WHITE PAPER

AMBISONICS

Preface

This White Paper was created to rectify the lack of a brief yet rigorous exposition of the central ideas and praxis of first-order Ambisonics. It was originally created for internal use by Pspatial Audio and has been made available in the hope that others might find this straightforward, yet thorough, explanation helpful.

Richard Brice . October 2014

Introduction

The British *Ambisonics* recording and reproduction system grew out of the heady success of the record industry in the nineteen-seventies - rich with technical developments like quadraphonics. Brainchild of various British academics including Peter Fellgett and the mathematician Michael Gerzon, Ambisonics builds upon Blumlein's original work on stereophony (Blumlein 1933) to propose a complete system for the acquisition, synthesis and reproduction of enveloping sound fields from a limited number of loudspeakers.

Essential to Ambisonics is the concept that the transmitted (and/or recorded) signals in Ambisonics are *not* the loudspeaker signals - as is the case with quadraphonic systems or 5.1 surround sound. Instead Ambisonics encodes directional sound within a set of signals: three for horizontal only Ambisonics and four for Ambisonics which includes the height sensation as well as the "surround sound". The power of this concept is that these loudspeaker-independent signals may be manipulated to drive a variety of loudspeaker arrangements of four loudspeakers or more. (Four is the minimum number.) In a nutshell: *Ambisonics encodes direction as a property of the recorded sound*.

Ambisonics *virtual microphones*

Let us imagine an event being recorded in Ambisonics is a physical space (say, a concert hall) and being replayed in another space (say a domestic living room.) Central to being able to derive an appropriate set of loudspeaker signals from the Ambisonics encoded signals is the concept of the *virtual microphone*. Because of the way the signals are encoded in Ambisonics, it is possible to decode the signals such that we effectively "steer" an imaginary microphone from the position of the Ambisonics microphone array in the concert hall so that any particular loudspeaker in the listening room has a directional microphone pointing in the appropriate direction to energise it.

For example, imagine a setup in which 4 loudspeakers are positioned in a diamond: front, left-side, back and right-side. In this case, it is possible, via straightforward, linear signal manipulations of the Ambisonics signals to derive the outputs of four microphones pointing: front, left side, back and right side. But for the listener next door, who has his speakers arranged in a square: left front, left back, right back and right front, it is equally possible to derive virtual microphones pointing appropriately left front, left back, right back and right front. Moreover, the directional pattern of the microphones may be manipulated too from cardioid through to a much narrower hyper-cardioid. For the enthusiastic listener who has the budget, for example, for eight amplifiers loudspeakers in an octagonal arrangement, eight virtual microphone signals may be derived from the Ambisonics signals; each arranged to energise each loudspeaker with an appropriate virtual microphone signal and with the appropriate directional pattern.

Because of this virtual microphone concept, Ambisonics supports a number of practical arrangements of microphones as we shall see. The most elegant being the tetrahedral array as illustrated in practical form in Figure 1 (left).

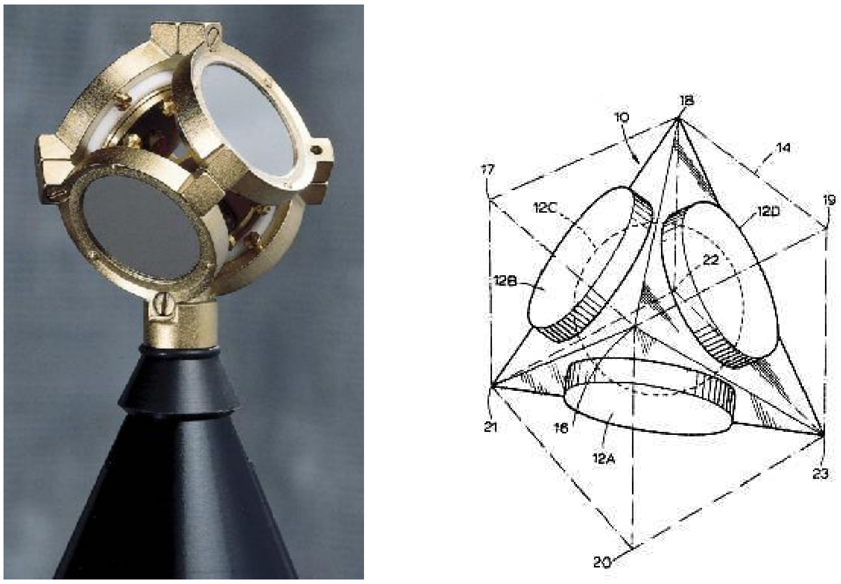


Figure 1 - The practical tetrahedral microphone (left) and design from patent document (right) which more clearly reveals its structure

In this type of microphone, four hyper-cardioid microphones capsules are arranged on the four faces of a regular tetrahedron respectively 35.3° below and above the four corner directions as illustrated in Figure 1 (right). This tetrahedral microphone array is termed the *Soundfield* microphone and was invented by Peter Craven and Michael Gerzon (Craven and Gerzon 1977). The four signals emanating from these microphones are known as: left-back down; right-back up; left-front up; right-front down. This is known as Ambisonics *A-format*.

The signal formats at various stages in the Ambisonics signal chain are labelled with a series of letters, starting at A-format for the microphone signals, through to the set of signals known as D-format which are the signals to be fed to the physical loudspeakers. The different formats will be explained as we work our way through the Ambisonics system chain[1].

Ambisonics signals and processing

In the following, we will confine ourselves to considering horizontal-only Ambisonics; that's to say, without the height signal. This simplifies the explanation without sacrificing much, because it's easy, once the simpler horizontal system is understood, to grasp how a third dimension may be added to the theory.

[1] A, B, C and D are the important signal formats and are Gerzon’s original nomenclature. *Some references refer to other formats. For example, loudspeaker signals-feed decodes of B-format material suitable for 5.1 arrays recorded on a digital, multichannel medium (like SA-CD or DVD-A) have been termed G-format but this isn’t completely standard.*

A-format

As stated above, in the beginning, there are A-format signals. These are the signals which emanate from the microphones. According to Gerzon (1975), when we are considering horizontal only Ambisonics, the signals should ideally emanate from a "flattened out" tetrahedral array in which,

> *.... the outputs of four hypercardioids each having nulls 120 degrees off-axis pointing in the four corner directions.* [ie. 90 degrees apart and labelled: Lb; Lf; Rf; and Rb.]

Now we already know that a microphone with a cardioid directional response may be fabricated by adding the contributions from a velocity microphone to the output of an omnidirectional microphone as illustrated in Figure 2.

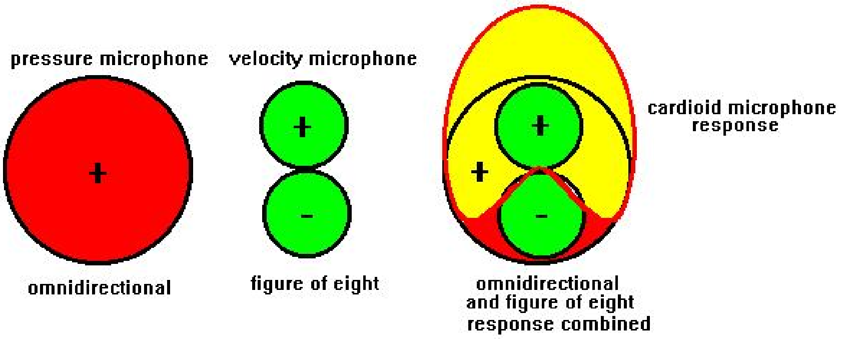

pressure microphone
velocity microphone
+
+
-
omnidirectional
figure of eight
+
+
-
cardioid microphone response
omnidirectional and figure of eight response combined

Figure 2 – A cardioid polar pattern is derived from the combination of an omnidirectional pattern with a figure-of eight pattern

A pure cardioid response is described by the following equation,

$$(0.5 + 0.5 \cos A)$$

where ***A*** is the angle of incidence. In words: the equal mixture of a cosine (figure of eight) response and an omnidirectional response. But the cardioid has a significant response at 120° off axis. In fact it's 25% (-12dB) of its on axis response. See figure below.

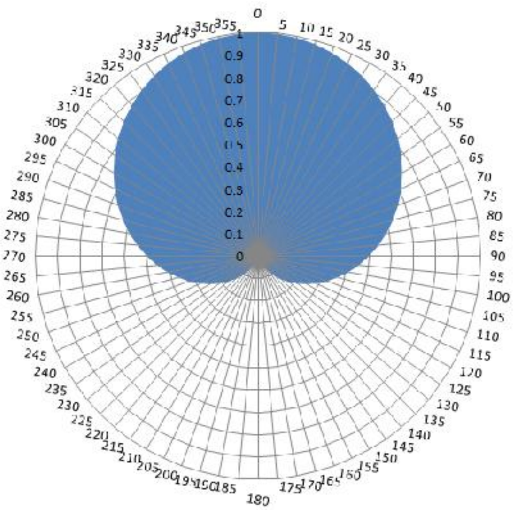


Figure 3 – Cardioid polar response

For a hyper-cardioid to have a null at 120 degrees off axis, it has the equation,

$$(\mathbf{0.25 + 0.5 \cos} \, \boldsymbol{A})$$

In words, the contribution of the omnidirectional microphone is reduced by one-half (-6dB). Here is the polar response of a hyper-cardioid with a null at 120 degrees off-axis[2].

