrix} \\ &= \begin{pmatrix} \frac{\sqrt{2\pi}}{n} & \frac{2\sqrt{\pi}}{n} x_1 & \frac{2\sqrt{\pi}}{n} y_1 \\ \frac{\sqrt{2\pi}}{n} & \frac{2\sqrt{\pi}}{n} x_2 & \frac{2\sqrt{\pi}}{n} y_2 \\ \vdots & \vdots & \vdots \\ \frac{\sqrt{2\pi}}{n} & \frac{2\sqrt{\pi}}{n} x_n & \frac{2\sqrt{\pi}}{n} y_n \end{pmatrix}. \end{aligned}$$

Let out_i be the output signal for the ith speaker. The above equations mean that the decoding is

$$\begin{pmatrix} out_1 \\ out_2 \\ \vdots \\ out_n \end{pmatrix} = M_d \begin{pmatrix} \sigma(W_P) \\ \sigma(X_P) \\ \sigma(Y_P) \end{pmatrix}.$$

That is, for $i = 1, \ldots, n$

$$out_i = \frac{\sqrt{2\pi}}{n} \sigma(W_P) + \frac{2\sqrt{\pi}}{n} x_i \, \sigma(X_P) + \frac{2\sqrt{\pi}}{n} y_i \, \sigma(Y_P). \tag{1}$$

This is what [Daniel *et al.*] refer to as the *basic decoding* and [Furse] calls the *idealised response*. We will call it the *ideal decoding*.

We consider only symmetric speaker arrays in this note because the fact that $M_e M_e^T$ is a multiple of the identity matrix makes its inverse trivial to compute and thus simplifies calculations enormously.

For a point source of unit intensity in the direction (x_0, y_0) we have

$$\sigma(W_P) = \frac{1}{\sqrt{2\pi}}, \quad \sigma(X_P) = \frac{1}{\sqrt{\pi}}\, x_0, \quad \sigma(Y_P) = \frac{1}{\sqrt{\pi}}\, y_0.$$

The ideal decoding for this point source then gives

$$\begin{aligned} out_i &= \frac{\sqrt{2\pi}}{n}\frac{1}{\sqrt{2\pi}} + \frac{2\sqrt{\pi}}{n}\, x_i \,\frac{1}{\sqrt{\pi}}\, x_0 + \frac{2\sqrt{\pi}}{n}\, y_i \,\frac{1}{\sqrt{\pi}}\, x_0 \\ &= \frac{1}{n}(1 + 2x_i x_0 + 2y_i y_0). \end{aligned}$$

[Daniel *et al.*] show that close to the centre of the listening space, the ideal decoding gives the optimum result for the velocity vector of the resulting soundfield.

Rescaling for the Furse-Malham set

In equation (1), the coefficients $\frac{\sqrt{2\pi}}{n}$, $\frac{2\sqrt{\pi}}{n}\, x_i$, $\frac{2\sqrt{\pi}}{n}\, y_i$, of $\sigma(W_P)$, $\sigma(X_P)$, $\sigma(Y_P)$ respectively, are entries in the decoding matrix M_d, which was calculated with respect to the orthonormal harmonics $W_P = \frac{1}{\sqrt{2\pi}}$, $X_P = \frac{1}{\sqrt{\pi}}\, x$, $Y_P = \frac{1}{\sqrt{\pi}}\, y$.

Consider (for the first-order planar case) a set of differently scaled harmonics

$$W^* = \frac{K_W}{\sqrt{2\pi}}, \quad X^* = \frac{K_X}{\sqrt{\pi}}\, x, \quad Y^* = \frac{K_Y}{\sqrt{\pi}}\, y.$$

If we follow through the basic encoding and decoding theory using these differently scaled harmonics, we find firstly that the new signals are

$$\sigma(W^*) = K_W \sigma(W_P), \quad \sigma(X^*) = K_X \sigma(X_P), \quad \sigma(Y^*) = K_Y \sigma(Y_P)$$

and secondly that the coefficients in equation (1) become

$$\frac{\sqrt{2\pi}}{nK_W}, \quad \frac{2\sqrt{\pi}}{nK_X}\, x_i, \quad \frac{2\sqrt{\pi}}{nK_Y}\, y_i.$$

The calculated output out_i of the ith speaker is thus unchanged, as one would expect.

The practical conclusion: to obtain the decoding matrix for the rescaled harmonics, take the decoding matrix for the orthonormal harmonics and divide the entries by K_W, K_X or K_Y as appropriate. The same calculation goes through for higher-order harmonics.

The rescaling constants for the Furse-Malham harmonics with respect to the orthonormal harmonics are:

2D: $K_W = K_X = K_Y = K_Z = K_R = K_S = K_T = K_U = K_V = \sqrt{\pi}$.

3D: $K_W = \sqrt{2\pi}$; $\quad K_X = K_Y = K_Z = \sqrt{\frac{4\pi}{3}}$; $\quad K_R = \sqrt{\frac{4\pi}{5}}$;

$K_S = K_T = K_U = K_V = \sqrt{\frac{16\pi}{15}}$.

In-phase decoding

As an example of the ideal decoding, consider four speakers arranged in a square at $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, $(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, $(-\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ and $(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}})$ (i.e. front right, rear right, rear left and front left). The decoding of a point source of unit intensity is, according to equation (1)

$$out_1 = out_4 = \frac{1}{4}(1+\sqrt{2}) \approx 0.6036; \quad out_2 = out_3 = \frac{1}{4}(1-\sqrt{2}) \approx -0.1036.$$

The two rear speakers are thus transmitting signals out of phase with respect to the signals from the front two speakers.

[Daniel *et al.*] introduce *correcting gains* g_0, g_1, giving a more general decoding

$$out_i = g_0 \frac{\sqrt{2\pi}}{n} \sigma(W_P) + g_1 \left(\frac{2\sqrt{\pi}}{n} x_i \, \sigma(X_P) + \frac{2\sqrt{\pi}}{n} y_i \, \sigma(Y_P) \right).$$

The general decoding for a point source of unit intensity in the direction (x_0, y_0) works out as

$$out_i = \frac{1}{n}\big(g_0 + 2g_1(x_i x_0 + y_i y_0)\big).$$

If we write α for the angle between the position vectors (x_0, y_0) and (x_i, y_i), this general decoding becomes

$$out_i = \frac{1}{n}(g_0 + 2g_1 \cos\alpha).$$

For in-phase decoding we require $out_i \geq 0$ for all i. It is clear that the worst case is where $\cos\alpha = -1$, and this leads to

$$g_1 = \frac{g_0}{2}. \tag{2}$$

Equation (2) is Malham's in-phase decoding correction.

