In-phase corrections for Ambisonics

Gordon Monro

School of Mathematics and Statistics
University of Sydney F07
NSW 2006
Australia
G.Monro@maths.usyd.edu.au
http://www-personal.usyd.edu.au/~gmonro/

INTRODUCTION

Ambisonics 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 lay-outs. (For a general discussion of approaches to spatialisation of sound, see (Malham 1998).)

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 (1992) proposed in the context of first-order Ambisonics an in-phase correction which ensures that all the speakers contributing to a single sound image are in phase. This paper extends Malham’s in-phase correction to the second-order case, finding a number of interesting solutions for both two-dimensional and three-dimensional decodings.

The basis for this work is the treatment of (Daniel et al. 1998); 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.

SPHERICAL HARMONICS AND AMBISONIC ENCODING

Ambisonics is based on the notion of a “sound-field”, which in this context means a distribution of intensity on a sphere surrounding the listener. Such a distribution can be approximated by a sum of spherical harmonics (Hobson 1931). If we consider only the horizontal plane, we are dealing with “circular harmonics” (i.e. Fourier series).

In first-order 3D Ambisonics the encoding process produces four signals: one omnidirectional (zeroth order) signal and three first-order signals (x, y and z). For 2D encoding the z signal is omitted.

In second-order 3D Ambisonics, we have one omnidirectional signal, three first-order signals, and five second-order signals (see (Malham 1999) and the references therein). For 2D second-order Ambisonics we have an omnidirectional signal, two first-order signals and two second-order signals. Unfortunately different authors have used different scaling factors for the various spherical harmonics. A comparison of systems is given in (Monro 2000). Happily, the scaling factors are not relevant for a discussion of in-phase corrections.

2D FIRST-ORDER DECODING

Consider a point source of unit intensity. It can be shown (Daniel et al. 1998) that for n speakers evenly spaced around a circle, the “ideal” decoding is as follows. Let \( \alpha_i \) be the angle between the direction of the ith speaker and the direction of the source. The output from the ith speaker should be

\[
\text{out}_i = \frac{1}{n}(1 + 2 \cos \alpha_i). \tag{1}
\]

This “ideal” decoding is what (Daniel et al. 1998) refer to as the basic decoding and (Furse 2000) calls the idealised response.

If we have four speakers at front left, front right, rear left and rear right, the output according to equation (1) is 0.6036 from the two front speakers and −0.1036 from the two rear speakers. Thus the output from the rear speakers is out of phase. To avoid this, put a correcting gain \( p \) on the first-order signals. It can be shown that the output then becomes

\[
\text{out}_i = \frac{1}{n}(1 + 2p \cos \alpha_i). \tag{2}
\]

If we set \( p = 0.5 \), \( \text{out}_i \) is never negative. This is Malham’s in-phase correction (Malham 1992).

2D SECOND-ORDER DECODING

For decoding material with second-order signals, the “ideal” decoding is

\[
\text{out}_i = \frac{1}{n}(1 + 2 \cos \alpha_i + 2 \cos 2\alpha_i). 
\]
We introduce two independently adjustable correcting gains $p_1$ and $p_2$: $p_1$ for the first-order signals and $p_2$ for the second-order signals. It can be shown that the output becomes

$$\text{out}_i = \frac{1}{n} (1 + 2p_1 \cos \alpha_i + 2p_2 \cos 2\alpha_i).$$  \hspace{1cm} (3)$$

We drop the factor of $\frac{1}{n}$ and write

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

For in-phase decoding we need to find values of $p_1$ and $p_2$ such that $f(\alpha) \geq 0$ for all $\alpha$. It turns out (Monro 2000) that the solutions are values of $p_1$, $p_2$ corresponding to points in the shaded region in Figure 1 below. Part of the boundary is the ellipse

$$2p_1^2 + 16(p_2 - \frac{1}{4})^2 = 1.$$  \hspace{1cm} (4)$$

The rest of the boundary is made up of the coordinate axes and the line $p_1 = 0.5 + p_2$. There are two possible shapes for the graph of $f(\alpha)$ against $\alpha$: below the line $p_1 = 4p_2$ there is only a single minimum (as in Figure 2); above this line there are two minima (as in Figure 3).

Figure 1

The two lines in the figure intersect at $\left(\frac{3}{4}, \frac{1}{2}\right)$, which is on the ellipse.

