Ambisonic Primer

Dave Hunt
05/01/2014
T: +44 (0)208743 3055
M: +44 (0)7798 844 091
E: davehuntaudio@btinternet.com
dave@3d-audioscape .com

Ambisonics

Ambisonics was invented by Michael Gerzon (and others) in the 1970s. It is a blend of mathematical theory, extending Alan Blumlein's theory of stereo, and psychoacoustic research into directional hearing. Although not widely known or used, work on ambisonics has continued since then, notably by David Malham and Jerome Daniel. There has been a resurgence of interest in it over the last ten years due to its capability, efficiency, scalability and compatibility with current and foreseen surround systems.

Ambisonics attempted to recreate for a central listener, surrounded by a number of loudspeakers, the complete sound field of an original recording. The "Soundfield" microphone was produced to make such recordings. This has been available for around four decades, and is generally considered to be a high quality microphone. It is often used to record normal "stereo" due to the ability to use its outputs to manipulate its pick-up pattern and direction. It uses four closely spaced sub-cardioid directional microphone capsules in a tetrahedral array. The four-channel microphone signals are known as "A-Format", and there a few modern microphones, including one from Soundfield and the TetraMic from Core-Sound, which output A-Format.

The signals from these are combined to produce four output signals:
W, equal to the sum of all four capsules, an omni-directional microphone,
X, a fure-8 microphone pointing forward,
Y, a figure-8 microphone pointing left,
Z, a figure-8 microphone pointing up.
W was reduced in gain by 3 dB to make it of similar amplitude to the other three signals. For 2-D horizontal work, Z is ignored.

This four-channel audio signal was dubbed "B-Format", and is an ambisonic encoding of the sound field. Ambisonically encoded material must be decoded before being sent to loudspeakers.

Figure-8 microphones are "pressure gradient" (difference of pressure between two closely spaced points at the front and the rear) microphones. They are also known as "velocity" or "bidirectional" microphones. An important factor is that the output signal is of opposite polarity at the rear of the microphone relative to the front. Possibly because of this ambisonics is often described as using amplitude and phase, though it is more accurate to talk of amplitude and polarity. Polarity is invariant with frequency, whereas phase is not.

A "cardioid" microphone can be produced by combining an omni and a figure-8 microphone, the opposing polarity of the figure-8 resulting in a null at the rear of the combined microphone. This does not mean that a Soundfield microphone is equivalent to three cardioid microphones, as the W,X,Y,Z signals are "orthogonal" in that they represent four different dimensions: pressure and three directional pressure gradient dimensions. They remain distinct through the encoding/decoding process, and are only combined in different ratios to produce loudspeaker signals. Here the signal from a given loudspeaker might be regarded as that of a directional microphone, pointing at that speaker, to the whole sound field.

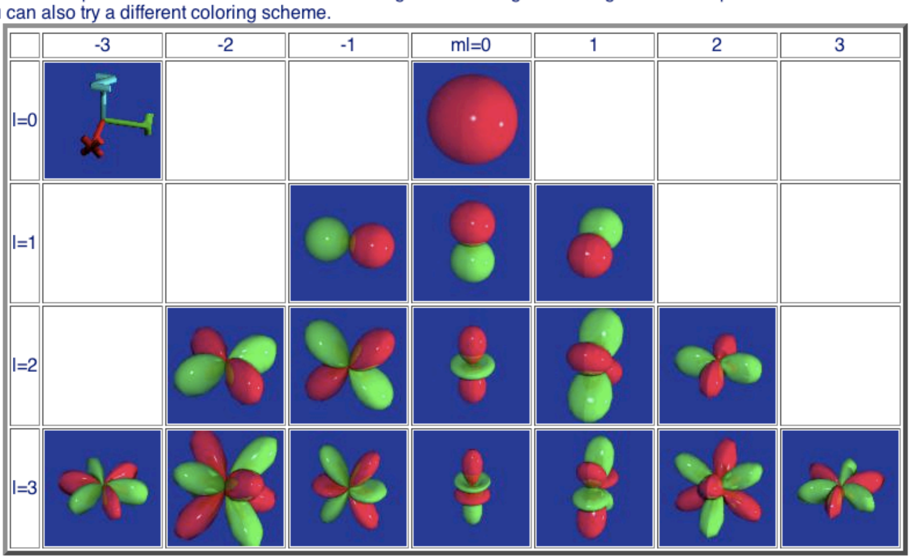
An ambisonic encoding of a mono audio signal can be made by applying differing gains to this signal to produce the W,X,Y,Z signals. A forward X signal for example will have a positive gain, a rearward signal a negative gain. Forward and back have opposite polarities.