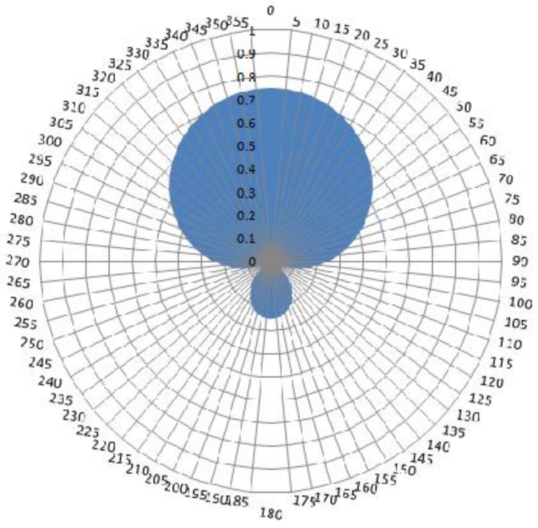


Figure 4 – Hyper-cardioid response with null at 120°

[2] It's worthwhile noting that a hyper-cardioid with a null at 120 degrees off-axis has a response which is 25% (-12dB) at 90 degrees off-axis which is where the adjacent microphone has its maximum response. That's to say that the maximum adjacent channel separation is 12dB in 4-channel Ambisonics captured with a microphone.

B- format

In fact, A-format signals are not used in signal processing, transmission or recording within the Ambisonics system. Instead the A-format signals are matrixed to B-format in the microphone base-station according to the following linear (and aperiodic) equations:

X = ½(-Lb+Lf+Rf-Rb)
W = ½(Lb+Lf+Rf+Rb)
Y = ½(Lb+Lf-Rf-Rb)
Z = ½(-Lb+Lf-Rf+Rb)

Practically this is achieved the circuit given below, which in Ambisonics parlance is referred to as the *AB module*. Note that all the resistor values are identical.

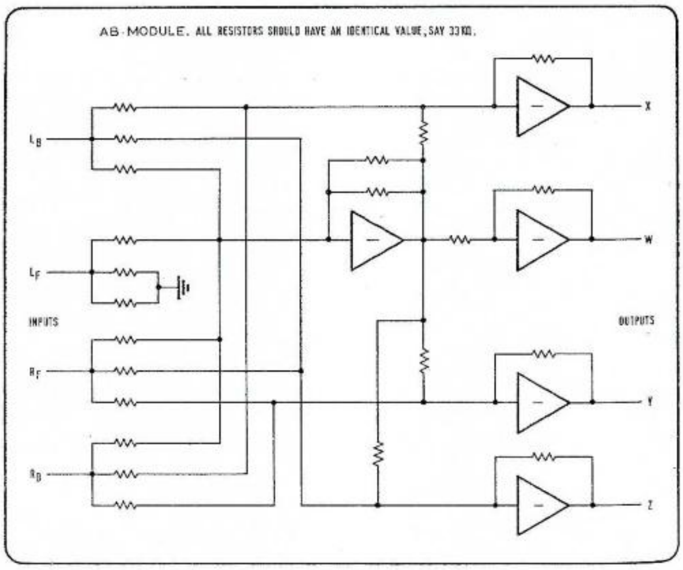

AB MODULE. ALL RESISTORS SHOULD HAVE AN IDENTICAL VALUE, SAY 33kΩ.
LB
LF
INPUTS
RF
RB
X
W
OUTPUTS
Y
Z

Figure 5 – The Ambisonics AB module

This electrical manipulation produces three new signals as if they were derived from two figure-of-eight microphones rigged perpendicularly to one another and a third signal derived from an omnidirectional microphone rigged coincidently. If we take the phase of the omnidirectional microphone (the ***W*** signal) as the reference:

- The ***X*** signal is equivalent to the output of front-back rigged cosine microphone (front being the positive-phase lobe);

- The ***Y*** signal is the equivalent of a left-right rigged cosine microphone (with left being its positive-phase lobe).

If the third dimension of height is to be recorded as well, a third orthogonally rigged cosine microphone is derived with respect to the other two so that it points up-down, with up being its positively phased lobe. Where present, this is known as the ***Z*** signal. This hypothetical microphone and its signals are illustrated in Figure 6.

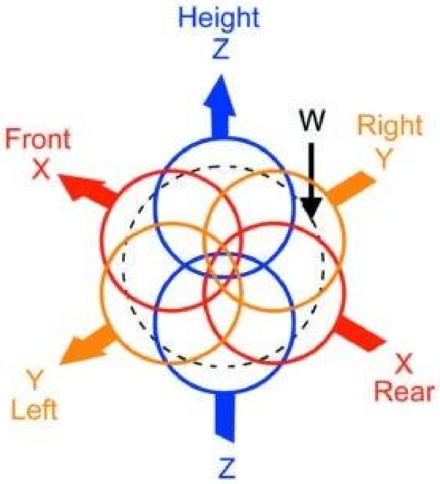


Figure 6 – The B-format signals

Returning to horizontal-only Ambisonics, note that a tetrahedral-array Soundfield microphone is not essential to make Ambisonics recordings. Some Ambisonics recording engineers (notably from the record company Nimbus) actually use three microphones, one omni' and two figure of eights, to derive B-format signals directly[3]. Their microphone rig is illustrated in Figure 7. And other arrangements are also possible.

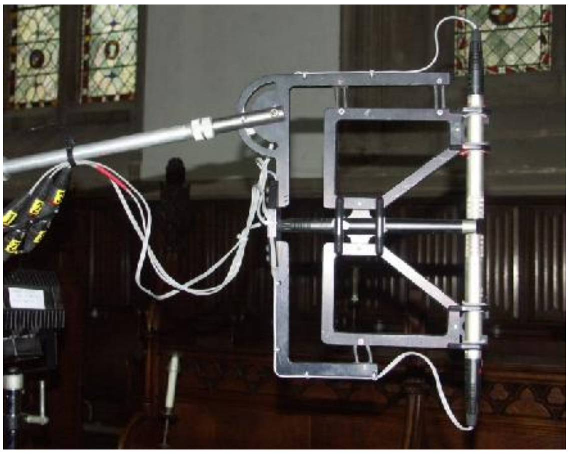


Figure 7 – The Nimbus (B-format) array

[3] How do you set the levels of such an array? The rule is: if the three (or two) velocity microphones and the omni (pressure) microphone are rigged in a reverberant environment such that the sound energy is arriving equally from all directions, the output from each microphone should amplified and arranged to read equally on a VU meter. If this rubric is employed, the 0.707 factor in the W signal (see text) is automatically taken into account.

If B-format signals are derived electronically (as with a pan-pot), then the circuit below is proposed.

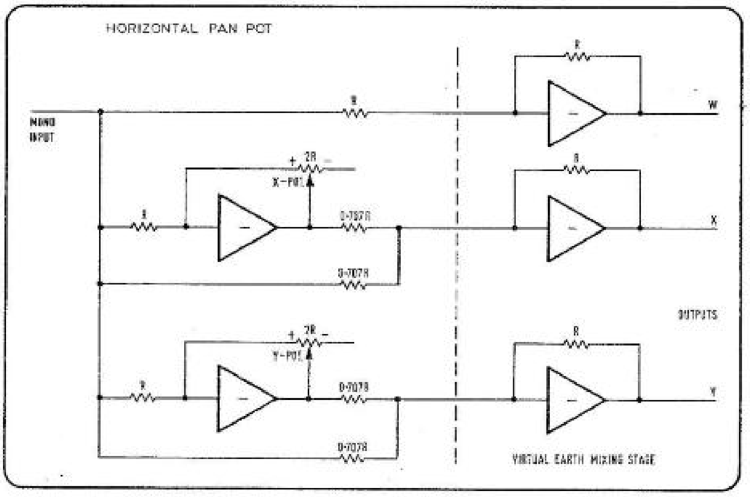


Figure 8 – Ambisonics horizontal pan-pot[4]

Gerzon (*ibid*) writes,

> *A second method of producing B-format signals it to use a panpot....... [Fig. 8] shows a circuit of a panpot (feeding a virtual earth mixing stage) that uses a joystick control to meet accurately the encoding specification for B-format for horizontal sounds (so that the* ***Z*** *signal is zero). The '****X****-pot and* ***Y****-pot' are the potentiometers that respond to the ..... [front-back] .....and 'left-right' motions respectively of the joystick.*

This synthetic approach is consistent with the classic exposition of Ambisonics that the position of a sound (***S***) within a three dimensional sound-field is encoded in the four signals which make up the B-format thus;

> ***X = S* . cos *A* . cos *B* (front-back)**
> ***Y = S* . sin *A* . cos *B* (left-right)**
> ***Z = S* . sin *B* (up-down)**
> ***W = S* . 0.707 (pressure signal)**

where ***A*** is the anti-clockwise angle from centre front and ***B*** is the elevation. These expressions are usually given as the theoretical "pan" expressions for (first-order) Ambisonics.