The theory of correcting gains is independent of the scaling used for the harmonics.

Planar second-order decoding

We proceed as in the first-order case, but there are two additional harmonics,

$$U_P = \frac{1}{\sqrt{\pi}}(x^2 - y^2) \quad \text{and} \quad V_P = \frac{1}{\sqrt{\pi}} 2xy.$$

The encoding of a point source of unit intensity in the direction (x_0, y_0) now has two additional signals

$$\sigma(U_P) = \frac{1}{\sqrt{\pi}}(x_0^2 - y_0^2); \quad \sigma(V_P) = \frac{1}{\sqrt{\pi}} 2x_0^2 y_0^2.$$

The matrix $M_e M_e^T$ 5×5 for this case, and for speakers arranged in a regular polygon, $M_e M_e^T$ is $n/2\pi$ times the 5×5 identity matrix. It follows that the ideal decoding for the second-order planar case is

$$\begin{aligned} out_i = &\frac{\sqrt{2\pi}}{n}\,\sigma(W_P) + \frac{2\sqrt{\pi}}{n}\,x_i\,\sigma(X_P) + \frac{2\sqrt{\pi}}{n}\,y_i\,\sigma(Y_P) \\ &+\frac{2\sqrt{\pi}}{n}\,(x_i^2 - y_i^2)\,\sigma(U_P) + \frac{2\sqrt{\pi}}{n}\,2x_i y_i\,\sigma(V_P). \end{aligned}$$

Inserting correcting gains g_0, g_1, g_2 gives a general decoding

$$\begin{aligned} out_i = g_0\,&\frac{\sqrt{2\pi}}{n}\,\sigma(W_P) + g_1\left(\frac{2\sqrt{\pi}}{n}\,x_i\,\sigma(X_P) + \frac{2\sqrt{\pi}}{n}\,y_i\,\sigma(Y_P)\right) \\ &+g_2\left(\frac{2\sqrt{\pi}}{n}\,(x_i^2 - y_i^2)\,\sigma(U_P) + \frac{2\sqrt{\pi}}{n}\,2x_i y_i\,\sigma(V_P)\right). \end{aligned}$$

The general decoding for a point source of unit intensity in the direction (x_0, y_0) is thus

$$out_i = \frac{1}{n}\big(g_0 + 2g_1(x_i x_0 + y_i y_0) + 2g_2(x_i^2 - y_i^2)(x_0^2 - y_0^2) + 4x_i y_i x_0 y_0)\big). \qquad (3)$$

Solutions for planar second-order in-phase decoding

As before, let α be the angle between the position vectors (x_0, y_0) and (x_i, y_i). Equation (3) becomes

$$out_i = \frac{1}{n}(g_0 + 2g_1\cos\alpha + 2g_2\cos 2\alpha).$$

Now write $p_1 = g_1/g_0$ and $p_2 = g_2/g_0$. Then

$$out_i = \frac{g_0}{n}(1 + 2p_1\cos\alpha + 2p_2\cos 2\alpha).$$

Let us drop the factor of $\frac{g_0}{n}$ and set

$$f(\alpha) = 1 + 2p_1\cos\alpha + 2p_2\cos 2\alpha. \qquad (4)$$

For in-phase decoding we need to find values of p_1 and p_2 such that $f(\alpha) \geq 0$ for all α. We proceed by finding the minimum value of $f(\alpha)$, and to that end we calculate

$$\frac{df}{d\alpha} = -2p_1\sin\alpha - 4p_2\sin 2\alpha = -2\sin\alpha(p_1 + 4p_2\cos\alpha).$$

For $\frac{df}{d\alpha}$ to be zero, either $\sin\alpha = 0$ or $\cos\alpha = \frac{-p_1}{4p_2}$. If $\sin\alpha = 0$, either $\alpha = 0$, giving in fact a maximum value for $f(\alpha)$, or $\alpha = \pi$, giving $f(\alpha) = 1 - 2p_1 + 2p_2$. Thus we obtain the condition

$$p_2 \geq p_1 - \frac{1}{2}$$

for $f(\alpha)$ to be non-negative at $\alpha = \pi$.

We note that we need $p_1 \leq 4p_2$ for the equation $\cos\alpha = \frac{-p_1}{4p_2}$ to have real solutions. In fact there are two possible shapes for the graph of $f(\alpha)$ against α: if $p_1 \geq 4p_2$ the shape is similar to that in Figure 2 below and the minimum value occurs at $\alpha = \pi$; if $p_1 < 4p_2$ the shape is similar to that in Figure 3 and there are two minima, occurring at the two values of $\cos^{-1}(\frac{-p_1}{4p_2})$.

If $p_1 \leq 4p_2$ and we set $\cos\alpha = \frac{-p_1}{4p_2}$ in equation (4), the condition $f(\alpha) \geq 0$ becomes

$$1 + 2p_1\left(\frac{-p_1}{4p_2}\right) + 2p_2\left(\frac{2p_1^2}{16p_2^2} - 1\right) \geq 0.$$

This reduces to

$$\frac{p_1^2}{(1/\sqrt{2})^2} + \frac{(p_2 - 1/4)^2}{(1/4)^2} \leq 1, \tag{5}$$

which is the equation of the interior of an ellipse.

We obtain the following picture.