This is known as "First Order Ambisonics" or POA (Plain Old Ambisonics). More generally, ambisonic synthesis uses a number of directionally encoded signals, in which sound directions are encoded into a number of channels with gains and polarities that vary with direction, specified by directional coding equations. The encoded signals represent the sound field at a central listening position.

These signals can be described as "spherical harmonics", of which the fundamental is a sphere, the first a set of three figures-of-eight at right angles to each other, and the second and beyond by more and more exotic directional shapes with all sorts of lobes.

Departamento de Química Física y Analítica
Universidad de Oviedo, 33006-Oviedo, Spain

Real combinations of spherical harmonics. Click on each image to see a large rendering of the same plot. Colors indicate the sign of the function. You can also try a different coloring scheme.



http://web.uniovi.es/qcg/harmonics/harmonics.html (May 2007)

The ambisonic order of the encoding corresponds to the value of l in the above diagram, though it usually designated as "M" in the literature.

Zero order is just equivalent to an omni-directional microphone, which effectively turns the intensity of air pressure variations (sound) into electricity.

First order (including the l = 0 and l = 1 harmonics) ambisonics is thus equivalent to an omni-directional microphone, and three figure-8 microphones, one pointing forward, another to the left and the third upwards.

Second order ambisonics uses the l = 0, 1, and 2 harmonics, the last usually labelled as R,S,T,U,V, a total of nine signals. They are equivalent to second-order microphones, which would have similar directional pick-up patterns. No such microphones have ever been manufactured, probably because they would have little practical use in audio recording or sound reinforcement, and present engineering challenges that would make them expensive to produce. The term B-Format has been extended to include second and higher orders.

A second order or above B-Format microphone would be difficult, or even impossible, to manufacture. Recently mhacoustics have produced the Eigenmike, based on research work in this direction by several parties. This uses an array of 32 omni-directional microphone capsules mounted in the surface of a rigid sphere, which also houses microphone preamplifiers and digital to analogue converters. An "Eigenmike Micophone Interface Box" converts the signals to a Firewire audio interface standard, and control software running on a computer processes the signals to produce those of various-order microphones simultaneously. The microphone can be "steered" and "focused"., even after recording its outputs This is very much a niche product, and thus very expensive. Even first order ambisonic microphones are not cheap, and beyond the budgets of casual users.

Higher order ambisonics (aka HOA) have been investigated by many parties. The aim is to improve spatial precision, and make ambisonics work more effectively over larger areas. Higher orders require more signal channels and more loudspeakers. Third order requires sixteen channels.

A mono signal can be encoded to B-Format of any order, and thus spatially positioned, with a software encoder in the form of a 'panner'. A stereo signal, which contains some 2D spatial information, can also be encoded to B-Format. Several such encoded audio streams, and those recorded with a Soundfield microphone, can be mixed together to produce a composite B-Format output.

A B-Format signal can be manipulated in several ways. The relative balance of W to XYZ (and higher order components) can be altered, making the directionality more or less pronounced. The gains of X, Y, Z etc..can be altered independently, changing the depth, width or height of the sound field, or reversing Front/Back. Left/Right, Up/Down. The sound field can be rotated about any or all of the three major axes, something not possible with binaurally coded or 5.1 sound. It can be made ‘dominant’ in a given direction, a sort of ‘zoom’ control, though this is not the same effect as in a zoom camera lens. This can only be done with first-order ambisonics, the mathematics of doing it higher orders having been deemed insoluble.

B-Format signals can be mixed together, and with B-Format encoded mono, stereo, M/S, quad, ambisonic. 5.1 etc. signals. Thus an audio production system can work directly with sound in a variety of formats.

A decoder is used to transform the B-Format encoded signal into loudspeaker feed signals. It produces signals for an array of loudspeakers that (of necessity) surround the central point. The loudspeakers are assumed to lie on the surface of a sphere. The sum of these loudspeaker signals reproduces the sound field at the central point.

The encoding and decoding are completely separate processes. A B-Format signal can be decoded for nearly any array of any number of speakers, and thus the B-Format signal is a ‘universal’ transmission medium, in contrast to a ‘channel per loudspeaker’ scheme such as quad, 5.1, 7.1 etc.

The sound field is produced by all the loudspeakers, working together. Usually all speakers output some signal, even to reproduce a sound apparently emanating from only one of them. Thus the effect of the sound being pulled to the nearest speaker is diminished, and the panning is smoother. The listening area for ambisonic surround sound is comparable with that for conventional stereo, but larger.

The average distance of the speakers from the listener, depends on the size of the loudspeaker array. Thus, when moving a mono sound source with a ‘distance panner’ such as those provided with 3DAudioScape, the position is not represented by distances measured in metres or feet and inches, but by ratios to the average distance of the loudspeakers. This is to ensure that decoders for different sizes of loudspeaker array do not change the spatial quality of the reproduced sound field.