Note here that the ***W*** signal is 0.707 the size of the maximum output of either the ***X*** or ***Y*** signals from this electronic stage. Reading Ambisonics literature, it's easy to get confused about this 0.707 multiplier in the encoding of the ***W*** signal in B-format. Note that there is *no 0.707 factor applied to the* ***W*** *signal if the B-format signals are derived from the A-format signals from the microphones in*

[4] Gerzon does make the important point regarding this circuit that the two pots are sine/cosine, so they must not be both at extreme travel at the same time. This is, in any case, covered in the equations for Ambsionics panning

an array of hyper-cardioids. This is because this multiplier is already taken into account in the physical response of the microphones themselves.

To understand this better, consider the Soundfield microphone as genuinely consisting of three figure of eight microphones and an omnidirectional microphone as illustrated in Figure 6. As we have seen, the cosine microphones are rigged front-back and side to side and are not disposed diagonally, as in stereo crossed-eights. But we could add the front-back (***X***) signal to the left-right (***Y***) signal and derive a diagonal figure of eight as shown in Figure 9.

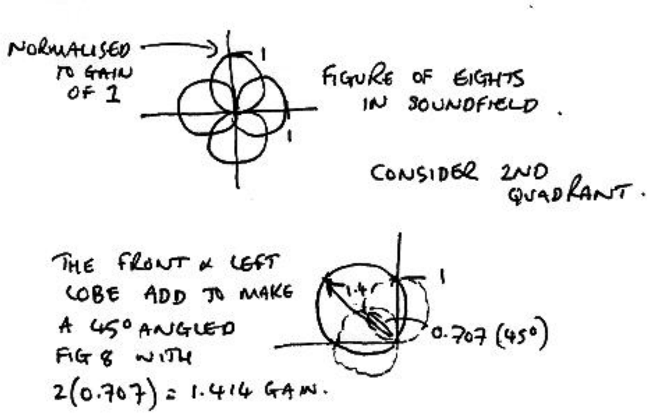

NORMALISED TO GAIN OF 1
1
1
FIGURE OF EIGHTS IN SOUNDFIELD.
CONSIDER 2ND QUADRANT.
THE FRONT & LEFT LOBE ADD TO MAKE A 45° ANGLED FIG 8 WITH 2(0.707) = 1.414 GAIN.
1.41
1
0.707 (45°)

Figure 9 – see text

And we could add the ***X*** signal to a reverse-phased ***Y*** signal, to create a diagonal, right facing cosine response in a similar way. In each case (as illustrated in the figure), the output of the microphone would be 1.414 times the output of the individual microphone. (Because, at 45°, the output of the eights is 0.707 times the on-axis response and the two are added together: thus 2 × 0.707 = 1.414.)

Now, in order to get back to our hyper-cardioid response (with a null at 120 degrees off-axis), we need to add half as much of the omni's signal to the figure-of-eight signal. And there we have the reason for the 0.707 factor in the W signal: because 0.707 is half of 1.414.

Ambisonics signals in B-format are the signals which are transmitted or recorded. The most common form in which you encounter B-format signals is in an Ambisonics audio file with a **.amb** extension. These are readable by many DAWs, including the freeware Audacity. When an .amb file opens in Audacity, the audio tracks open in the interface in the following order: ***W, X, Y*** (and where present), ***Z***. It is therefore a very simple matter of setting pans and gains prior to rendering to derive stereo or D-format (loudspeaker) signals

C-format (or Ambisonics UHJ)

An alternative to B-format transmission or recording is *C-format* or *Ambisonics UHJ*. This is stereo compatible format for horizontal-only Ambisonics. The format is spatially lossy, but has the obvious advantage that is may be broadcast and stored and replayed just like any other stereo signal. And, if the appropriate decoder is used to derive horizontal Ambisonics, the results are very good.

In C-format, the three channels of B-format, ***W, X*** and ***Y***, are encoded into a two in a similar way in which the four signals of quadraphonic are encoded into a stereo compatible signal, by means of a phase amplitude matrix. (Although Ambisonics enthusiasts would be quick to point out that the properties of the phase-amplitude matrix are optimised compared with the various quadraphonic arrangements!)

Rather wonderfully, despite having been matrixed to two channels, with appropriate decoding, C-format signals maintain their independence from the loudspeaker signals and different loudspeaker signals may still be derived for different loudspeaker arrangements. By far the greatest catalogue of Ambisonics recordings (in both number and in artistic merit) exist in the UHJ format[5].

UHJ is derived from ***W, X*** and ***Y*** in the following way. Firstly, sum and difference signals are created such that,

$$S = 0.9396926*W + 0.1855740*X$$
$$D = j(-0.3420201*W + 0.5098604*X) + 0.6554516*Y$$

Then the left-total (***Lt***) and right-total (***Rt***) signals are derived such that,

$$Lt = (S + D)/2.0$$
$$Rt = (S - D)/2.0$$

where **j** is a +90 degree phase shift (phase advance)[6].

UHJ decoding is the inverse of the encode so that,

$$S = (Lt + Rt)$$
$$D = (Lt - Rt)$$

and,

$$W = 0.982*S + j*0.164*D$$
$$X = 0.419*S - j*0.828*D$$
$$Y = 0.763*D + j*0.385*S$$

where **j** is a +90 degree phase shift (phase advance)[7].

[5] Nimbus have made this statement regarding UHJ, "*With a couple of exceptions, all of our recordings are UHJ encoded Ambisonics. We have been using essentially the same recording technique for almost thirty years now....... our own B-format microphone [combines] two Shoeps figure-of-eights and a B&K omni*."

[6] There is however a reference which states that C-format is derived from the following formula:

$$S = 0.9396926*W' + 0.1855740*X$$
$$D = j(-0.3420201*W' + 0.5098604*X) + 0.6554516*Y$$

where ***W' = W*** × 1.414; in other words already *energy field optimised*. I don't believe this to be true, but - as with so much Ambisonics research - one is left feeling confused and frustrated!

[7] Intuitively we can see how this works: ***W*** is essentially mostly (***L + R***), which is intuitively logical. As is the fact that ***Y*** (leftwards) is mostly ***k(L-R)***. It's also pretty obvious that Stereo UHJ is stereo-compatible because the

Getting to D-format - decoding Ambisonics

As already stated, one of the powers of Ambisonics is that, once the spatial information has been encoded into the B-format channels (***W, X, Y*** and possibly ***Z***), there is no particular prescribed loudspeaker layout for reproduction. However, this moves the system complexity to the decoder. The encode side of Ambisonics is straightforward and reasonably well documented (with the possible exception of C-format). It is rather on the decode side where the eventual D-format loudspeaker signals are derived from B-format or C-format signals where the complications arise.

And here, it is unfortunate that no, definitive - let alone straightforward - exposition of Ambisonics theory exists. Gerzon was too cerebral for such a dreary task. And his background as a mathematician ensured that his writings on Ambisonics are often not in the easiest form to digest and apply. This, along with a certain amount of deliberate obscuration during the period that the inventors and investors thought they could make a fortune, has resulted in an Ambisonics literature which is not only often confusing, but frequently itself confused.

Moreover, Gerzon's highly theoretical approach and style has ensured that Ambisonics as a subject attracts more theorists than experimenters and engineers. So that, whilst there is no shortage of papers discussing the solutions simultaneous, non-linear, differential equations[8] , there is a paucity of: definitive formulae, data based on properly conducted experimental listening tests; few practical reference designs or test material; and no commercial hardware whatsoever! And people wonder why Ambisonics failed.....

The problems in implementing Ambisonics exist even with the simplest decoders (specified in the earliest days, and substantially quadraphonic in form, with no height). These are called *Regular Polygon decoders*, the most straightforward being a square of four loudspeakers arranged on the perimeter of a circle so as to be equidistant from a centrally seated listener; just like quadraphonics.

Some Ambisonics references (the Wikipedia article for *Ambisonic decoding* for example) says the decoded signals, in these straightforward arrangements should follow the following, simple rule, such that the output of speaker ***Pn*** is,

$$Pn = a\,(W + X \cos \mathrm{Sn} + Y \sin Sn)$$

where ***Sn*** is the direction of the speaker under consideration. The factor ***a*** is just some arbitrary scaling constant to avoid overloads (in fact not present in the Wikipedia reference). This suggests that the ***W*** signal should be 3dB greater than either the ***X*** or ***Y*** signal if we assume the speaker was set to 45° from the listener (as was the case in quadraphony). Another Wikipedia article (*Ambisonics*) says the signals of a four channel, "naive" encoder should be the following,

signals are set up this way. The only counterintuitive part is the ***X*** signal (the front-back signal) which is encoded as phase information.