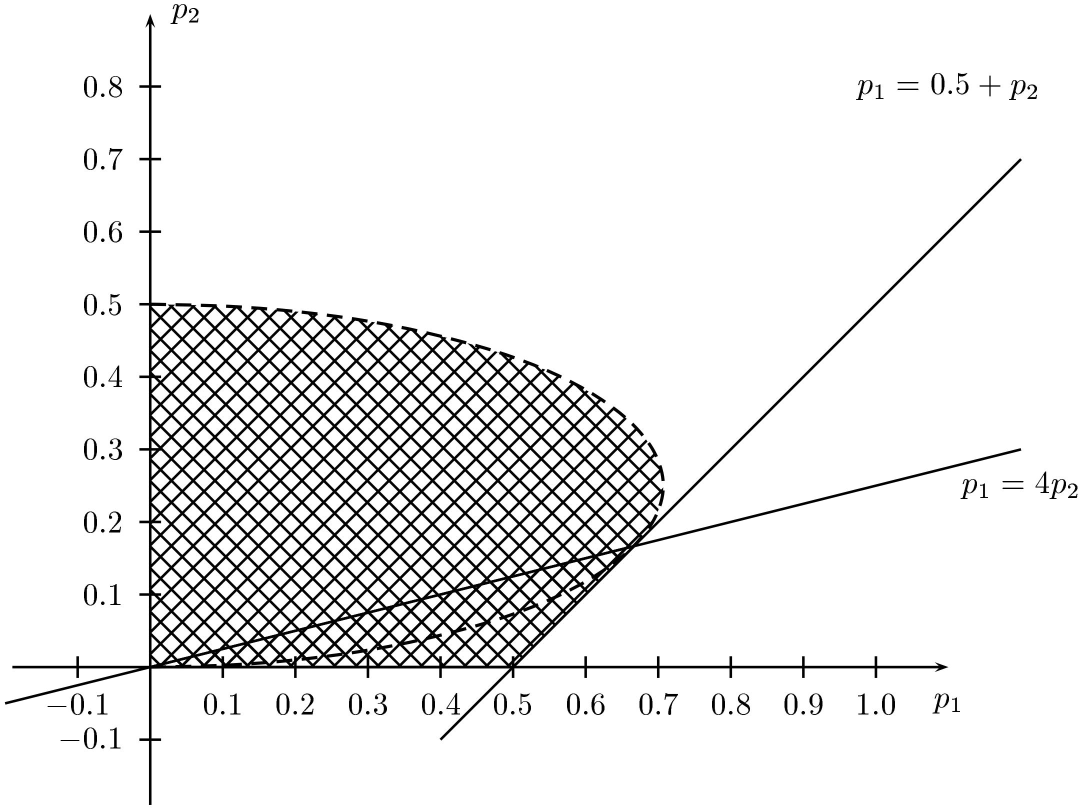

Figure 1

The two lines intersect at $(\frac{2}{3}, \frac{1}{3})$, which is on the ellipse. In-phase solutions are shown shaded.

Interesting solutions are on the boundary of the ellipse. We discuss the following below.

- The "smooth" solution: $p_1 = 0.6667$, $p_2 = 0.1667$.
- The "extends first order" solution: $p_1 = 0.5000$, $p_2 = 0.4268$.
- The "maximum energy" solution: $p_1 = 0.6795$, $p_2 = 0.3192$.
- The "maximum front-back ratio" solution: $p_1 = 0.6667$, $p_2 = 0.3333$.
- The "maximum integrated front-back ratio" solution: $p_1 = 0.7071$, $p_2 = 0.25$.
- Furse's solution: $p_1 = 0.658$, $p_2 = 0.342$.

The "smooth" solution

This occurs at the point in Figure 1 where the two lines and the ellipse all intersect. It is the "last" point where $f(\alpha)$ has only a single minimum. Figure 2 shows the graph of $f(\alpha)$ against α in both Cartesian and polar forms (solid line), and also the first-order in-phase solution $p_1 = 0.5$, $p_2 = 0$ (dotted line).

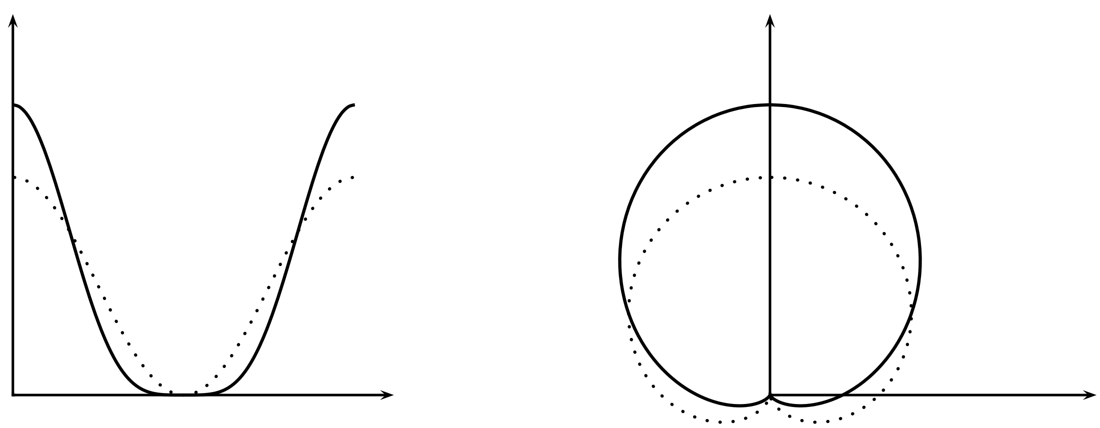

Figure 2

The polar plot is a "source directivity response", showing the amount of signal sent to speakers in various directions for a source located straight ahead.

The "extends first order" solution

This solution has the same value of p_1 as for first-order in-phase decoding (see equation (2)). It is the "lumpiest" of the solutions discussed here: see Figure 3. A possible use for this solution will be discussed later.

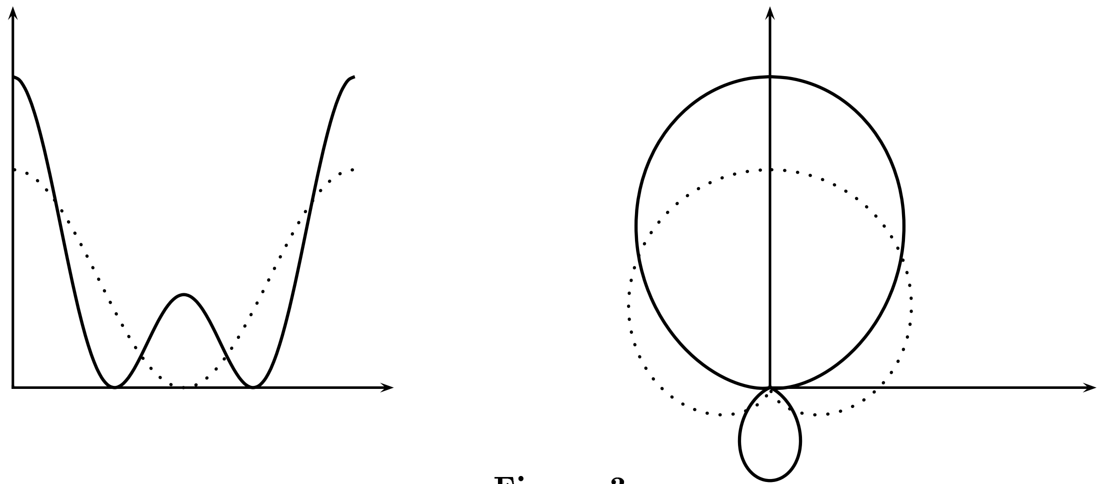

Figure 3

It is not clear how important it is for the source directivity response to have a single minimum. Intuitively, the relatively large contribution from the rear speakers in the above example will have undesirable effects, but listening tests are needed to determine how much "lumpiness" can be tolerated.