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: $p_1 = 0.5000, p_2 = 0.4208$.
- 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: $p_1 = 0.7071, p_2 = 0.2500$.
- 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 $\alpha$ 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. 1998) that the quantity

$$r_E = \frac{2(p_1 + p_1p_2)}{1 + 2(p_1^2 + p_2^2)}$$

compares the magnitude of the energy vector (also known as the intensity vector) of the reconstruction to that of the 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 (4) that maximises $r_E$. This was 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,
A straightforward calculation shows that in our situation the front-back ratio is

\[
\frac{1 + 2p_1 + 2p_2}{1 - 2p_2} \quad \text{if } p_2 \leq p_1/2
\]

\[
\frac{1 + 2p_1 + 2p_2}{1 - 2p_1 - 2p_2} \quad \text{if } p_2 \geq p_1/2
\]

The point on the ellipse (4) 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.

**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 can be calculated, for in-phase solutions, to be

\[
\frac{\pi + 4p_1}{\pi - 4p_1}.
\]

The maximum value of this quantity is obtained when \( p_1 \) has its maximum value of \( 1/\sqrt{2} \), and with this value of \( p_1 \), the integrated front-back ratio is 19.063, or approximately 25.6 dB. This solution, with \( p_2 = 1/\sqrt{2} \) and \( p_2 = 0.25 \), 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.

**Figure 4**

It may seem surprising that \( p_2 \) 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 \( p_2 \) means that \( p_1 \) 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; one can argue that the integrated front-back ratio is the better thing to maximise. Altogether this appears to be the best solution.

**Furse’s solution**

Furse (Furse 2000) 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. 1998). 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 (4), and to a rather lumpy source directivity response.

**3D SECOND-ORDER DECODING**

The results for 3D decoding are similar to those for 2D decoding, though there are differences in detail. I will use the letter \( q \) for correcting gains in the 3D case.

The counterpart to equation (2) is

\[ \text{out}_i = \frac{1}{n} (1 + 3q \cos \alpha_i), \]

so the first-order in-phase correction is \( q = 1/3 \) (as opposed to \( p = 1/2 \) in the 2D case). For a sufficiently symmetric speaker array (of which the main example is 12 speakers at the face centres of a dodecahedron), the 3D version of equation (3) is

\[ \text{out}_i = \frac{1}{n} \left( 1 + 3q_1 \cos \alpha_i + \frac{15q_2}{4} \left( \frac{1}{3} + \cos 2\alpha_i \right) \right). \]

As we did before with \( f \) in the 2D case, define a function \( g \) by

\[ g(\alpha) = 1 + 3q_1 \cos \alpha + \frac{15q_2}{4} \left( \frac{1}{3} + \cos 2\alpha \right). \]

We need to find values of \( q_1 \) and \( q_2 \) such that \( g(\alpha) \geq 0 \) for all \( \alpha \). The result is the shaded area in Figure 5. This time the ellipse has the equation

\[ 3q_1^2 + 25(q_2 - 1/5)^2 = 1. \]

The two lines in the figure intersect at \( (\frac{1}{2}, \frac{1}{10}) \), which is on the ellipse.

**Figure 5**

Four noteworthy solutions are
The “extends first order” correcting gains respectively 9 and 10 channels are as follows.

We will only discuss the “maximum integrated front-back ratio” solution here. For the 3D case the integrated front-back ratio for parameters \( q_1, q_2 \) works out to be

\[
\frac{2 + 3q_1}{2 - 3q_1}
\]

without explicit dependence on \( q_2 \). As in the 2D case, the maximum value of this ratio occurs at the end of the semimajor axis of the ellipse. Here \( q_1 = 1/\sqrt{3} \), and the ratio is then 13.928. This solution is graphed in Figure 6, and appears to be the best 2nd-order in-phase correction for 3D Ambisonics.

**Figure 6**

**MIXING 1ST AND 2ND 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* 1998). 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) and for the second-order material (9 channels), making 13 channels altogether. Two compromise solutions which use respectively 9 and 10 channels are as follows.

The “extends first order” correcting 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 rather lumpy 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. This solution uses the minimum of 9 channels.

An improved solution with 9 channels can be obtained by multiplying the \( x, y \) and \( z \) signals from the second-order material by 1.4142 before mixing them with the first-order material. Then for 2D decoding use \( p_1 = 0.5, p_2 = 0.25 \); for 3D decoding use \( q_1 = 0.3333, q_2 = 0.3155 \). This gives in-phase decoding in all cases and good 2D decoding for both the first- and the second-order material. The 3D decoding of the 2nd-order material is still somewhat “lumpy”.

**An extra W channel**

Malham (1999) 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. This solution requires 10 channels.

**Acknowledgements**

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, Dave Malham for numerous enlightening conversations, and the University of Sydney for support.

“Ambisonics” is a registered trademark of Nimbus Communications International.

**References**