[8] To align vectors of Makita theory with Energy theory in non-regular speaker layouts.

$$LF = 2.828(2W + X + Y)\text{... etc.}$$

In other words, that the ***W*** signal should be *twice* the ***X*** and ***Y*** signal amplitudes.

Gerzon (1985), on the other hand, gives the following equations for a regular array of n loudspeakers.

$$Pn = 1/\sqrt{(n)}\ .\ \{W + (1.414.\ \cos Sn) + (1.414\ .\ \sin Sn)\ \} \quad \text{.......... (Eq. A)}$$

In which case, for speakers set at the corners of a square (Sn = 45°, 135°, 225°, 315°), the following equations would be used:

$Lf = aW' + bX' + bY'$
$Rf = aW' + bX' - bY'$
$Lb = aW' - bX' + bY'$
$Rb = aW' - bX' - bY'$

such that $a = b$, and the weights of ***W*** and ***X*** and ***Y*** are all equal.

Even more confusingly, none of the previous schemes appear to have been the case in practice when Ambisonics was in its heyday. If we reference the hardware decoder designed by Gerzon (1977), neither of these solutions are satisfied. The decoder (in block-schematic form) is illustrated in Figure 10.

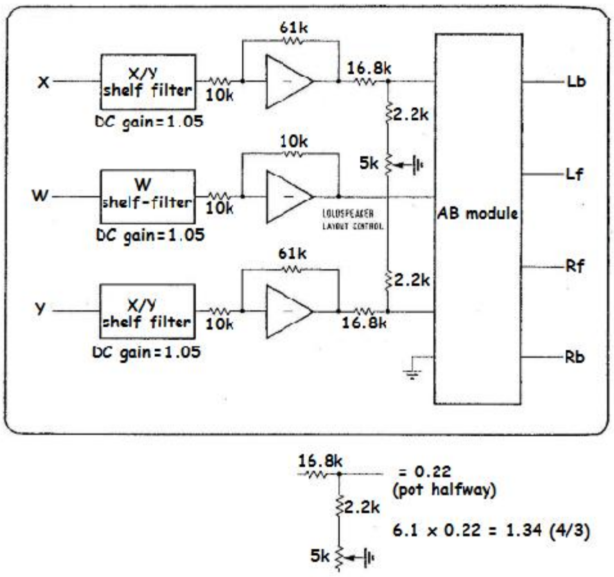


Figure 10 – Block schematic of the Integrex decoder (designed by Gerzon)

Here we can see that ***W*** (following the shelf-filters, which we'll ignore for the moment) is multiplied by one, and ***X*** and ***Y*** are multiplied by 6.1, only to be subsequently divided by 0.22 (assuming the layout control is central for a square layout). That makes an overall gain (ignoring the shelf filters) of

6.1 × 0.22 = 1.34.

Therefore, the ***X*** and ***Y*** signals are larger than the ***W*** signal: not smaller...... and not equal either.

So, in summary, we have references which maintain that in deriving the loudspeaker signals in Ambisonics, the following situations should apply:

- ***W*** should be twice (6dB) the amplitude of the ***X*** and ***Y*** signals.
- ***W*** should be 1.414 times (3dB) above the value of the ***X*** and ***Y*** signals.
- ***W*** should be the same amplitude of the ***X*** and ***Y*** signals.
- ***W*** should be 75% (-2.5dB) below the amplitude of the ***X*** and ***Y*** signals.

Which is correct? After all, we are talking about disagreements on an overall amplitude ratio of 2.7 times (8.5dB) here! Here is my interpretation.....

If we assume that, despite all the manipulations which might happen in between, the desired end result is that the original signals (the A-format microphone signals) feed the four loudspeakers of a square layout (which is a reasonable assumption) then clearly Gerzon's equation in the 1985 article makes the most sense. Because all the scaling is unity and the left front loudspeaker (at 45° to the listener) is *après tout* fed with the signal from the microphone pointing 45° left. One could say that the transfer functions of the ***W, X*** and ***Y*** channel are unity. In fact, things are not that simple as we shall see, however, let's proceed on that basis for the time being and imagine a decoder based on these equations. But before we do, we shall have to make a little detour in psychology and consider the mechanisms of spatial hearing.

Velocity and energy

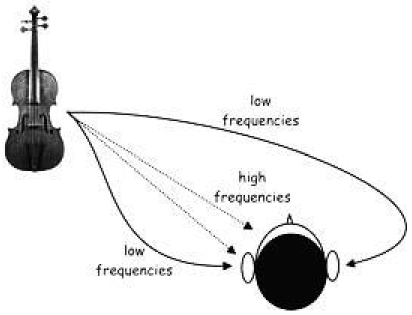


Figure 11 - Directional hearing cues

Consider the situation shown above, in which an experimental subject is presented with a source of sound located at some distance from the side of the head. The two most important cues the brain uses to determine the direction of a sound are due to the physical nature of sound and its propagation through the atmosphere and around solid objects. We can make two reliable observations:

- At high-frequencies, the relative loudness of a sound at the two ears is different. Because the head acts as a baffle, the nearer ear receives a louder signal compared with the remote ear.

- At low frequencies, the baffling effect of the head is inoperative, but a delay exists between the sound reaching the near ear and the further ear.

It may be demonstrated that both effects aid the nervous system in its judgement as to the location of a sound source. At high frequencies, the head casts an effective acoustic "shadow" which acts like a low-pass filter and attenuates high frequencies arriving at the far ear, thus enabling the nervous system to make use of *interaural intensity differences* to determine direction. At low frequencies, sound diffracts and bends around the head to reach the far ear virtually unimpeded. So, in the absence of intensity-type directional cues, the nervous system compares the relative delay of the signals at each ear. This effect is termed interaural delay difference. Because when we speak of low-frequencies here, we refer to sound waves with wavelengths larger than the head, the low-frequency delay manifests itself as an *interaural phase-difference* between the signals arriving at either ear. The idea that sound localisation is based upon interaural phase-differences at low frequencies and interaural intensity differences at high frequencies has been called Duplex theory and it originates with Lord Rayleigh at the turn of the twentieth century. We shall see that these two mechanisms complicate Ambisonics decoding.

An experiment

Let us imagine an experimental Ambisonics listener being requested to turn her head in the direction of reproduced sound events so that each event she is asked to face the direction of the reproduced sound. The experiment would be arranged so that some of the sound events contain only high frequencies (HF), and some contain only low frequencies (LF). We can conjecture that the subject will be using a "nulling" technique to do this to minimise phase differences at LF and amplitude differences at HF[9].

A reasonable measure of the degree of success of the Ambisonics reproduction would then be to compare the direction to which the listener turns with the direction of the original sound in relation to the microphone array in the original venue; or to the theoretical pan position if the sounds were artificially positioned.

Remember here that we are assuming that the decoder is using equal proportions of ***W, X*** and ***Y*** to reproduce the D-format speaker signals using the speaker equation (A).

If this experiment is performed, it is found that the low frequency sound events do indeed appear to come from very close to their original positions. The listener does turn to the correct direction. However, the same cannot be said for the position of the high frequency events in which the listener

[9] The two theories of sound localisation based on these mechanisms are known as *Makita theory* and the *energy vector theory*. The Makita localisation of a reproduced sound is that direction in which the head has to face in order that the interaural phase difference is zero. And the Energy vector locali-sation is the direction the head has to face in order that there be no interaural amplitude difference at high frequencies.

often does not turn to the correct direction. Furthermore, if a reproduced LF sound and a subsequent HF sound are contrived to come from identical encoded direction, the listener will not turn to the same direction for the two sounds except under certain circumstances. We can say that the phase null at LF does not coincide with the energy null at HF for the listener over a substantial proportion of the reproduced azimuth. Experiment and theory agree well at LF - where the wavelengths of the low-frequency events are substantially greater than the distance between the ears. But above that threshold (which is above about 500Hz[10]), theory and experience diverge. The reason for this is relatively simple: Ambisonics theory, for all its mathematics, is really based on an *entirely low-frequency model of human hearing:* just as was Blumlien's original work on stereophony from which it derives. (For a mathematical description of Blumlein's end-to-end system see Brice 2012.) The theory of Ambisonics virtual microphones is conceptualised for low-frequency sounds with modifications to the LF theory to allow for better HF imaging.

Blumlein took the view that what was really required was the capturing of all the sound information at a single point and the recreation of this local sound field at the final destination - the point where the listener is sitting. He demonstrated that for this to happen, it required that the signals collected by the microphones and emitted by the loudspeakers would be of a different form from those we might expect at the listener's ears; because we have to allow for the effects of crosstalk. He developed a complete theory to do this for sounds below the 500Hz threshold. That theory is, in effect, identical in Ambisonics: one might say that it's a subset of the more Gerzon's generalised theory.

Blumlein was forced to an empirical approach to high-frequency localisation in stereo - for which his team invented the *Shuffler* (Brice 2012) and Ambisonics is similarly forced to various empirical "dodges" and compromises to cope with HF reproduction. Nearly all the complications in Ambisonics derive from this central fact.

