Multichannel Audio Technologies: Lecture 16

Introduction to Ambisonics

Ambisonics can be considered as an extension of Blumlein’s stereophonic principles to full three dimensional imaging. From the beginning, it was designed as a surround sound system that would overcome the major problems of the so-called ‘quadraphonic’ systems that were its predecessors -- the main problem being that they simply didn't work very well. Research rapidly indicated, however, that in addition to providing full surround sound in an encode/decode environment (where the original recording is encoded into a stereo/mono-compatible form for transmission and later decoded by the listener into multiple speaker feeds), Ambisonics could also offer a significant ‘super stereo’ capability without decoding.

Ambisonics was the brainchild of a small group of British academics, notably Michael Gerzon of the Mathematical Institute in Oxford, and Professor P B Fellgett of the University of Reading. Gerzon (pictured left messing with an early Ambisonics Soundfield microphone) was an absolute genius in the field of mathematics and recording technologies, and was pioneer not only in Ambisonics, but also in surround matrixing, lossless compression algorithms, and dither (and it is a little known fact that he was a major contributor in the design of the Waves plugins).

Gerzon and co. built on the astonishing work on stereo recording and reproduction performed by Britain’s early audio genius, Alan Dower Blumlein. Blumlein was working on stereo recording and disc-cutting in the Twenties, and as well as developing the stereo cutting system introduced over 30 years later for microgroove stereo LPs, he also invented what is one of the simplest and most accurate of all stereo recording techniques: M-S coincident pair recording.

Ambisonics Background:

We have shown how pair-wise panning for images to the side (lateral imaging) does not work, and this is one of the reasons why the quadraphonic system was such a massive failure. Ambisonics however, is a system that does not use pair-wise panning, but rather has it’s origins in the coincident microphone techniques outlined by Blumlein.
Blumlein’s Definition of Stereo

Since our hearing mechanism relies heavily on timing information, Dr Harvey Fletcher thought it reasonable to use microphones to capture similar timing differences, and that is exactly what the spaced microphone system does. However, Blumlein (shown left) developed coincident techniques to overcome the inherent deficiencies (as he saw them) of the spaced microphone systems being developed in America.

When sound is replayed over loudspeakers, **both ears hear both speakers**, so we actually receive a very complex pattern of timing differences, involving the real timing differences from each speaker to both ears, plus the recorded timing differences from the microphones. This arrangement tends to produce rather vague positional information, and if the two channels are combined to produce a mono signal, comb-filtering effects can often be heard.

Blumlein demonstrated that by using only the amplitude differences between the two loudspeakers, it was possible to fool the human hearing system into translating these into perceived timing differences, and hence stable and accurate image positions. We all take this entirely for granted now, and are quite happy with the notion that moving a pan-pot or balance control to alter the relative amplitudes of a signal in the two channels will alter its position in the stereo image in an entirely predictable and repeatable way.

This process is used every day to create artificial stereo images from multi-miked recordings, but contrary to popular belief, the level difference between the two channels which is necessary to move a sound image all the way to one loudspeaker is not very much. Typically, a 12 to 16dB difference between channels is sufficient to produce a full left or right image, and about 6dB will produce a half-left or right image -- although the exact figures vary with individual listeners, the monitoring equipment and the listening environment.

To create stereo images directly from real life, Blumlein wanted to develop a microphone technique which captured level differences between the two channels, but no timing differences. To avoid timing differences, the two microphones must be placed as close together as is physically possible -- hence the term ‘Coincident Stereo’. The normal technique is to place the capsule of one microphone immediately above the other, so that they are coincident in the horizontal plane, which is the dimension from which we are trying to recreate image positions. Amplitude differences between the two channels are created through the microphone's own polar patterns, making them more or less sensitive to sounds from various directions. The choice of polar pattern is the main tool we have for governing the nature of the recorded sound stage.

In the Blumlein setup shown below, we utilize an omnidirectional microphone with a leftward facing ‘Figure of 8’ microphone. This is commonly known as the Mid-Side configuration, (and variations exist using a forward facing cardioid or figure of 8 instead of an omnidirectional microphone).
A Blumlein coincident pair—an omni mic crossed with a figure-of-eight pointing left.

In terms of the stereophonic reproduction, instead of left and right channels, we have ‘sum’ and ‘difference’ channels. Let's say we take that omni Mid mic and add it to the Side mic - just bus them to the same bus on the console. Well, if we look at the polar plots of the two patterns something interesting happens. (Remember the figure 8 is aimed to the left and is positive on that side and negative on the right.)

As we add the Mid to the Side around the left side of the graph, the signals add together and draw a nice round polar pattern. At 0° the Mid microphone is full power, but the Side is null, at 270° the Side is full, and at 180° the side is null again. But around the right side
of the microphone the Side mic cancels the Mid, so you get much less signal on the right side.

In fact, adding the two microphone directionalities results in an equivalent virtual cardioid microphone aimed 90° to the left of center. Thus when we add the two microphone signals together we get the feed for the left channel of our stereo setup.