The "maximum energy" solution

It is shown in [Daniel *et al.*] that the quantity

$$r_E = \frac{2(g_0 g_1 + g_1 g_2)}{g_0^2 + 2(g_1^2 + g_2^2)} = \frac{2(p_1 + p_1 p_2)}{1 + 2(p_1^2 + p_2^2)}$$

compares the magnitude of the energy vector (also known as the intensity vector) of the reconstruction to that of the the energy vector of the original source; for a good reconstruction r_E should be close to 1. We therefore seek the solution on the ellipse (5) that maximises r_E. This is analytically messy, but is easily found numerically to be $p_1 = 0.6795$, $p_2 = 0.3192$.

The "maximum front-back ratio" solution

The *front-back ratio* is defined to be, for the decoding of a source located straight ahead,

$$\frac{\text{max. signal in front semicircle/hemisphere}}{\text{max. signal in rear semicircle/hemisphere}}.$$

A straightforward calculation shows that in our situation the front-back ratio is

$$\frac{1 + 2p_1 + 2p_2}{1 - 2p_2} \quad \text{if } p_2 \le p_1/2$$

$$\frac{1 + 2p_1 + 2p_2}{1 - 2p_1 + 2p_2} \quad \text{if } p_2 \ge p_1/2$$

The point on the ellipse (5) giving the maximum front-to-back ratio was found to be $p_1 = 2/3$, $p_2 = 1/3$, and for these values of p_1, p_2 the maximum front-back ratio is 9, or approximately 19 dB. Malham (private communication) has obtained higher front-back ratios by allowing small out-of-phase signals in some speakers.

The "maximum integrated front-back ratio" solution

Define the *integrated front-back ratio* as

$$\frac{\text{average signal in front semicircle/hemisphere}}{\text{average signal in rear semicircle/hemisphere}}.$$

In our situation the integrated front-back ratio is easily calculated, for in-phase solutions, to be

$$\frac{\pi + 4g_1}{\pi - 4g_1}.$$

The maximum value of this quantity is obtained when g_1 has its maximum value of $1/\sqrt{2}$, and with this value of g_1, the integrated front-back ratio is 19.063, or approximately 25.6dB. This solution, with $g_1 = 1/\sqrt{2}$ and $g_2 = 0.25$, is graphed in Figure 4. The solution is an attractive one, appearing in a conspicuous spot at the end of the semi-major axis of the ellipse, and with only slight "lumpiness" at the straight-behind location.

It may seem surprising that g_2 does not appear in the result for the maximum integrated front-back ratio, but this result is only valid for in-phase solutions, and the presence of g_2 means that g_1 can take values greater than 0.5, and still yield an in-phase solution.

The ordinary front-back ratio is arguably an easy-to-measure proxy for the integrated front-back ratio.

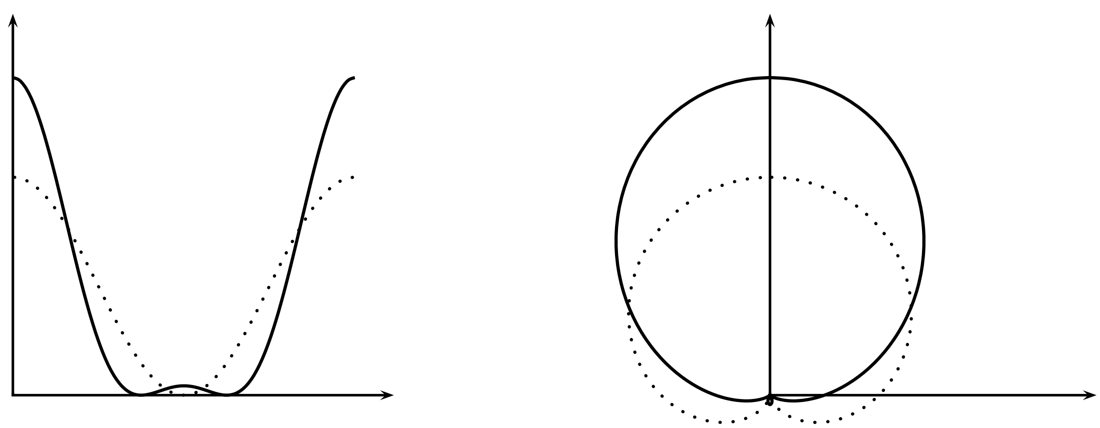

Figure 4

Furse's solution

Furse [Furse] has posted numerical results for several 2D and 3D speaker configurations; these results include in-phase solutions (which Furse calls *controlled opposites*). Unfortunately Furse has not released his source code, so his method of calculation remains obscure. Nonetheless, it can be observed that Furse's procedures correspond to those of [Daniel *et al.*]. His ideal decodings are exactly those given by the calculation $M_d = M_e^T(M_eM_e^T)^{-1}$, scaled for the Furse-Malham harmonics. Furse's in-phase solutions are evidently obtained by using correcting gains as above, and for sufficiently symmetric polygons, Furse's correcting gains are

$$p_1 = 0.658; \quad p_2 = 0.342.$$

This solution corresponds to a point on the ellipse (5).

Three-dimensional first-order decoding

The zeroth- and first-order orthonormal spherical harmonics are

$$W_S = \sqrt{\frac{1}{4\pi}}; \quad X_S = \sqrt{\frac{3}{4\pi}}\,x; \quad Y_S = \sqrt{\frac{3}{4\pi}}\,y; \quad Z_S = \sqrt{\frac{3}{4\pi}}\,z.$$

Here (x, y, z) is a point on the sphere $x^2 + y^2 + z^2 = 1$; as before we regard such a point as a direction.