In order fundamentally to improve Ambisonics reproduction at HF, more loudspeakers are required and ultimately, more microphone channels are required too. These *higher-order* Ambisonics systems as they are known (with more than 4 channels) move Ambisonics into the area of wave-front reconstruction systems but, just as with wave-front reconstruction, theoretical considerations prove that a great many channels are required to give correct performance at HF. (It has been calculated that 32 channels and a thousand loudspeakers would be required for a full-bandwidth Ambisonics solution!)

Practical approaches to HF reproduction

Nevertheless, just as with Blumlein's Shuffler (or my own FRANCINSTIEN), signal manipulations and compromises may be employed greatly to improve HF reproduction compared with the simplest decoder described above which, because it recreates the velocity vector field is known as a *velocity optimised decoder*.

[10] 500Hz is really a nominal figure and represents the middle of the "Twilight Zone" (from about 350Hz upwards to about 2kHz), in which the hearing system is neither very effective at determining direction by phase or by amplitude.

The most basic technique is to increase the proportional contribution of the ***W*** in the derivation of the speaker signals. This has the effect (as we saw in the encoding section) of transforming the virtual microphones from hyper-cardioid towards cardioid. It can be demonstrated that the best solution for HF reproduction is satisfied when the ***W*** signal is 3dB higher than the ***Y*** and ***X*** signals. If this is the case, virtual microphones are nearly cardioid (as shown in Figure 12) and the reproduced HF energy field is the best approximation.

If no shelf filters are employed (see below), this recipe is considered the best as a general purpose decoder. (Presumably, this explains why these weightings are the ones referred to in the Wikipedia "Ambisonics decoding" article: because this is the simplest general-purpose decoder.) This technique is referred to as an *energy optimised decoder.*

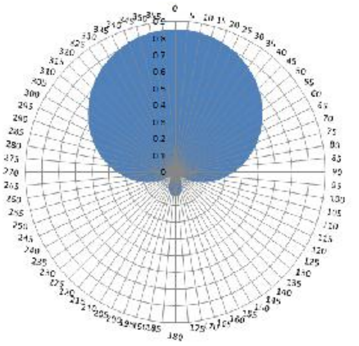


Figure 12 - The polar response of the virtual microphones when W is 3dB greater than the velocity signals

Of course, in employing this simple technique, the LF image is compromised, since it is no longer ideal. What is therefore ideally required is the ability to decode using a smaller proportion of ***W*** at LF than at HF. And that is where the shelf-filters come in.

Shelf filters - the best of both worlds

A sensible place to start analysing the shelf-filters in Ambisonics is from our only design reference[11] (Gerzon 1977, for the Integrex horizontal-only hardware decoder) which perform as illustrated in Figure 13.

[11] Not entirely true! Hand drawn circuits of the MINIM decoder exist too. Interestingly, circuit analysis of this schematic reveals that the decoder is velocity optimised at LF (equal contributions of ***W, X***, and ***Y***) and that ***X*** degenerates to -6dB at HF relative to ***W*** and ***Y***. (The ***W*** and ***Y*** shelf filters are simply phase compensators to match the ***X*** channel). So, in this design, the left-right directivity is retained at HF, but front-back is sacrificed. This is presumably the result of some empirical optimisation.

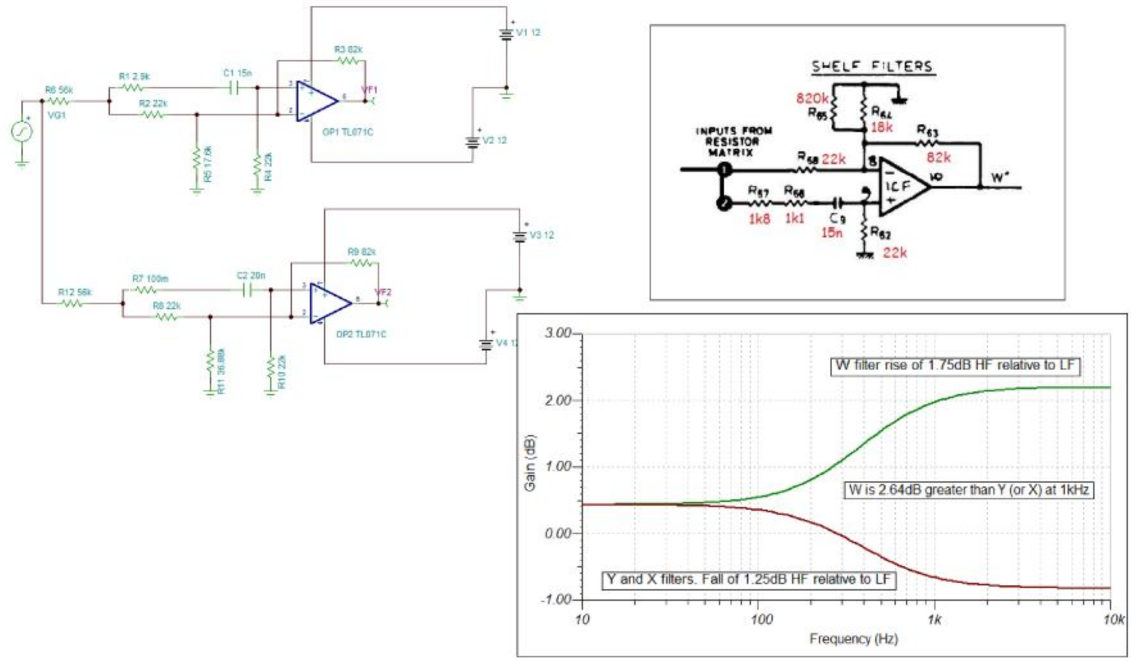


Figure 13 – The shelf filters in the Integrex decoder

As the curves in Figure 13 attest, the gain of the ***W*** channel at low frequencies is raised by approximately 3dB relative to the ***X*** and ***Y*** channel at HF. If the decoder is arranged to provide equal proportions of ***W, X*** and ***Y*** in the D-format decoding at low frequencies (in other words, to function as a *velocity optimised* decoder), the role of the shelf-filters will be to raise the ***W*** proportion - and reduce the ***X,Y*** proportion - so as to function as an *energy optimised* decoder at HF. Thus the best of both worlds is achieved relatively simply. The only complication is that the shelf filters must be arranged to be linear-phase (or at least have a matching phase response), otherwise the subsequent D-format matrixing will be affected because it relies on phase relationships between the ***W, X*** and ***Y*** signals to work properly[12].

Controlled opposites

Ambisonics was unashamedly designed for domestic reproduction with one (or a few) centrally placed listeners sitting equally between four or more loudspeakers. However, some of Ambisonics more tempting commercial and artistic applications exist in larger venues. The great practical difference between these two applications is the need, in the case of the large venue, to recreate a reasonable spatial illusion across a large space with many listeners sitting nearer one loudspeaker than the other three or more loudspeakers. Fundamentally, Ambisonics is "broken" for such applications because it relies *au font* on the separate sources being equally perceived. Yet, it has been noted (by Malham in his paper *Experience with large area 3-D Ambisonics Sound Systems*) that

[12] Sadly, the mystery of why the Integrex decoder actually reduces the ***W*** signal by 25% in relation to the ***X*** and ***Y*** signals at LF prior to D-format matrixing remains a mystery. My suspicion is that the decoder design contains an error of correction being applied twice...... Surely, we do not really want a narrower hyper-cardioid response at LF? We know that the vector velocity-field is best served by the hyper-cardioid (according to the theory) and the shelf filters do not adequately compensate for an optimised vector energy field at HF.

the system does appear to work better than the theory predicts. However, Malham discovered in listening tests that any hyper-cardioid pattern in the virtual microphones is destructive, because the reverse lobe has the effect of introducing a diagonal crosstalk component. In this case, it is better to degenerate the hyper-cardioid virtual microphone even further and have them based on simple cardioid patterns.

In order to degenerate the hyper-cardioid to a pure cardioid at HF, the ***W*** signal needs to be raised relative to the ***X,Y*** gain by,

1.414 / 0.707 = 2 times,

Because a cardioid has its null at 180 degrees to its frontal response, such a decoder is termed a *controlled opposites decoder*, or a *cardioid decoder*, or even - for enigmatic reasons - a *naive decoder.*

So, with our more complete knowledge, we can revisit the four apparently conflicting recommendations for the ratio of ***W*** to ***X*** & ***Y*** again in a regular polygonal decoder and say:

- ***W*** should be twice (+6dB) the amplitude of the ***X*** and ***Y*** signals.
    - This is for a *controlled opposites* decoder, suitable for large venues.
- ***W*** should be 3dB above the value of the ***X*** and ***Y*** signals.
    - This is for an *energy optimised decoder* which is a best case compromise or for best HF localisation is small spaces.
- ***W*** should be the same amplitude of the ***X*** and ***Y*** signals.
    - This is for a *velocity optimised* decoder which offers the best LF localisation in small spaces.