Likewise if we subtract the mid and side signals we get the feed for the right channel of our stereo setup. We accomplish subtraction on our mixer by inverting the signal from the side microphone and adding it to the mid signal. Again the resulting microphone directionality is a cardioid microphone facing right.

The question arises at this point as to why we should go to all this effort and why don’t we just use XY cardioids instead? It is true that we have created an XY pair, but as we increase or decrease the volume of the Mid microphone in the mixdown, we can change the microphone directionalities i.e. Since the Mid microphone is an omni pattern, then when you adjust Mid vs. Side volume the patterns shift from cardioid when the microphones are equal, to wide cardioid when there is more Mid, to hypercardioid when there is more Side.

**Ambisonics: Three-Dimensional Stereo**

Ambisonics can on one level be viewed as the aforementioned Blumlein stereo system, extended into three dimensions. Three? Yes, Ambisonics is capable of encoding sound sources from any direction in space, including vertically. The technique employed in an Ambisonic microphone is to use the equivalent of a single omnidirectional capsule plus three figure-of-eight capsules: one pointing left-right, one front-back, and the other up-down as shown below:

![Ambisonics Diagram](image)

If a 'traditional' Blumlein M-S coincident pair gives you two signals which need to be decoded to derive the left and right speaker feeds, it's fairly obvious that a three-
dimensional Blumlein system will give you more of the same. In fact, the ‘studio format’ for Ambisonics, generally known as B-Format, is exactly this: a mono (sum) signal from the omnidirectional component (Left + Right + Front + Back + Up + Down), known as the 'W' component, plus three difference signals: Front - Back (known as the ‘X’ component), Left - Right (the ‘Y’ component), and Up - Down (the ‘Z’ component). Notice that only four channels are needed to encode not only surround information, but also height (Ambisonics with height is also called ‘Periphony’—“sound around the edge”. This phrase was coined by Michael Gerzon in his classic AES paper ‘Periphony – With-Height Sound Reproduction’).

To explain how this configuration works, imagine that we have placed a single, omnidirectional microphone in front of an orchestra. The recording made with it will be mono, but it will contain all the acoustical information about the sound of the orchestra, and the sound of the hall, at the mike's position. The only thing missing is directionality. This can be obtained with three added signals that specify front/back, left/right and up/down positioning. In Ambisonics, these three signals are called X, Y and Z velocity components, respectively; the omni (mono) component is called W.

So, we wind up with four signals (omni, plus front/back, left/right and up/down directionality). Quite different from conventional Quadraphonics or Dolby Surround where the four signals are the speaker feeds! So how do we get the signals we have obtained to feed the speakers?

This is handled by an Ambisonic decoder. (It really should be called a “directional processor”; “decoder” has the unfortunate connotation of matrixed Dolby or Quadraphonics.) We could simply say that the four signals are added and subtracted to produce the speaker signals, but it isn’t hard to understand the process on a slightly deeper level.

Imagine a playback system in which all four Ambisonic signals are available to us. We’ve placed eight speakers in the room, four in the corners on the floor and four in the corners at the ceiling. What does the decoder send to the speakers?
The omni (W) signal goes to all eight speakers with the same amplitude and phase. The front/back (X) signal is sent to the four front speakers in-phase with the omni signal and out-of-phase with the omni to the four rear speakers. Likewise, the left/right (Y) directional signal is sent in-phase to the four left speakers and out-of-phase to the right four. (If you can’t figure out what happens with the up/down (Z) component, then you haven’t been paying attention.)

In practice, the X, Y and Z components are adjusted in relative level to compensate for the exact position and layout of the speakers. This is done with a control on the front of the decoder, after measuring the length, width and height of the speaker array.

Is this really an improvement over Quadrrophonic? Consider that Ambisonics satisfies all the following directional hearing characteristics:

1. Correct positioning of sound sources at all positions around the listener, both for low and high frequencies.
2. Zero phasiness. (Phasiness is that strange effect you hear when your speakers are out of phase. The image is smeared and vague, and there may be a “pressure in the ears” feeling.
3. Correct imaging no matter which way the listener turns.

The necessary W, X, Y and Z signals can be generated by the Soundfield Microphone for live recordings, or by means of the simplest forms of signal processing in the studio. There is no limitation in combining studio effects with live sound. Almost anything the producer imagines can be created. He can make sounds come from anywhere around the listener and have any apparent size; he can rotate the sound field smoothly in any direction, make it come closer or go further away, or even cause it to pass through the listener’s head and come out the other side! Needless to say, these effects are difficult or impossible to achieve with quadraphonic systems. For Ambisonics, they are simple.

The system we’ve been describing normally requires three transmission channels for surround sound, plus a fourth if height reproduction is desired.

**Basic Ambisonic Technology**

Since Ambisonics was envisaged as a total solution to surround sound recording and reproduction, there are several Ambisonic ‘formats’ which relate to particular chains throughout the recording and reproduction workflow. The A-format is describes microphone pickup, the B-format for studio equipment and processing, the C-format for transmission, and the D-format for decoding.