If we consider a point source of unit intensity in the direction (x_0, y_0, z_0), the corresponding signals are

$$\sigma(W_S) = \sqrt{\frac{1}{4\pi}}; \quad \sigma(X_S) = \sqrt{\frac{3}{4\pi}}\,x_0; \quad \sigma(Y_S) = \sqrt{\frac{3}{4\pi}}\,y_0; \quad \sigma(Z_S) = \sqrt{\frac{3}{4\pi}}\,z_0.$$

Suppose that there are n speakers located in the directions $(x_1, y_1, z_1),\ldots,(x_n, y_n, z_n)$. As before, we form the matrix

$$M_e = \begin{pmatrix} \sqrt{\frac{1}{4\pi}} & \sqrt{\frac{1}{4\pi}} & \cdots & \sqrt{\frac{1}{4\pi}} \\ \sqrt{\frac{3}{4\pi}}\,x_1 & \sqrt{\frac{3}{4\pi}}\,x_2 & \cdots & \sqrt{\frac{3}{4\pi}}\,x_n \\ \sqrt{\frac{3}{4\pi}}\,y_1 & \sqrt{\frac{3}{4\pi}}\,y_2 & \cdots & \sqrt{\frac{3}{4\pi}}\,y_n \\ \sqrt{\frac{3}{4\pi}}\,z_1 & \sqrt{\frac{3}{4\pi}}\,z_2 & \cdots & \sqrt{\frac{3}{4\pi}}\,z_n \end{pmatrix}$$

For a sufficiently symmetric arrangement of speakers $M_eM_e^T$ is $\frac{n}{4\pi}$ times the 4×4 identity matrix and our ideal decoding matrix $M_d = M_e^T(M_eM_e^T)^{-1}$ then becomes

$$M_d = \begin{pmatrix} \frac{4\pi}{n}\sqrt{\frac{1}{4\pi}} & \frac{4\pi}{n}\sqrt{\frac{3}{4\pi}}\,x_1 & \frac{4\pi}{n}\sqrt{\frac{3}{4\pi}}\,y_1 & \frac{4\pi}{n}\sqrt{\frac{3}{4\pi}}\,z_1 \\ \frac{4\pi}{n}\sqrt{\frac{1}{4\pi}} & \frac{4\pi}{n}\sqrt{\frac{3}{4\pi}}\,x_2 & \frac{4\pi}{n}\sqrt{\frac{3}{4\pi}}\,y_2 & \frac{4\pi}{n}\sqrt{\frac{3}{4\pi}}\,z_2 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{4\pi}{n}\sqrt{\frac{1}{4\pi}} & \frac{4\pi}{n}\sqrt{\frac{3}{4\pi}}\,x_n & \frac{4\pi}{n}\sqrt{\frac{3}{4\pi}}\,y_n & \frac{4\pi}{n}\sqrt{\frac{3}{4\pi}}\,z_n \end{pmatrix}$$

Thus for ideal decoding the output from speaker i is

$$out_i = \frac{4\pi}{n}\left(\sqrt{\frac{1}{4\pi}}\,\sigma(W_S) + \sqrt{\frac{3}{4\pi}}\,x_i\sigma(X_S) + \sqrt{\frac{3}{4\pi}}\,y_i\sigma(Y_S) + \sqrt{\frac{3}{4\pi}}\,z_i\sigma(Z_S)\right).$$

We introduce correcting gains h_0, h_1, corresponding to the gains g_0, g_1 introduced in the planar case, and obtain the more general decoding

$$out_i = \frac{4\pi h_0}{n}\sqrt{\frac{1}{4\pi}}\,\sigma(W_S) + \frac{4\pi h_1}{n}\sqrt{\frac{3}{4\pi}}\big(x_i\sigma(X_S) + y_i\sigma(Y_S) + z_i\sigma(Z_S)\big).$$

The general decoding for a point source of unit intensity located in the direction (x_0, y_0, z_0) works out as

$$out_i = \frac{h_0}{n} + \frac{3h_1}{n}(x_ix_0 + y_iy_0 + z_iz_0).$$

If we switch to vector notation and write $\mathbf{u}_i = (x_i, y_i, z_i)$, $\mathbf{u}_0 = (x_0, y_0, z_0)$, then the above equation can be written

$$out_i = \frac{1}{n}(h_0 + 3h_1\mathbf{u}_i \cdot \mathbf{u}_0). \qquad (6)$$

In-phase decoding

Since $\mathbf{u}_0$ and $\mathbf{u}_i$ are unit vectors, $\mathbf{u}_i \cdot \mathbf{u}_0 \geq -1$. Thus equation (6) gives the following in-phase decoding condition for 3D first-order decoding:

$$h_1 = h_0/3.$$

We note that even if the source signal lies in the xy plane, (so $\sigma(Z_S) = 0$), the condition for in-phase decoding differs from that for the planar case.

Three-dimensional second-order decoding

We proceed as in the three-dimensional first-order case, but there are five additional harmonics:

$$R_S = \sqrt{\frac{15}{16\pi}} \cdot \sqrt{\frac{1}{\sqrt{3}}}\,(3z^2 - 1); \quad S_S = \sqrt{\frac{15}{16\pi}}\,2xz; \quad T_S = \sqrt{\frac{15}{16\pi}}\,2yz;$$

$$U_S = \sqrt{\frac{15}{16\pi}}\,(x^2 - y^2); \quad V_S = \sqrt{\frac{15}{16\pi}}\,2xy.$$

These are orthonormal harmonics; the rescalings required for the Furse-Malham harmonics have been discussed above.

For a sufficiently symmetrical arrangement of speakers the matrix $M_e M_e^T$ is $\frac{n}{4\pi}$ times the 9×9 identity matrix. The main example of such a symmetrical arrangement is a dodecahedral array, i.e. 12 speakers at the face centres of a regular dodecahedron; a cubical arrangement does not have the required symmetry. In the sufficiently symmetrical case, the ideal decoding for the second-order 3D case, for a speaker in the direction (x_i, y_i, z_i), is

$$\begin{aligned} out_i = \frac{4\pi}{n}\Bigg(& \sqrt{\frac{1}{4\pi}}\,\sigma(W_S) + \sqrt{\frac{3}{4\pi}}\,x_i\sigma(X_S) + \sqrt{\frac{3}{4\pi}}\,y_i\sigma(Y_S) + \sqrt{\frac{3}{4\pi}}\,z_i\sigma(Z_S) \\ & + \sqrt{\frac{15}{16\pi}}\cdot\sqrt{\frac{1}{\sqrt{3}}}\,(3z_i^2-1)\sigma(R_S) + \sqrt{\frac{15}{16\pi}}\,2x_iy_i\sigma(S_S) + \sqrt{\frac{15}{16\pi}}\,2y_iz_i\sigma(T_S) \\ & + \sqrt{\frac{15}{16\pi}}\,(x_i^2-y_i^2)\sigma(U_S) + \sqrt{\frac{15}{16\pi}}\,2x_iy_i\sigma(V_S)\Bigg). \end{aligned}$$

