Ambisonics - A Technique for Low Cost, High precision, Three Dimensional Sound Diffusion (Full Paper)

D.G. Malham
Music Technology Group
University of York
Heslington
York YO1 5DD

Email: dgm2@uk.ac.york.vaxa

(original shorter version given at ICMC GLASGOW 1990)

ABSTRACT: Ambisonic systems store information about three dimensional soundfields that can be used to produce via a suitable decoder/loudspeaker array the illusion of true three dimensional sound images. It can be shown from information theory that to reproduce the soundfield present in a two metre sphere 400.000 channels are needed. Ambisonics is based on storing only the information that would be available from a simple combination of microphones and then providing a way of using that information to fool the ear into perceiving a full soundfield. The simplicity of the system means that complex manipulations of soundfields are easily accomplished using computers.

When faced with the need to position sounds in space, other than over the somewhat limited sound stage of the conventional stereo presentation, composers of amplified music have been limited in their choice of technique. A multitrack tape or other multichannel source of raw material can be diffused through a multispeaker array in the performane space. This is equivalent to treating the sound diffussion system as a musical intrument. As such it needs a highly skilled performer to get the most out of it. At its best, the results can be highly satisfying, at its worst it can be disastrous. Either way, the composer may well not have any control over the diffusion process. The only other technique which has been widely used up to now is to have sounds panned into a multichannel system, each channel of which is allocated to one loudspeaker. This is most widely available as the four channel output option in Csound and other systems.

This so-called 'quad' system is based on the premise that if stereo ("two holes in a wall") is better than mono ("one hole in a wall") then quad must be better still. Unfortunately, as is now well known, this is untrue. Stereo is best when the two speakers subtend an angle of 60 at the listener. If, however, a four channel system is set up so as to meet this criterion the side speakers will subtend 120 angles so at best, side images will be severely unstable if they are attainable at all. Using a square format, this can be mitigated somewhat, at the expense of reduced front and rear image quality. Composers can work round these problems by carefully tailoring the sounds in the studio but this work can all be wasted if the performance speaker array does not closely match the original studio speaker layout. Even with this tailoring it has been shown that the pairwise panned quad format contains redundant information, conflicts within which causes the instability in the images. It is also, of course, wasteful of channel capacity.

During the early seventies, techniques were evolved which removed the redundancy whilst extending the information carried by the four channels of the quad format to include height formation. The conflicts which degraded image stability were also removed.

Research into directional perception resulted in methods of using the four channels of information to provide a psychoacoustical optimised set of speaker drive signals for any sensible set of symmetrical arranged speakers. Systems which are based on techniques which have these features include the BMX system from Nippon-Columbia and Ambisonics which was developed in Britain. Of these, Ambisonics is the most highly developed. It was concieved as a recording system to allow the capturing and reproduction the full acoustic image ('soundfield') that a listener would percieve in a live situation. There is even a microphone available - the Soundfield microphone - which produces all four required signals simultaneously. It should be noted that the Dolby surround sound system "soundfield" is not a soundfield in this technical sense of the term.

Basic signal definitions

Whilst the Ambisonic system was evolved as a recording technology, the equations which govern the relationships between the four channels which make up the studio format signals (technically known as 'B' format signals) are simple enough to allow easy manipulation with a computer program. The signals are equivalent to those that would be produced by three mutually perpendicular figure of eight microphones and one omnidirectional microphone. These may be written as :-

X	=	$\cos(A).\cos(B)$	(forward facing microphone) (1)
W	=	0.707	(omnidirectional microphone) (2)
Y	=	$\sin(A).\cos(B)$	(sideways facing microphone)(3)
Z	=	$\sin(B)$	(upwards facing microphone) (4)

for a sound source with a horizontal displacement angle of A measured anticlockwise from a centre front position and a vertical angle of B. The constant factor of 0.707 in the omnidirectional channel is based on engineering considerations, reflecting the need to have roughly equal levels in all channels, and is not required for the basic theory to work.

If we constrain the sound source to move on the surface of a notional unit sphere around the listener, we can simplify these equations even further to

$$X = x$$

$$W = 0.707$$

$$Y = y$$

$$Z = Z$$

where x,y,z are the ordinary Cartesian coordinates of the sound source. Constraining it to move on the surface of a sphere is equivalent to requiring that $(x^2 + y^2 + z^2)^{0.5} = 1$.