The latter two schemes may be fused together by means of a shelf filter so that the same decoder is optimised for velocity vector field at LF and energy at HF. As to the last recommendation (from praxis in the Integrex decoder):

- W should be 75% (-2.5dB) below the amplitude of the X and Y signals.
    - This appears to be a mistake.

As we now see, none of the formula is downright wrong (except the Integrex reference which appears to have a genuine error). But there is a lot in each case which has gone unexplained.

Equation for virtual microphone

Now, finally, we are in a position to understand and use the equation for the virtual microphone intelligently. It is the following:
A horizontal virtual microphone[13] at horizontal angle Θ with pattern $0 \leq p \leq 1$ is given by

$$M(\Theta, p) = p\sqrt{2}W + (1-p)(\cos\Theta X + \sin\Theta Y).$$

The ***p*** parameter may be selected thus:

Parameter *p*	**Pattern**
0	Figure of eight
0 to 0.5	Hyper-cardioid and super-cardioid
0.5	Cardioid
0.5 to 1	Wide cardioids
1.0	Omnidirectional

Other issues

Distance Compensation (or Near-field Compensation) filters

It is well known that a velocity microphone will exhibit a bass-heavy response if used close to a sound-source such that a pressure component exists between the front and rear face of the microphone membrane. This is usually termed *bass tip-up*. In an Ambisonics system - and where the decoder has been optimised for velocity optimisation - the same effect happens in reverse so that, in small spaces (and in sound wavelength terms, most domestic rooms are small), the reproduction can become "muddy" and "boomy" if the D-format signals are derived from the primary ***W, X*** and ***Y*** channels without further processing. The solution is very simple and involves high-pass filtering the ***X*** and ***Y*** signals (*not* ***W***) with a simple, first-order RC network with a breakpoint around 20Hz. In a digital system, these could be implemented with simple IIR filters.

5.1 decoding

Because of the commercial success of 5.1 surround sound, many more homes now have a multi-channel, multi-speaker system than ever was the case at the height of quadraphonics' short life. So, there is great interest in reproducing B- or C-format Ambisonics material on the ITU-R BS 775 loudspeaker set-up as illustrated below.

The naïve approach, of course, would be to feed each speaker with a virtual microphone angled in the direction of the loudspeakers; according to Figure 14. But, for various reasons, this approach doesn't work well. The first problem is that this arrangement is far from a regular pentagon (or even a regular square - if the *Center* is ignored). Moreover, the Center loudspeaker is often badly matched

[13] This virtual microphone is free-field normalised, which means it has a constant gain of one for on-axis sounds.

to the Left and Right loudspeakers thereby destroying the frontal image which remains the most important for all music applications at least.

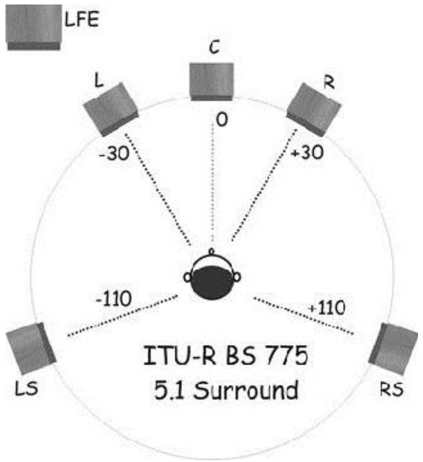


Figure 14 – ITU-R BS775 5.1 speaker arrangement

Gerzon and Barton produced an analytical recommendation to how virtual microphone signals might be defined for such a speaker layout (Gerzon & Barton 1992). These are termed *Viennese decoders* because they read the original paper at an AES Convention in Vienna. These solutions are still covered by a patent. However Wiggins has published a solution (Wiggins 2008) which is derived from an iterative computer optimisation method, the results of which are illustrated in Figure 15. What is more, Wiggins demonstrates that his solution substantially outperforms Gerzon's analytical solution.

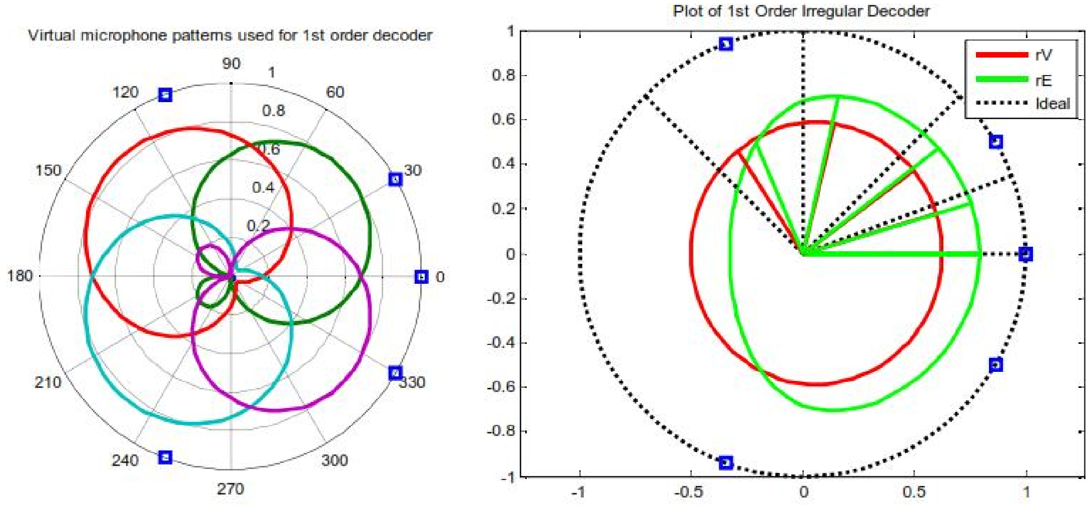


Figure 15 - Wiggins optimised first-order Ambisonics decoder for 5.1 speaker layout: results illustrated left and virtual microphone arrangement illustrated left.

Essentially, Wiggins retains the square array of microphones to energise the four speakers: Left, Right, Left Surround and Right Surround. These are still angled at 45°, 135°, 225°, 315° degrees with respect to the front, but the front loudspeakers are fed with hyper-cardioids and the rears with cardioids. The *Center* channel is left silent. This is the solution used in **Stereo Sauce** for 5.1 optimisation.

Patents and LOGO

All the original Ambisonics patents have now expired and the technique is public domain, including the patent for the tetrahedral microphone. The Ambisonic Logo, a trademark formerly owned by Wyastone Estade Ltd, also expired in 2010.

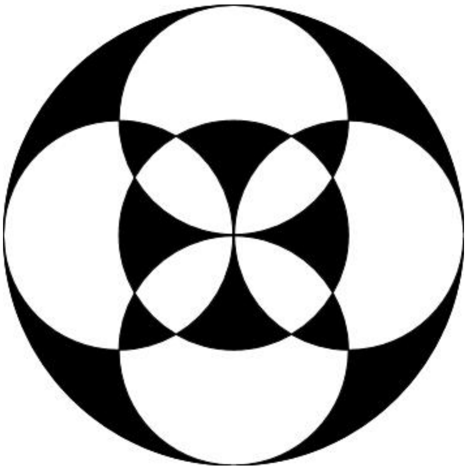


Figure 16 – The Ambisonics logo which – like the rest of the original patents – are now in the public domain

REFERENCES

Blumlein, A. (1933) British Patent 394,325 June 14th

Craven, P.G. and Gerzon, M.A. (1977), Coincident microphone simulation covering three dimensional space and yielding various directional outputs. United States Patent US 4,042,779.

Gerzon, M (1975), Ambisonics, Studio Sound, Vol. 17, pp 24-26, 28, 40 (August)

Brice, R (2012), Music Electronics Vol. 1 & 2, Transform Media Ltd.

Gerzon, M. (1977) Multi-System Ambisonic Decoder, Part 1: Basic Design Philosophy, Wireless World, vol. 83 no. 1499, pp. 43-47 (1977 July) Part 2: Main Decoder Circuits, Wireless World, vol. 83 no. 1500, pp. 69-73 (1977 Aug.)

Gerzon, M. (1985), Ambisonics in Multichannel Broadcasting and Video, JAES, Vol. 33, No. 11 (November)

Gerzon, M. A. & Barton, G. J. (1992) Ambisonic Decoders for HDTV. Proceedings of the 92nd International AES Convention, Vienna. 24 – 27 March.

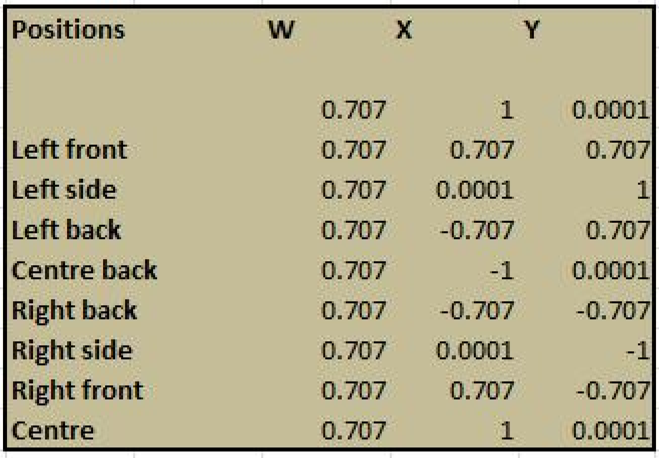
Wiggins, B. (2008), Has Ambisonics come of Age?, Proceedings of the Institute of Acoustics, Vol. 30, Pt. 6.