Inserting correcting gains h_0, h_1, h_2 gives a general decoding

$$\begin{aligned} out_i = {} & \frac{4\pi h_0}{n}\sqrt{\frac{1}{4\pi}} + \frac{4\pi h_1}{n}\big(x_i\sigma(X_S) + y_i\sigma(Y_S) + z_i\sigma(Z_S)\big) \\ & + \frac{4\pi h_2}{n}\sqrt{\frac{15}{16\pi}}\Bigg(\frac{1}{\sqrt{3}}(3z_i^2-1)\sigma(R_S) + 2x_iy_i\sigma(S_S) + 2y_iz_iy_0\sigma(T_S) \\ & \qquad + (x_i^2-y_i^2)\sigma(U_S) + 2x_iy_i\sigma(V_S)\Bigg). \end{aligned}$$

The general decoding for a point source of unit intensity in the direction (x_0, y_0, z_0) is then

$$\begin{aligned} & \frac{4\pi h_0}{n}\cdot\frac{1}{4\pi} + \frac{4\pi h_1}{n}\cdot\frac{3}{4\pi}(x_ix_0 + y_iy_0 + z_iz_0) \\ & + \frac{4\pi h_2}{n}\cdot\frac{15}{16\pi}\big(\frac{1}{3}(3z_i^2-1)(3z_0^2-1) + 4x_iy_ix_0y_0 + 4y_iz_iy_0z_0 \\ & + (x_i^2-y_i^2)(x_0^2-y_0^2) + 4x_iy_ix_0y_0\big). \end{aligned} \tag{7}$$

Solutions for three-dimensional second-order in-phase decoding

Let us write $\mathbf{u}_i = (x_i, y_i, z_i)$ and $\mathbf{u}_0 = (x_0, y_0, z_0)$ as before, and set α to be the angle between $\mathbf{u}_i$ and $\mathbf{u}_0$. It can be shown (see Appendix A) that

$$\frac{1}{3}(3z_i^2-1)(3z_0^2-1)+4x_iy_ix_0y_0+4y_iz_iy_0z_0+(x_i^2-y_i^2)(x_0^2-y_0^2)+4x_iy_ix_0y_0 = \frac{1}{3}+\cos 2\alpha.$$

The decoding of the point source (equation (7)) then becomes

$$out_i = \frac{1}{n}\big(h_0 + 3h_1 \cos\alpha + \frac{15}{4}h_2(\frac{1}{3} + \cos 2\alpha)\big).$$

Set $q_1 = h_1/h_0$ and $q_2 = h_2/h_0$ (so q_1, q_2 are the 3D analogues of the planar p_1, p_2). In the 3D case

$$out_i = \frac{h_0}{n}\big(1 + 3q_1 \cos\alpha + \frac{15}{4}q_2(\frac{1}{3} + \cos 2\alpha)\big).$$

Let us drop the factor of $\frac{h_0}{n}$ and set

$$g(\alpha) = 1 + 3q_1 \cos\alpha + \frac{15}{4}q_2(\frac{1}{3} + \cos 2\alpha).$$

For in-phase decoding we need to find values of q_1, q_2 such that $g(\alpha) \geq 0$ for all α. As in the planar case, we find the minimum value of $g(\alpha)$ by calculating $\frac{dg}{d\alpha}$.

$$\frac{dg}{d\alpha} = -3q_1 \sin\alpha - \frac{15}{4} \cdot 2\sin 2\alpha = -3\sin\alpha(q_1 + 5q_2\cos\alpha).$$

If $\frac{dg}{d\alpha} = 0$, either $\sin\alpha = 0$ or $\cos\alpha = \frac{-q_1}{5q_2}$. The situation is very similar to that for the planar case; we find that for in-phase solutions we need

$$q_2 \geq \frac{3q_1 - 1}{5}$$

and, if $q_1 \leq 5q_2$ we also need

$$\frac{q_1^2}{(1/\sqrt{3})^2} + \frac{(q_2 - 1/5)^2}{(1/5)^2} \leq 1.$$

This leads to the following picture.