These definitions do not include a distance component. We will see later how distance cues may be accurately coded, but for now we will just note that the technique that is used in analog Ambisonic systems is to reduce the X,Y,Z signals and increase the W signal. This gives an impression of the sound moving towards the listener. If the X,Y,Z signals are allowed to go to zero and then increase again but with opposite sign, the sound appears to pass through the listener and emerge the other side. In this case the RMS value, if we may call it that, of the X,Y,Z values must be equal to or less than 1. Typically the omnidirectional signal is made to to vary as:

$$W = 1 - 0.293(x^2 + y^2 + z^2) \quad (5)$$

which keeps the overall volume reasonably constant. It might, however, be more appropriate to have the overall level increase as the sound is panned towards the listener as this is more natural.

It is possible to place as many sound sources as you like in a soundfield, simply by mixing their individual B format signals into a master B format. The only mandatory requirement is that if you wish the sounds to remain in their original relative placings then any processing (fading, equalisation etc.) that is done to one of the channels in the B format signals of that sound must also be done to all its other channels. So if you want to reduce the level of one particular sound in the final image you must reduce all four of its B format channels equally. Similarly, if you wish to reduce the high frequency content, you must use the same low pass filter on each of the four channels. The only exception is reverberation. If you add reverb to all the components of a sound source's B format, the reverberation will tend to appear to be coming from the direction of the sound and not from the diffuse acoustic field of the room you are simulating. Under some circumstances this might be a useful effect but it does not correspond to natural reverberation. The topic of reverberation will be covered more fully later.

Apart from the advantages that we have seen so far, the most attractive thing about Ambisonics from the computer musicians' point of view is the extraordinary ease with which manipulations of whole sound images containing arbitrary numbers of sources with arbitrary positions can be manipulated.

<u>Soundfield manipulations.</u>

In a conventional quad system with pairwise panned material, if the composer decides after completing some complex sound image to move the image say 25 degrees to the right, the only way it can be achieved is to recompile the whole thing again with all the panning positions shifted 25 degrees. If the sound image was in Ambisonic format, the following transformation of the four channel B format soundfile would do the same job and take a lot less computing power.

$$X' = X.\cos(-25) - Y.\sin(-25) \qquad (5)$$
$$Y' = Y.\cos(-25) + X.\sin(-25) \qquad (6)$$
$$W' = W$$
$$Z' = Z$$

where X',W',Y',Z' are the B format signals in the rotated output soundfile and X,W,Y,Z are the corresponding signals in the input soundfile. The minus sign for the angle is because it is clockwise rotation.

This can be extended to the vertical axis as well with only a moderate increase in computational complexity. To give you some idea, here is the formula for rotate then tilt:

$$X' = X.\cos(a) - Y.\sin(a) \qquad (7)$$
$$W' = W$$
$$Y' = Y.\cos(a).\cos(b) + X.\sin(a).\cos(b) - z.\sin(b) \qquad (8)$$
$$Z' = Z.\cos(b) + X.\sin(a).\sin(b) + Y.\cos(a).\sin(b) \qquad (9)$$

where *a* is the angle of rotation and *b* is the angle of tilt. It should be noted at this point, that the formulae presented in this paper hold true for soundfields of any complexity, even those produced by from live soundfield recordings which, for reverberant spaces, can have an essentially infinite number of separate sound sources.

B format signals can be manipulated to provide even more exotic transformations. For instance, by cross blending the X and W signals in appropriate amounts it is possible to zoom in on sounds that are either to the front or rear of the listener. This effect has been called 'dominance' and can be described by the formulae:

$$X' = X + \sqrt{2}.d.W \qquad (10)$$
$$W' = W + \frac{1}{\sqrt{2}}d.X \qquad (11)$$
$$Y' = \sqrt{1-d^2}\,.\,Y \qquad (12)$$
$$Z' = \sqrt{1-d^2}\,.\,Z \qquad (13)$$

where d varies between +1 for frontal sounds dominate to −1 for back sounds dominate.

In analog systems it is usually to allow front-back or up-down dominance only, due to the complexity of the controls required. In a computer based digital system it is a relatively simple matter to point the dominance function in any direction. While this might not necessarily be of any use when dealing with synthesized soundfields. when dealing with natural soundfields it can be of great use in emphasizing particular sounds.