Reproduction requires four or more loudspeakers depending on whether it is pantophonic or periphonic, size of area etc. Practical minimums are four for horizontal only, eight if you require height as well. The important thing to note is that there is no need to
consider the actual details of the reproduction system when doing the original recording or synthesis, since if the B format specifications are followed and suitable loudspeaker/decoder setups are used then all will be well. In all other respects the two parts of the system, encoding and decoding, are completely separate.

**A-Format Microphone Pickup**

In most Ambisonic microphones, such as the **Calrec Soundfield microphone** and its successors, the four polar diagrams are simulated by a tetrahedral array of capsules (shown below). This has the benefit of also allowing them to be electronically corrected for true coincidence -- because the closer together the capsules are, the more accurate the localisation is, particularly at high frequency. This is one of the reasons that the Soundfield microphones are excellent M-S stereo mics as well as having their (somewhat under-exploited) surround sound benefits. A soundfield microphone is just that--a device for capturing all the sounds in an environment so that they can be stored in such a way as to make it possible to regenerate in the listening environment the original pattern of sound waves falling on the microphone.

![Soundfield microphone](image)

Soon after the introduction of the Soundfield microphone, developments began to be made in the field of simulating soundfields as well as capturing them. The result is that today there are comprehensive mixing systems that allow individual multitrack signals to be panpotted into an Ambisonic picture -an area we'll look at later.

**Producing B-Format Signals:**

The outputs of the sound field microphone give us the omnidirectional pressure component and the directional information of the source. Mono sources which are not
recorded with a sound field microphone can be rendered to B-Format using a specific set of encoding equations.

Decoder designs are predicated on the basis that sounds being positioned in Ambisonic B-format are placed on the surface of or within a notional unit sphere where the maximum radius a sound may be placed at can be thought of as 1 - this is frequently referred to as the 'Unit Sphere'. If the sound is moved outside this sphere the directional information will not be decoded correctly and sounds will tend to pull to the nearest speaker.

Initially all transformations will place sounds on the surface of the unit sphere:

\[
\begin{align*}
x &= \cos A \cdot \cos B \\
y &= \sin A \cdot \cos B
\end{align*}
\]

Graphical representation of the Unit Sphere, where A = the angle of rotation , B = the angle of elevation.

If a monophonic signal is to be placed on the surface of a unit sphere, then its coordinates will be, referenced to centre front:

\[
\begin{align*}
x &= \cos A \cdot \cos B \\
y &= \sin A \cdot \cos B
\end{align*}
\]
\[ z = \sin B \]

These coordinates directly relate to the B-format signal levels thus;

\[ X = \text{input signal} \times \cos A \times \cos B \]
\[ Y = \text{input signal} \times \sin A \times \cos B \]
\[ Z = \text{input signal} \times \sin B \]
\[ W = \text{input signal} \times 0.707 \]

The 0.707 multiplier on W is there as a result of engineering considerations related to getting a more even distribution of signal levels within the four channels when taking live sound from a Soundfield microphone. A is the anticlockwise angle of rotation from the centre format and B is the angle of elevation from the horizontal plane. These multiplying coefficients, \( (\cos A \times \cos B \text{ etc.} ) \), will position the monophonic sound anywhere on the surface of the soundfield, producing the B-format encoded output signals.

**Decoding Ambisonics**

On decoding, each individual speaker is fed a combination of the B-format signals corresponding to its position with respect to the center of the array. For instance, for a square (horizontal only) array of four speakers, arranged left front (LF), right front (RF), left back (LB), and right back (RB), the signals are:

\[
\begin{align*}
LF &= W + 0.707(X + Y), \\
RF &= W + 0.707(X - Y), \\
LB &= W + 0.707(-X + Y), \\
RB &= W + 0.707(-X - Y).
\end{align*}
\]

For a cubic array, the signals for the four planar corners in the "up" (U) and "down" (D) planes are:

\[
\begin{align*}
LFU &= W + 0.707(X + Y + Z), \\
RFU &= W + 0.707(X - Y + Z), \\
LBU &= W + 0.707(-X + Y + Z), \\
RBU &= W + 0.707(-X - Y + Z), \\
LFD &= W + 0.707(X + Y - Z), \\
RFD &= W + 0.707(X - Y - Z), \\
LBD &= W + 0.707(-X + Y - Z), \\
RBD &= W + 0.707(-X - Y - Z).
\end{align*}
\]

The directivity factor of 0.707 in the equations above results in a cardioid source directional response for each loudspeaker. Remember how in the mid-side technique we changed the directional response of the ‘virtual microphone’ by changing the ‘side’ signal level? This is exactly what we are doing here. Except that this directionality now refers to
the loudspeaker feeds. In other words, for each loudspeaker, we have a corresponding **virtual** microphone!

(a) Diagram showing amplitude/phase and virtual microphone response for 1 loudspeaker  
(b) Same as (a) but with increased directivity factor.

Where the intended listening area is significantly smaller than the speaker array, a more hypercardioid shape can be employed by increasing the directivity factor, which results in improved imaging for centrally located listeners.