Appendix – B-format test sequence

Appreciation of the various decode optimisations is greatly aided by having a standardised test sequence of tones in various azimuthal positions so as to compare the results of the different optimisations. Moreover, such a sequence is absolutely essential for testing decoding equipment. None apparently being available, I generated one myself. Details of the sequence, and the results due to different decode optimisations are given here.

The test sequence

A 1kHz tone was electronically panned to eight cardinal azimuthal positions: centre; left-front, left-side, left-back; centre-back; right-back; right-side; and right-front. The details of the values of the W, X and Y channel are given in this table*.



Positions	W	X	Y
	0.707	1	0.0001
Left front	0.707	0.707	0.707
Left side	0.707	0.0001	1
Left back	0.707	-0.707	0.707
Centre back	0.707	-1	0.0001
Right back	0.707	-0.707	-0.707
Right side	0.707	0.0001	-1
Right front	0.707	0.707	-0.707
Centre	0.707	1	0.0001

** The values 0.0001 mean ù 0 but are there to stop divide by zero problems.*

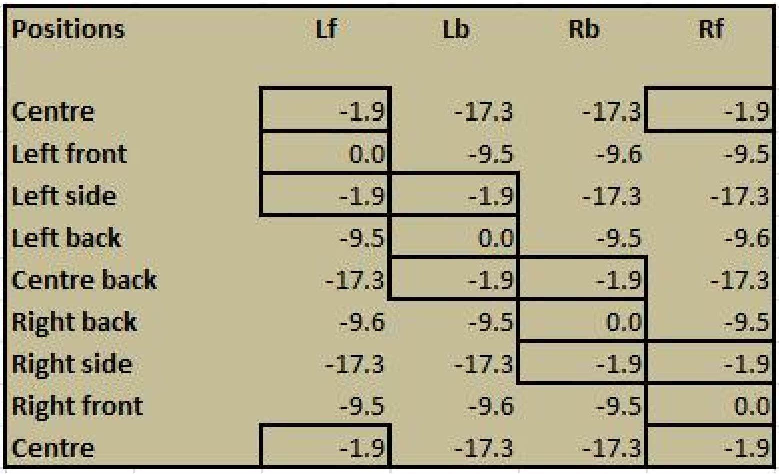
These values are as specified in the Ambisonics pan equations:

$X = S \,.\, \cos A \,.\, \cos B$ (front-back)
$Y = S \,.\, \sin A \,.\, \cos B$ (left-right)
$Z = S \,.\, \sin B$ (up-down)
$W = S \,.\, 0.707$ (pressure signal)

Where, because this is a horizontal only test sequence, ***B*** is 0° and therefore **cos *B*** = 1 and **sin *B*** = 0, thereby ensuring that ***Z*** is always zero.

Velocity optimised

The results of velocity optimisation to a four-speaker, regular square decoder (a = 0.334, and b = 0.333 : ***W*** is unity gain, pure LF Ambisonics) is given in the table below. The figures are in dB relative to an arbitrary reference. In each case, the strongest signal(s) is indicated by a box around the resulting figure. Note that polarity (ie. phase is not indicated: just the level of the signal for the four loudspeakers for a given source direction.



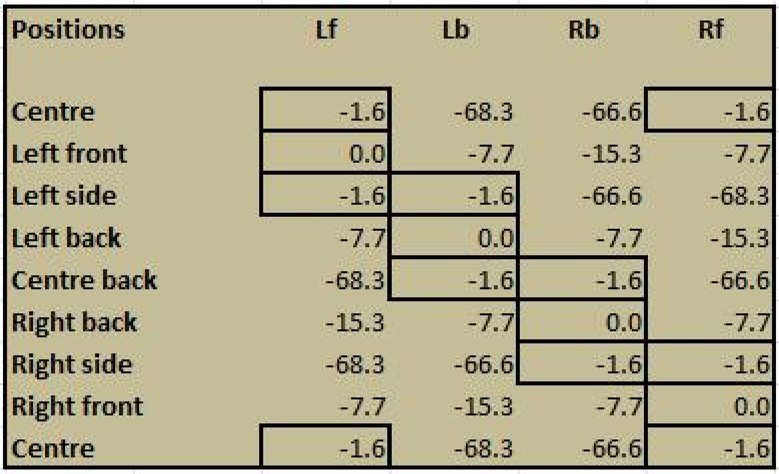
Positions	Lf	Lb	Rb	Rf
Centre	-1.9	-17.3	-17.3	-1.9
Left front	0.0	-9.5	-9.6	-9.5
Left side	-1.9	-1.9	-17.3	-17.3
Left back	-9.5	0.0	-9.5	-9.6
Centre back	-17.3	-1.9	-1.9	-17.3
Right back	-9.6	-9.5	0.0	-9.5
Right side	-17.3	-17.3	-1.9	-1.9
Right front	-9.5	-9.6	-9.5	0.0
Centre	-1.9	-17.3	-17.3	-1.9

Notable points:

- Front back separation is reasonable ≈ 15dB.
- There is substantial output from the diametrically opposite loudspeaker wrt. the source position (but investigation would show it is anti-phase).
- Adjacent channel separation is ≈ 9dB or more.

Energy optimised

Energy optimisation (a = 0.414, and b = 0.293 : ***W*** is 3dB greater than ***X*** or ***Y***) produces quite a different set of results - something of a surprise given that the parameter value changes appear quite small. This is the table for energy optimised decoding.



Positions	Lf	Lb	Rb	Rf
Centre	-1.6	-68.3	-66.6	-1.6
Left front	0.0	-7.7	-15.3	-7.7
Left side	-1.6	-1.6	-66.6	-68.3
Left back	-7.7	0.0	-7.7	-15.3
Centre back	-68.3	-1.6	-1.6	-66.6
Right back	-15.3	-7.7	0.0	-7.7
Right side	-68.3	-66.6	-1.6	-1.6
Right front	-7.7	-15.3	-7.7	0.0
Centre	-1.6	-68.3	-66.6	-1.6

Notable points:

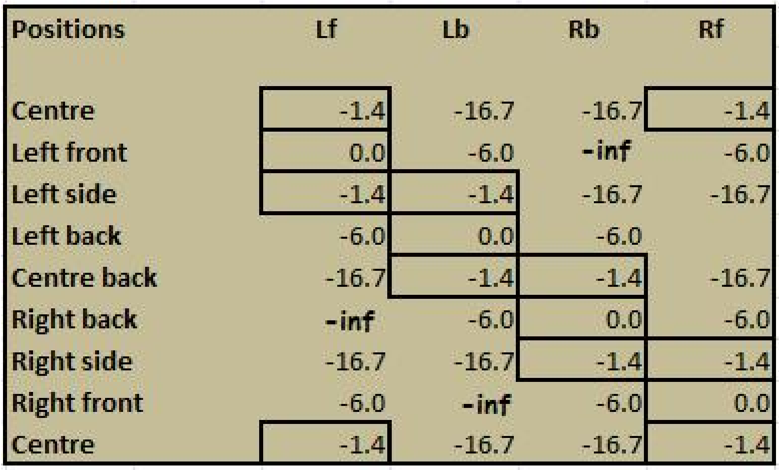
- Front back separation is very high ≈ 70dB.
- There is a small amount of output from the diametrically opposite loudspeaker wrt. the source position: about ≈ -15dB.

- Adjacent channel signal is medium: ≈ 8dB.

One can see how energy optimised is the best default decoder (at least, without shelf filters.)

Controlled opposites

These are the results from a controlled opposites, or cardioid, decoder (a = 0.5, and b = 0.25 : where ***W*** is twice as big as ***X*** or ***Y***).



Positions	Lf	Lb	Rb	Rf
Centre	-1.4	-16.7	-16.7	-1.4
Left front	0.0	-6.0	**-inf**	-6.0
Left side	-1.4	-1.4	-16.7	-16.7
Left back	-6.0	0.0	-6.0	
Centre back	-16.7	-1.4	-1.4	-16.7
Right back	**-inf**	-6.0	0.0	-6.0
Right side	-16.7	-16.7	-1.4	-1.4
Right front	-6.0	**-inf**	-6.0	0.0
Centre	-1.4	-16.7	-16.7	-1.4

Notable points:

- Front back separation is reasonable ≈ 17dB.
- Unsurprisingly (given the name), there is no output from the diametrically opposite loudspeaker wrt. the source position.
- Adjacent channel separation is low-ish: ≈ -6dB.