Another effect that might be of interest is the soundfield mirror. If you multiple the Y channel, for instance, by a factor which varies from +1 to −1 the whole soundfield will vary from normal through an image on a single vertical plane going vertically front-back (when the multiplier equals zero) to an image with left and right exchanged. Again, it is easier to extend this to arbitrarily oriented mirroring planes when manipulating the B format signals in the digital domain.

As I have reported elsewhere (ref 15), all these effects can be applied simultaneously to the B format signals by using a matrix multiplication of the form:

$$X_{out}=k_1X+k_2W+k_3Y+k_4Z$$
$$X_{out}=k_5X+k_6W+k_7Y+k_8Z$$
$$X_{out}=k_9X+k_{10}W+k_{11}Y+k_{12}Z$$
$$X_{out}=k_{13}X+k_{14}W+k_{15}Y+k_{16}Z$$

where the channel identifiers have the same meaning as before and the coefficients $k_1 - k_{16}$ are formed by appropriate matrix operations from the multipliers of such equations as (5) to (13) as required. If spatial manipulations are coded at a k-rate lower than sample rate, as is normal, this represents a considerable saving in computation when more than one manipulation is needed on any given soundfield.

All the manipulations so far covered have been frequency independent. Furthermore, sounds panned into the sound image have sharp, pin-point localisation, but it will often be desirable to have a more diffuse presentation. We will be presenting a paper at a later date on means of synthesizing extended radiating surfaces, but here we just note that by adding phase shift networks in the signal paths it is possible to make to produce frequency dependent encoding of direction. This causes components with different frequencies to come from different directions. The effect of this is best on complex sounds when a diffuse spread of sound is desired but it may be that the automatic change of source localisation with pitch that is obtained with simple sounds could be a useful operator. The precise mechanisms and applications of techniques such as this are far too extensive to cover here but figure 1 is included to give some indication of the possibilities. This takes the two input signals and spreads one around and around a circle above the listener and the other around below the listener.

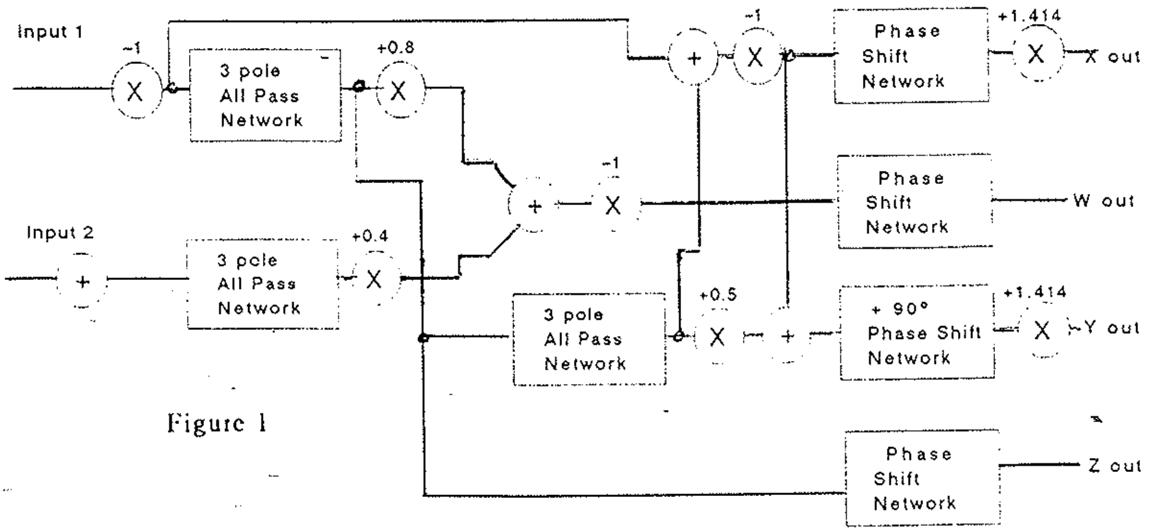

Figure 1

Reverberation

One simplistic way of adding reverberation is to take the W signal, pass it to a reverberator and mix the output of the reverberator back into the W channel. While this does work to some extent, it does break one of the main conditions for Ambisonics systems, namely that for any competently designed loudspeaker array the sound image should be the same, within the limits of the acoustics of the performance space.