Generally, a minimal number of speakers is required. This minimum depends on the “ambisonic order, M”, and whether the reproduction is planar (2D) or “periphonic” (3D). In theory the minimum number.is equal to the number of ambisonic channels, though a few more gives better results with more stable sound localisation. Using too few will result in insufficient reproduction of sound sources placed between speakers, and the perception that sound is coming from the speakers.

Thus 2D first order ambisonic reproduction needs a minimum of three speakers, and 3D first order reproduction a minimum of four. This generally considered sub-optimal. For 2D at least four speakers are preferred, and eight for 3D.

“Regular” layouts, those with all speakers at the same distance from the central listening position, with equal angles between all pairs of adjacent speakers use the simplest decoder designs. The mathematics of decoder design for “irregular” loudspeaker layouts becomes very complex, whereas solutions for “regular” layouts are much simpler. For 2D ambisonics a regular polygon (a square, pentagon, hexagon, octagon etc.) with the speakers at the points or at the centre of the faces, are possible layouts.

For 3D surround possible regular arrays are a tetrahedron (sub-optimal), cube, dodecahedron or icosahedron. Pyramid layouts, such as a square with a single overhead loudspeaker and a single speaker below the listener, are simple and reasonably effective.

There are arguments which would suggest a practical limit to the maximum number of speakers for a given ambisonic order. The more speakers, the more the output from adjacent speakers becomes similar. As the output of all speakers combine at the listener, greater directional precision is not achieved by using more speakers, and the cost of these speakers plus amplification, cable and speaker mountings rises.

Higher order ambisonics, incorporating second-order and above harmonics, gives more precise positioning but requires more audio channels, more loudspeakers, and more processing.

Modern computers have enough power to provide an ambisonic audio production system, and a flexible and scalable ambisonic decoder and playback system for domestic or performance purposes. Live sources can be incorporated, and it can be used with video in a number of ways.

Ambisonic Loudspeaker Arrays

The decoding equations used within 3DAudioScape are those for ‘regular’ loudspeaker arrays,which often use a number of pairs of opposing speakers.

The array should ideally be a regular polygon, with the speakers at the points or at the centre of the faces, for 2D surround (square, pentagon, hexagon, octagon etc.).

For 3D surround possible regular arrays are a tetrahedron (generally considered sub-optimal), cube, dodecahedron or icosahedron. Pyramid layouts, such as a square with a single overhead loudspeaker and a single speaker below the listener, are simple and reasonably effective.

Another possibility is a horizontal square at mid height, and a vertical square aligned with the main axis across the listening area. This could be extended with a vertical square aligned with the main axis front-rear of the listening area.

Attempts should be made to get as near as possible to one of these layouts. Perfection is rarely achievable, and compromises always have to be made. Nearly every space has limitations as to where and how speakers can be mounted. In particular under-floor loudspeakers can only be housed successfully and safely under a steel open-grid floor. Even those mounted on the floor are a trip hazard. Similarly, speakers mounted at height require higher than normal ceilings, and solid structures. Do not be discouraged. Do the best you can, and the results should be usable and effective.

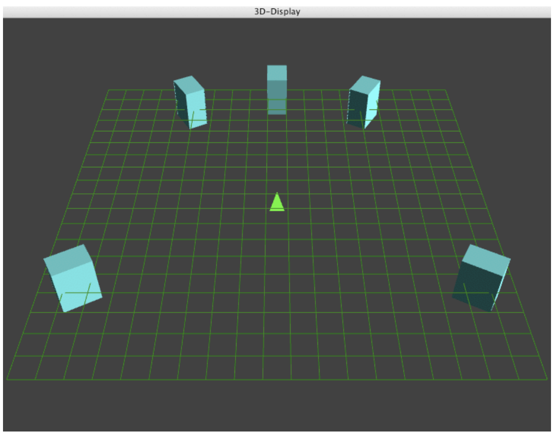
It may be possible to use a distributed ceiling loudspeaker system to convey height, where a suitable system exists and it is not feasible to rig any alternative.

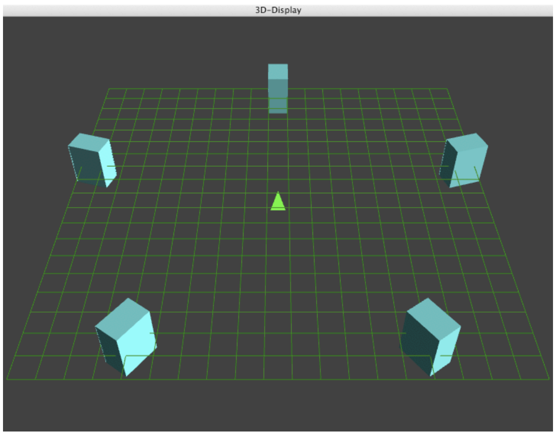
The mathematics of irregular arrays is fairly intractable, and for some cases there is no practical solution. In 3DAudioScape it is apparently possible, in the decoder, to place a speaker at will, but the results will not necessarily be or sound correct.