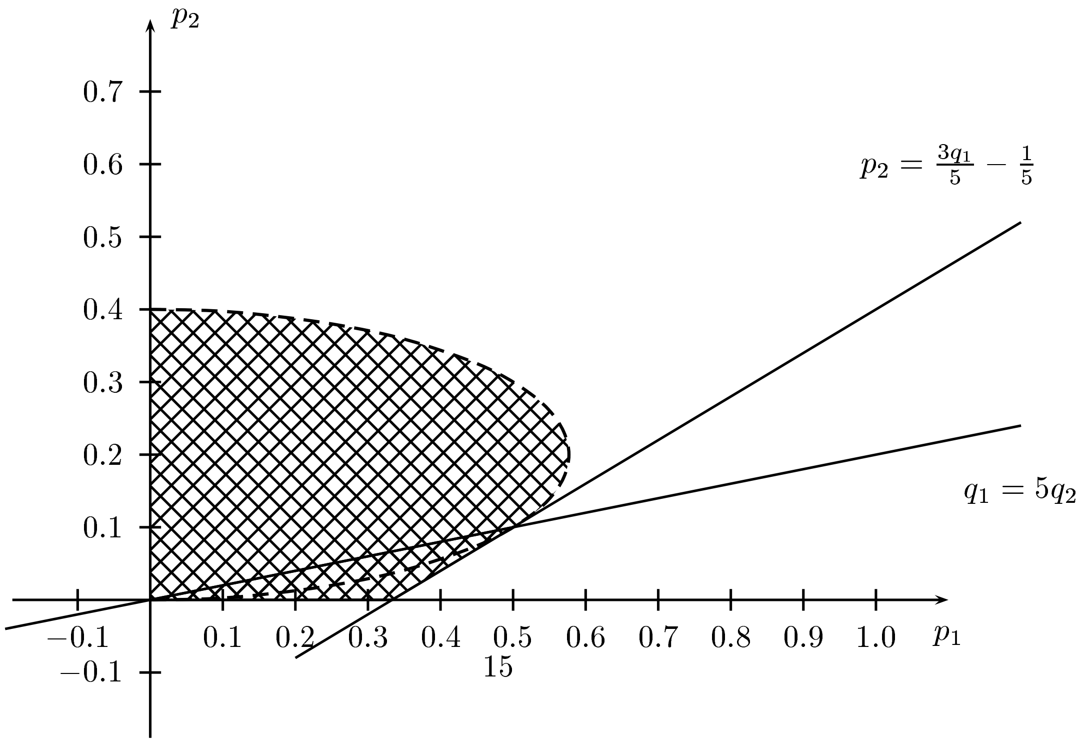

Figure 5

The two lines intersect at $(\frac{1}{2}, \frac{1}{10})$, which is on the ellipse. Again interesting solutions will be on the boundary of the ellipse, and we note the following, some of which are further discussed below.

- The "smooth" solution: $q_1 = 0.5$, $q_2 = 0.1$.
- The "extends first order" solution: $q_1 = 0.3333$, $q_2 = 0.3633$.
- The "maximum front-back ratio" solution: $q_1 = 0.5714$, $q_2 = 0.2282$.
- The "maximum integrated front-back ratio" solution: $q_1 = 0.5774$, $q_2 = 0.2$.
- Furse's solution: $q_1 = 0.504$, $q_2 = 0.102$.

The "smooth" solution

This solution is graphed in Figure 6.

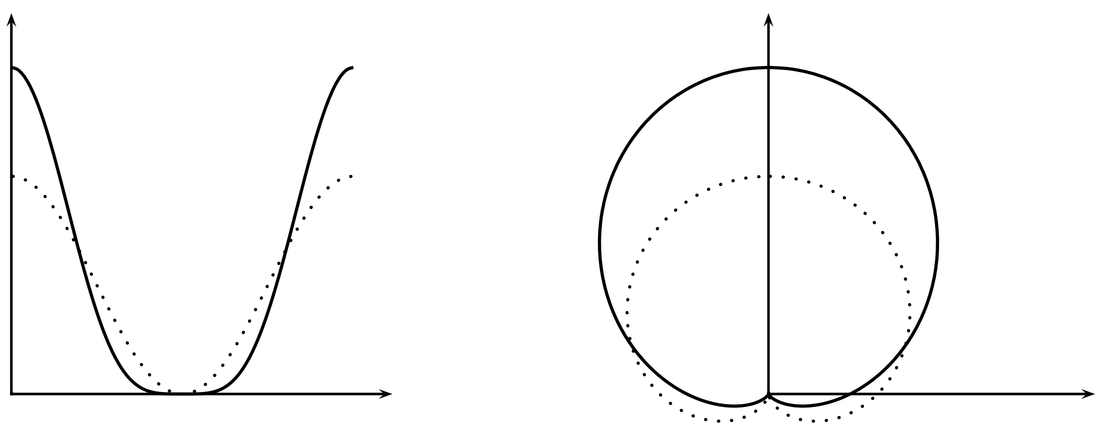

Figure 6

The "maximum front-back ratio" solution

Numerical exploration locates this at $q_1 = 0.5714$, $q_2 = 0.2282$; the value of the ratio at this point appears to be exactly 9.0, which is the same value as was obtained in the 2D case..

The "maximum integrated front-back ratio" solution

For the 3D case the integrated front-back ratio for parameters q_1, q_2 works out to be

$$\frac{2 + 3g_1}{2 - 3g_1},$$

again without explicit dependence on q_2. The maximum value of this ratio again occurs at the end of the semi-major axis of the ellipse, when $g_1 = 1/\sqrt{3}$, and the ratio is then 13.928. This solution is graphed in Figure 7.

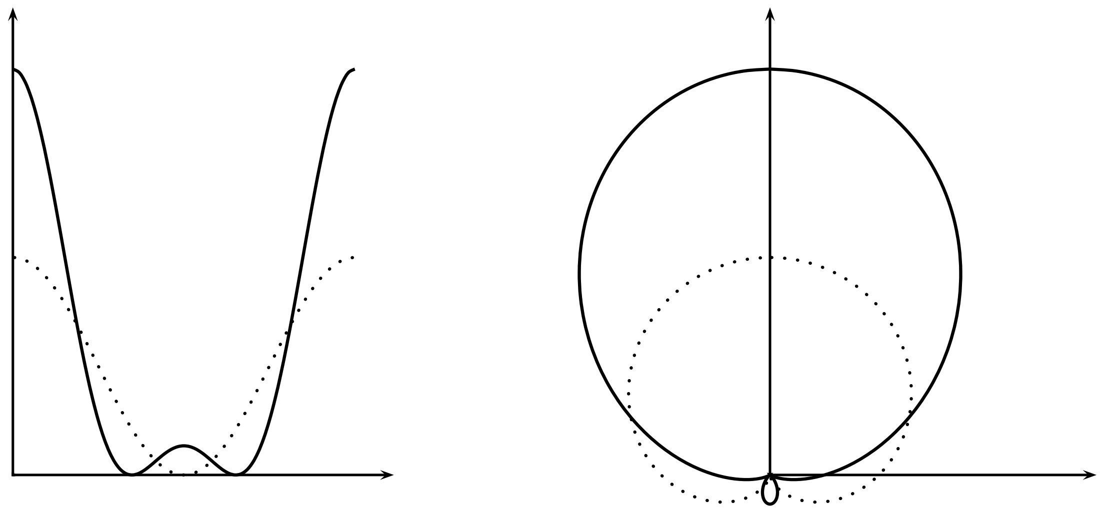

Figure 7

Furse's solution

Furse's solution for the 3D case, unlike his for the 2D case, is very close to the "smooth" solution; the numerical evidence indicates that it is not identical to the "smooth" solution, and in fact is just outside the ellipse.

Mixing first and second order material

At the time of writing, there are no second-order microphones, but there is a practical Ambisonics first-order microphone, the *SoundField* microphone [SoundField]. We may wish to mix ambient sound recorded with this microphone (first-order material) with sounds encoded by computer with both first- and second-order harmonics (second-order material). There is no difficulty if the ideal decoding is used. If an in-phase solution is sought, we have the problem that the X signal (say) of the first-order material may need to be scaled differently from the X signal of the second-order material. At first sight this requires using separate channels for the first-order material (4 channels) an for the second-order material (9 channels), making 13 channels altogether.