An alternative approach is to use the sound-spreader approach mentioned above. The two inputs of the spreader of figure 1 are fed with separate reverberated inputs derived from the W signal and then mixing the resultant B format signal with the dry B format. In order to avoid comb filtering effects it is important that the two inputs are not correlated to any significant degree. If this is a problem, a delay of 20-30 mS can be included in one channel.

While this approach does work, it is limited by its design origins, which were analog. It is possible to do much better in the digital domain by using an Ambisonics based room simulation. Conventional ray tracing methods for simulating the effects of early reflections by setting up multiple phantom images of the source can be enhanced by providing each phantom image with its own set of B format signals within the final mix that is sent to the overall reverberator. This kind of approach can be extended as far as the available computing power will allow. It is possible to use different reverberators for different directions in the soundfield, thus giving more control over the final sound image properties.

Simulation of Distance

As was intimated earlier, the technique used in analog systems for providing a sensation of varying sound source distance is workable but is not the best that can be achieved. Sound images created in that way have most distant positions which correspond roughly to the radius of the sphere on the surface of which the loudspeakers are placed. As with the simplistic reverberation using the W channel only, this does not make best use of the potentialities of Ambisonics. As the sound is moved towards the listener, the sound image becomes more and more indistinct as the amount of directional information, which is carried by the X,Y and Z channels, is reduced. A moments' thought will show that in real situations, the directional information does not reduce as a sound source moves towards you, except in the distinctly unnatural case of a sound moving inside your head. In fact, the X,Y and Z channels should increase in amplitude along with the W component until the centre point is reached, whereupon they should smoothly reverse in phase with respect to the W component and then all four should then start reducing in amplitude again.

Distance cues in natural soundfields are provided by a number of different mechanisms, most of which tend to be rather difficult to handle in the analog domain and relatively easy to simulate digitally. These mechanisms include:

a) Acquired knowledge.

Mental modelling of the loudness of known sound sources enables judgment of their distance on loudness alone.

b) Reverberation.

In a reverberant soundfield, the level of reverberation stays roughly constant, whereas the level of the direct sound varies with distance.

c) Early reflections.

The pattern of early reflections produced by a sound source in a non-anechoic acoustic changes with changing sound source position. This can provide cues not just for distance but for many other things as well. For instance, a phenomena that was observed early on by the author was that even on horizontal only Ambisonic presentations with familiar sounds such as pianos, it was possible to get a distinct impression of the height of the Soundfield microphone, *even though* *no height information was being reproduced.* This effect has since been reported in the literature. It could possibly be due to the brain interpreting the comb filter effect caused by the path length difference between the direct sound and the first reflection off the floor. There is a known phenomena, that may be related, whereby white noise presented over headphones can be made to produce the sensation of a tone moving up and down in space by using a filter to create a notch of varying frequency. All these cues are easy to synthesize using a reasonably detailed model of the acoustic space we are trying to produce. By working with Ambisonic signals, we can improve the modelling accuracy, particularly of the early reflection pattern, because the modelled sounds can have true three axis directions associated with them. The prospect is for both far closer modelling of real spaces together with a much wide range of possibilities for producing virtual spaces which are not physically realizable.

Further Developments.

It is interesting to note that even greater precision is possible if you add second order signals (W is a zero order signal and X,Y,Z are first order) to the B format (ref 16) to give a nine channel system. These extra channels are of the type x.y, x^2-y^2, $3z^2-1$, x.z, y.z and their presence does not affect the original four B format signals so the system is downwardly compatible. What you would do with all this extra precision is uncertain as no-one has published any decoder design information yet.

Ambisonic sound diffusion systems

Decoders are designed to take the B format signals and apply them to an array of speakers in such a way that the contribution from each speaker adds vectorially at the listening position to give a sound image with each source giving an appropriate sound vector.

Because the ear uses different directional mechanisms above 700Hz (ENERGY VECTOR THEORY) and below 700 Hz (MAKITA THEORY) the vectorial sums have to be done differently in the two different areas. This is accomplished by filtering the B format signals prior to applying them to the amplitude matrix which feeds the speakers.

For a three dimensional array there are some rules to be followed. You need at least three pairs of speakers - preferably more - distributed reasonably evenly around the centre. The members of each pair should be at opposite ends of a line running through the centre and all speakers should be the same distance from the centre.