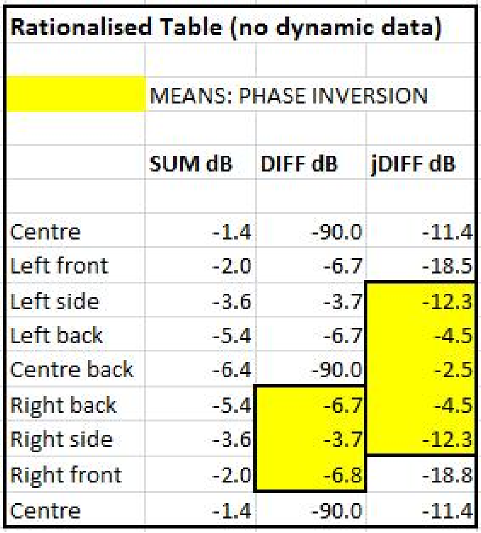
UHJ Test sequence signal

The B-format test signal was encoded into UHJ via signal manipulations according to the encoding equations:

S = 0.9396926*W + 0.1855740*X
D = j(-0.3420201*W + 0.5098604*X) + 0.6554516*Y
Left = (S + D)/2.0
Right = (S - D)/2.0

where j is a +90 degree phase shift

The resulting signal has the following characteristics:



Rationalised Table (no dynamic data)			
🟨	MEANS: PHASE INVERSION		
	SUM dB	DIFF dB	jDIFF dB
Centre	-1.4	-90.0	-11.4
Left front	-2.0	-6.7	-18.5
Left side	-3.6	-3.7	-12.3
Left back	-5.4	-6.7	-4.5
Centre back	-6.4	-90.0	-2.5
Right back	-5.4	-6.7	-4.5
Right side	-3.6	-3.7	-12.3
Right front	-2.0	-6.8	-18.8
Centre	-1.4	-90.0	-11.4

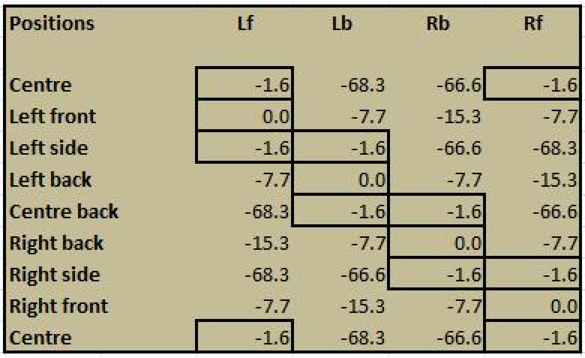
This signal was subsequently decoded (in ***Stereo Sauce***) according to the following equations:

S = (Left + Right) [*This is how it's done in **Stereo Sauce** (no divide by two)*.]
D = (Left - Right)
W = 0.982*S + j*0.164*D
X = 0.419*S - j*0.828*D
Y = 0.763*D + j*0.385*S

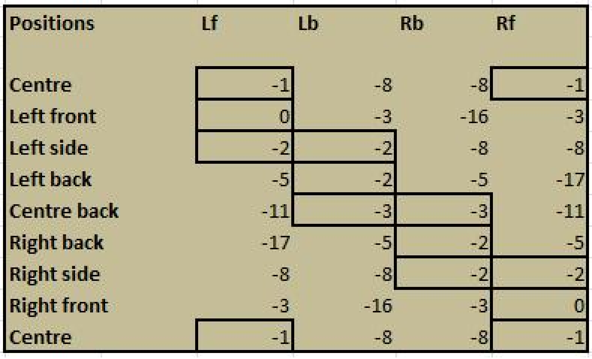
where j is a +90 degree phase shift

Comparison with UHJ results

It's instructive to compare the results of a pure WXY B-format decode with the measured results (from ***Stereo Sauce*** decode) of a UHJ encode-decode followed by WXY decode. The two tables below compare a pure *energy-optimised*, B-format decode with a UHJ decode, followed by B-format decode of a UHJ decoded test signal.



Positions	Lf	Lb	Rb	Rf
Centre	-1.6	-68.3	-66.6	-1.6
Left front	0.0	-7.7	-15.3	-7.7
Left side	-1.6	-1.6	-66.6	-68.3
Left back	-7.7	0.0	-7.7	-15.3
Centre back	-68.3	-1.6	-1.6	-66.6
Right back	-15.3	-7.7	0.0	-7.7
Right side	-68.3	-66.6	-1.6	-1.6
Right front	-7.7	-15.3	-7.7	0.0
Centre	-1.6	-68.3	-66.6	-1.6

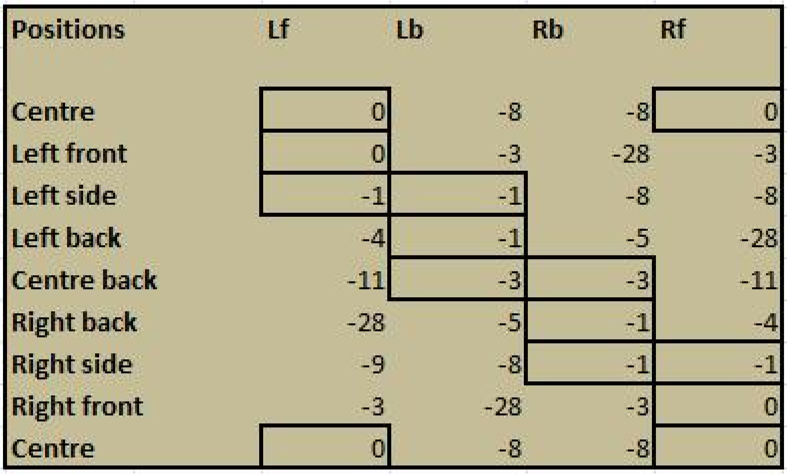


Positions	Lf	Lb	Rb	Rf
Centre	-1	-8	-8	-1
Left front	0	-3	-16	-3
Left side	-2	-2	-8	-8
Left back	-5	-2	-5	-17
Centre back	-11	-3	-3	-11
Right back	-17	-5	-2	-5
Right side	-8	-8	-2	-2
Right front	-3	-16	-3	0
Centre	-1	-8	-8	-1

Certainly, the same pattern of results is maintained. The lossy nature of UHJ has the following effects:

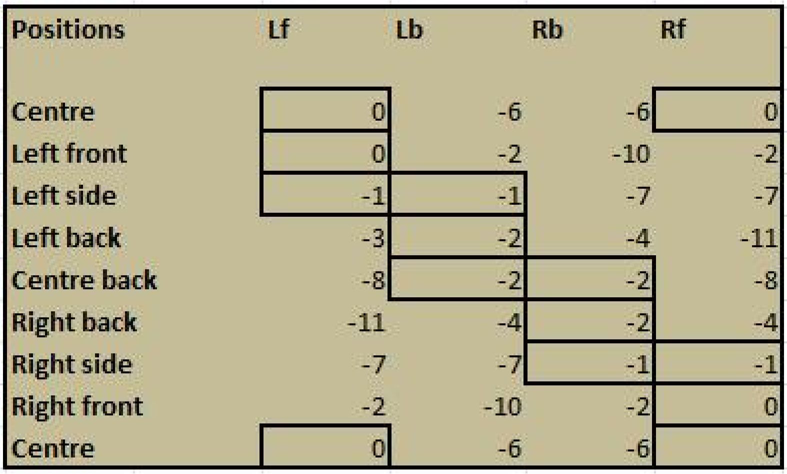
- Front-back separation is greatly reduced.
- Adjacent channel separation is reduced to 3dB (just adequate).

A UHJ decode followed by a velocity D-format decode produces the following results.



Positions	Lf	Lb	Rb	Rf
Centre	0	-8	-8	0
Left front	0	-3	-28	-3
Left side	-1	-1	-8	-8
Left back	-4	-1	-5	-28
Centre back	-11	-3	-3	-11
Right back	-28	-5	-1	-4
Right side	-9	-8	-1	-1
Right front	-3	-28	-3	0
Centre	0	-8	-8	0

Finally, a controlled-opposites decode (from the UHJ decoded test signal) produces these results.



Positions	Lf	Lb	Rb	Rf
Centre	0	-6	-6	0
Left front	0	-2	-10	-2
Left side	-1	-1	-7	-7
Left back	-3	-2	-4	-11
Centre back	-8	-2	-2	-8
Right back	-11	-4	-2	-4
Right side	-7	-7	-1	-1
Right front	-2	-10	-2	0
Centre	0	-6	-6	0

From which it is possible to observe that the various crosstalk mechanisms in UHJ encoding and decoding alter radically the nature of the decoding so that these options are not really relevant when it comes to decoding UHJ sources.

It may be said with certainty that, for UHJ-stereo sources, an energy decoder is most suitable. To gild the lily, it is even better to delay the rear signals of the UHJ encode by about 12mS. This has the effect of reducing the slightly over-reverberant nature of UHJ decodes and diminishes the “close” or “over-bearing” nature UHJ decodes often appears to manifest.

RAB 28th April 2014