Several compromise approaches to in-phase solutions are possible.

The "extends first order" gains

For 2D decoding we can use $p_1 = 0.5$, $p_2 = 0.4268$; for 3D decoding we can use $q_1 = 0.3333$, $q_2 = 0.3633$. Unfortunately these are not very good correcting gains for the second-order material, but they do provide in-phase decoding for both the first- and the second-order material, and any speaker layout can be used.

Fixed correcting gains

Suppose that we know in advance the correcting gains to be used, for example $p_1 = 0.5$ for the first-order material and $p_1 = 0.6667$, $p_2 = 0.1667$ for the second-order material. Let $\sigma_1(W)$, $\sigma_2(W)$ be the the W signals for the first- and second- order material respectively, and similarly for the other harmonics. The encoding we use is

$$\sigma(W) = \sigma_1(W) + \sigma_2(W);$$
$$\sigma(X) = \sigma_1(X) + \frac{0.6667}{0.5}\sigma_2(X), \text{ and similarly for } Y,\ Z.$$

In the decoding we now use $p_1 = 0.5$, $p_2 = 0.1667$, and both the first-order and the second-order material will be decoded correctly.

The disadvantage of this approach is that the material is encoded for planar layouts (in this example) and would not be suitable for 3D decoding. The independence of encoding and decoding which is a feature of Ambisonics has been compromised.

If a single encoding is needed in just 9 channels for both 2D and 3D decoding, probably the best that can be done is

$$\sigma(W) = \sigma_1(W) + \sigma_2(W);$$
$$\sigma(X) = \sigma_1(X) + 1.4142\sigma_2(X), \text{ and similarly for } Y,\ Z.$$

For 2D decoding we use $p_1 = 0.5$, $p_2 = 0.25$, which gives the effect of $p_1 = 0.5$ for the first-order material and $p_1 = 0.7071$, $p_2 = 0.25$ for the second-order material. For 3D decoding we use $q_1 = 0.3333$, $q_2 = 0.3155$, which gives the effect of $q_1 = 0.3333$ for the first-order material and $q_1 = 0.4714$, $q_2 = 0.3155$ for the second-order material. In each case the first-order material receives correct in-phase decoding and the parameters for the second-order decoding lie on the appropriate ellipse. However, in the 3D case, the second-order decoding is "lumpy".

An extra W channel

Malham [Malham 99] has proposed using two W channels, say W_1 for the first-order material and W_2 for the second-order material. This approach allows any decoding to be used; its only disadvantage is that the relative loudness of the first-order and second-order materials may be different in different decodings.

When practical second-order microphones become available and third-order decoding setups are common, no doubt further compromises, or additional channels, will be needed.

References

[Daniel *et al.*] Daniel, Jérôme; Rault, Jean-Bernard; Polack, Jean-Dominique. "Ambisonics encoding of other audio formats for multiple listening conditions" preprint no.4795, 105th Audio Engineering Society Convention, Sept. 1998,. (Corrected version available by contacting the authors at Centre Commun d'Etudes de Télédiffusion et Télécommunications, Cesson Sévigné, France)

[Furse] Furse, R. "First and second order Ambisonic decoding equations", `http://www.muse.demon.co.uk/ref/speakers.html`: retrieved 9th December 1999.

[Hobson] Hobson, E.W. "The Theory of Spherical and Ellipsoidal Harmonics", Cambridge University Press 1931, reprinted by Chelsea Publishing Co., 1955.

[Malham 92] Malham, D.G. "Experience with large area 3D Ambisonics sound systems", Proc. Institute of Acoustics vol 14 part 5 (Proc. Conf. on Reproduced Sound, no. 8), 1992, pp. 212–215.

[Malham 98] Malham, D.G. "Approaches to spatialisation", *Organised Sound* vol 3, no. 2, pp. 167-177.

[Malham 99] Malham, D.G. Proc. International Computer Music Conference, BNeijing, 1999.

[SoundField] "SoundField", `http://www.proaudio.co.uk/sndfield/`: retrieved 9th December 1999.

Appendix A: Calculation for second-order decoding

We need to show that

$$\frac{1}{3}(3z_i^2-1)(3z_0^2-1)+4x_iz_ix_0z_0+4y_iz_iy_0z_0+4x_iy_ix_0y_0+(x_i^2-y_i^2)(x_0^2-y_0^2) = \frac{1}{3}+\cos 2\alpha, \qquad (A1)$$

where α is the angle between $\mathbf{u}_i=(x_i,y_i,z_i)$ and $\mathbf{u}_0=(x_0,y_0,z_0)$.

If we expand out all the terms, the left hand side of (A1) works out to

$$3z_i^2z_0^2-z_i^2-z_0^2+\frac{1}{3}+4(x_iz_ix_0z_0+y_iz_iy_0z_0+x_iy_ix_0y_0)+(x_i^2-y_i^2)(x_0^2-y_0^2). \quad (A2)$$

Now $z_i^2z_0^2=(1-x_i^2-y_i^2)(1-x_0^2-y_0^2)$. Replace the term $3z_i^2z_0^2$ in (A2) by $2z_i^2z_0^2-(1-x_i^2-y_i^2)(1-x_0^2-y_0^2)$ and simplify. We obtain

$$2z_i^2z_0^2+1+\frac{1}{3}-(x_i^2+y_i^2+z_i^2)-(x_0^2+y_0^2+z_0^2)+2x_i^2x_0^2+2y_i^2y_0^2 + 4(x_iz_ix_0z_0+y_iz_iy_0z_0+x_iy_ix_0y_0). \qquad (A3)$$

Since $x_i^2 + y_i^2 + z_i^2 = x_0^2 + y_0^2 + z_0^2 = 1$, (A3) reduces to

$$\begin{aligned} &2(x_i^2 x_0^2 + y_i^2 y_0^2 + z_i^2 z_0^2 + 2x_i y_i x_0 y_0 + 2x_i z_i x_0 z_0 + 2y_i z_i y_0 z_0) - 1 + \frac{1}{3} \\ &\quad = 2(x_i x_0 + y_i y_0 + z_i z_0)^2 - 1 + \frac{1}{3} \\ &\quad = 2\cos^2\alpha - 1 + \frac{1}{3} \\ &\quad = \cos 2\alpha + \frac{1}{3}. \end{aligned}$$