Each speaker is then fed with the following sort of signal:

$W + aX + bY + cZ$

where for a speaker at position $X_iY_iZ_i$ where X,Y,Z ARE GIVEN IN TERMS OF THE RADIUS, you calculate a, b and c from the following matrix formulae. (M is the number of pairs)

$$\begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} = \frac{1\ m}{\sqrt{2}} \left[\sum_{h=1}^{m} \begin{pmatrix} x_h^2 & x_h y_h & x_h z_h \\ x_h y_h & y_h^2 & y_h z_h \\ x_h z_h & y_h z_h & z_h^2 \end{pmatrix} \right]^{-1} \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix}$$

For a periphonic system the gains of the B format signals should vary with frequency as follows:

Low Frequencies High Frequency

W 0.00 dB +3.01 dB

X,Y & Z +1.76 dB 0.00 dB

The filters do not need to be sharp but the phase characteristics of the two filter characteristics should be matched.

PLEASE NOTE

People planning to produce devices based on Ambisonic technology should note that commercial exploitation is covered by a comprehensive set of patents which are held by NIMBUS RECORDS. ALL enquiries should be directed to:

Stuart Garman

Nimbus Records Limited

Wyastone Leys

Monmouth NP5 3SR Phone 0600 890 682

B Format connector standards

5pin XLR (Cannon) connector
Input–male, Output–female

Gnd pin 1 screen
X pin 2 blue
W pin 3 clear
Y pin 4 yellow
Z pin 5 red

6 pin DIN
Chassis socket, cable plug
no explicit gender for input or output

Gnd pin 1 screen/green
X pin 2 blue
W pin 3 white
Y pin 4 yellow
Z pin 5 red
R pin 6 violet
(R is a mono reverb feed)

References

1) GERZON, M.A. "The Principles of Quadrophonic Recording"

Part one. STUDIO SOUND. August 1970. pp338–342

2) GASKELL. P.S. "Spherical Harmonic Analysis and Some Applications to Surround Sound" BBC RESEARCH DEPARTMENT REPORT NO. BBC RD 1979/25 (1979)

3) GERZON. M.A. "Ambisonics Part two: Studio Techniques" STUDIO SOUND. August 1975. pp24 – 30.

4) GERZON. M.A. "Panpot and Soundfield Controls" NRDC AMBISONIC TECHNOLOGY REPORT NO. 3. August 1975.

5) GERZON. M.A. "Artificial Reverberations and Spreader Devices" NRDC AMBISONIC TECHNOLOGY REPORT NO. 4. August 1975.

6) GERZON, M.A. ""Sound Reproduction Systems" BRITISH PATENT NO. 1 550 628 August 15, 1979.

7) FELLGETT, P.B., GERZON, M.A. "From Quadro to Surround Sound" ELEKTORMay 1977 pp 5/19 - 5/25

8) GERZON, M.A. "Surround Sound Decoders" in 7 parts, WIRELESS WORLD 1977

9) GERZON, M.A. "Design of Ambisonic Decoders for Multi Speaker Surround Sound" Presented at the 58th AES CONVENTION, New York, 4 November 1977.

10) GERZON, M.A. "PRACTICAL PERIPHONY: The Reproduction of Full-Sphere Sound" Presented at the 65th AES CONVENTION, London, 25 February 1980.

11) FARRAR, K. "Soundfield Microphone" WIRELESS WORLD, Otober/November 1979

12) MALHAM D.G. "Digitally Programmable Soundfield Controller" STUDIO SOUND, Vol 26 No 2 pp 75, 1984.

13) MALHAM D.G. "The Spacio-Acoustic Processor" PROCEEDINGS OF THE INSTITUTE OF ACOUSTICS AUTUMN CONFERENCE ON REPRODUCED SOUND, Windermere 1985, pp 165 - 170.

14) FELLGETT, P.B. "Ambisonic Reproduction of Directionality in Surround Sound Systems" NATURE, Vol 252 December 13, 1974 pp 534 - 538.

15) MALHAM, D.G. "Computer Control of Ambisonic Soundfields" Preprint No. 2463(H2) presented at the 82nd AES convention 1987 10-13 March, London

16) GERZON, M "Periphony: With Height Sound Reproduction" Journal of the Audio Engineering Society, Vol 21, No.1 Jan/Feb 1973, pp2-10

Other relevant (British) patent numbers include 2073556 (Periphonic decoders), 1411994 (irregular speaker arrays), 1550627, 1494751, 1494752, 1512514