Cinema 5.1 systems have three speakers, left, centre and right, behind the screen, and left-surround and right-surround channels, usually with a number of speakers distributed to the sides and rear of the audience. The aim is to produce a diffuse surround effect with no clear direction.

Studio and domestic 5.1 systems, in theory have speakers at 0 degrees, plus and minus 30 degrees, and plus and minus 110 degrees from front centre, though there are large variations particularly with domestic systems. Again, the aim is to produce a diffuse surround effect, although music producers may be aiming for something a bit more immersive and definite.

These arrays have no good solution to the ambisonic decoding equations. Better results can be obtained by moving the speakers so as to form a pentagon, with speakers 72 degrees apart.





Some seem to favour a layout with a centre-rear and left/right-front layout (which could be regarded as a front-rear mirror of this layout). The reason for this is unclear apart from a general distaste for a centre speaker in the musical community, this being regarded as useful only for dialogue, If a large video monitor or projection is present, it can be difficult to get a speaker in a good front-centre loudspeaker position. If a domestic 5.1 system is used, the front-centre loudspeaker may be different from the others and at a different height, neither of which is ideal.Dolby or DTS 6.1 systems have become reasonably common in the last few years, particularly as decoders for them and enough amplifiers to drive six speakers are built into many domestic ‘receiver/amplifiers’. Loudspeaker layout is usually identical to 5.1 systems, with an extra centre-rear speaker. Again, for ambisonics it would be better if the speakers were moved to form a hexagon, with the speakers 60 degrees apart. Similar considerations to those for 5.1 systems apply.

7.1 systems have two major variations. Cinema systems have five speakers, with left-left, left-centre, centre, right-centre, right-right speakers behind the screen, and a left-surround and right-surround channel, usually with a number of distributed speakers. Domestic systems have no clear specification, although manufacturers seem to agree generally on speakers at 0 degrees, plus and minus 30 degrees, plus and minus 110 degrees from front centre, with a further left-rear and right-rear pair of speakers at indefinite angles.

All speakers should ideally be identical and full-range. They should at least be tonally compatible, preferably from the same manufacturer and with similar components. The output level of each speaker should be the same, and 3DAudioScape allows attenuation of speakers to aid this. It is best achieved using pink noise and a sound level meter, backed up by the evidence of your own ears.

All speakers should be at the same distance from the centre point, or the feed to nearer speakers should be delayed to make them appear to be at the same distance from the centre. The decoder in 3DAudioScape contains an automatic delay compensation process, and speaker feeds can be delayed to match the acoustic delay of the speaker most distant from the central point.

No provision for tonal adjustment of loudspeaker signals is currently provided within 3DAudioScape. This could be achieved by the use of an external mixer, or equalisers, though use of such equalisers is not without potential problems.

A mono sub-woofer system should be sent a signal that is a sum of all the other loudspeaker signals (e.g. by using an auxiliary send bus on a mixing desk), or a separate loudspeaker output from 3DAudioScape. The other loudspeakers should not have a high-pass filter except for protection of the drivers, and the sub-woofer signal should be low-pass filtered at a suitable frequency. Two or more sub-woofer systems should be sent signals derived from suitable mixes of other loudspeaker signals, or dedicated loudspeaker outputs from 3DAudioScape, and should be placed appropriately. With domestic systems suitable signals can be derived using the ‘Bass Management’ controls of a domestic ‘Surround Receiver’.

Some acoustic treatment of the listening space, to at least diminish somewhat the most dominant effect of a room on sounds within it, is highly desirable. This is, of course, time consuming, expensive, and has consequences on visual interior design. Sound studios and theatrical performance spaces will tend to have some acoustic design, or the facilities to alter the acoustic properties, even if only heavy drapes and the means to hang them. A well known and designed “listening room’ is a good start. Again, don’t be discouraged. Start with what you can get together. If you like what you hear you’ll think of ways to improve it.

Ambisonic Equivalent Panning

“Ambisonic Equivalent Panning” was formulated by Martin Neukom and Jan Schacher in 2007. It combines encoding and decoding into one process for each sound source. It allows the use of higher order ambisonics in this process, and this can be different for different sound sources. Perhaps unexpectedly this offers a higher level of efficiency than an “encode to B-Format, mix, decode approach”. More sources can then be handled, or less latency required.

A disadvantage is that the processing has to be adjusted for each different loudspeaker array, and that there is no recordable B-Format signal that could be decoded for any array with a software or firmware decoder.