Ambisonics. Part two: Studio techniques

By Michael Gerzon

*Conventional methods of 'quadraphonic' reproduction may not convey the original sound field to best advantage. The studio techniques for ambisonics and its applications are discussed, in the light of both accepted 'quadraphonic' techniques, and the requirements of further accurate reproduction of the sound field. Areas of compatibility and of disparity are discussed, to be read in the light of Peter Fellgett's system description (*Studio Sound, *Vol. 17, pp. 20-22, 40 (August 1975).*

The main aim in the development of NRDC ambisonics technology has been to record, to convey to the consumer, and to reproduce an accurate and repeatable surround sound directional effect. It is now well known, both from controlled experiments and from the experience of recording engineers and producers, that existing surround sound approaches (including the four channel 'discrete' approach) give extremely poor image stability for all positions except the four corners, even under ideal conditions. With 'discrete' techniques, the front stage suffers from the 'hole in the middle' effect, and the sides are virtually unusable in any less-than-ideal situation (eg when the listener is not at the centre of the speaker layout, or when the layout is non-square).

The impracticality of existing approaches has led to a careful study of each stage of the multi-channel recording and reproduction of a sound field. [1 2 3 4]. The aim of a surround sound system is to reproduce at the listener's ear accurately, reliably and repeatably, the directional sound field created in the studio either by a sound field encoding microphone array,[1] or by artificial directionality encoding devices (pan-pots) or artificial surround-reverberation devices. This aim contrasts with the aim[2] of quadraphonic systems, which is to duplicate in the home the defects of a pair-wise mixed mastertape. Without accuracy and repeatability of directional effect, the recording producer's work will not be heard correctly by the domestic listener, and artistic communication will be compromised.

In order to maintain accuracy of effect, the process of encoding the sound field on to a mastertape must be accurately specified, and the specification accurately followed in either the microphone arrays or the pan pots. Existing 'pairwise' pan pots do not give a satisfactory encoding specification[3]. Similarly, most microphone

Reproduced from *Studio Sound*, Vol. 17, pp 24-26, 28, 40 (August 1975), by permission of IPC Media Ltd, publishers of *Hi-Fi News* (www.hifinews.co.uk)

clusters use microphones with poorly defined polar diagrams that are frequency-dependent and spaced apart, and so fail to satisfy a reasonable encoding specification. Conventional mastertape encoding assigns the four speaker feed signals to four tracks of a tape, but it is obvious that this is a sub-optimal procedure, even with correctly designed microphones and pan-pots. In order to recreate the illusion of a given directional field, it is obvious that the speaker feed signals will not be the same for a rectangular speaker layout as for a square one. Thus if we are to accommodate a variety of shapes of speaker layout appropriate to differently shaped listening rooms, we need to use a decoder to derive the speaker feed signals appropriate to the layout used.

As we shall illustrate later in this article, he optimal decoder even for a square speaker layout is a rather complex frequency-dependent matrix, and so we must face the fact that the mastertape merely encodes the sound field information, and not speaker feed signals.

Studies[3] in the design of decoders for four-speaker rectangle layouts show that three channels of information are optimal for accurate image localisation. The addition of a fourth channel always degrades image localisation quality (eg by giving a 'hole in the middle' effect). Thus, the basic horizontal encoding specification has to be three-channel. Most studio equipment used for surround sound mastering has four available channels, and so the problem arises of how to use this fourth channel. Since no useful extra horizontal information may be encoded in the forth channel, it may most fruitfully be used for encoding height or elevation information.

It is not suggested that periphonic (ie with-height) reproduction will be a serious commercial proposition at the present time, but periphonic technology is now understood, and not much more complex than horizontal-only technology. The adoption of a periphonic standard at this stage will prevent the premature obsolescence of valuable mastertapes in ten or fifteen years when periphony becomes commercial, and meanwhile will permit producers to gain periphonic experience without premature commercial pressures. It is also wise to ensure that existing media can change over smoothly to a periphonic standard in the future, so as to prevent a repetition of the chaos caused by quadraphonics. At the present time, the height information gives useful additional mixdown flexibility for stereo or horizontal surround effect. If height is not required, it may be omitted, as in most of the equipment described subsequently.

For studio use, there are two four channel signal formats, termed A-format and B-format. A-format consists of four channels L_B, L_F, R_F, R_B compatible with existing 'discrete' practice for the four corner positions. Technically, the A-format signals may be described for horizontal only sounds as the outputs of four hypercardioids each having nulls 120° off-axis (or their panpot equivalent signals) pointing in the four corner directions. When height is included the A-format signals are the outputs of four hypercardioids with nulls 114.1° off axis pointing respectively 35.3° below, above, below and above the four corner directions (ie towards regular tetrahedral axes[5]).

The second format is B-format, consisting of the four signals X, W, Y, Z, where X is a forward facing figure of eight signal with frontal gain $\sqrt{2}$, W is an omnidirectional signal of gain 1, Y is a sideways-facing figure-of-eight signal with leftward gain $\sqrt{2}$, and Z is an upward figure-of-eight signal with upward gain $\sqrt{2}$.

The circuit to convert A-format to B-format, known as an 'AB module' is shown in **fig. 1**, and performs the following matrixing:

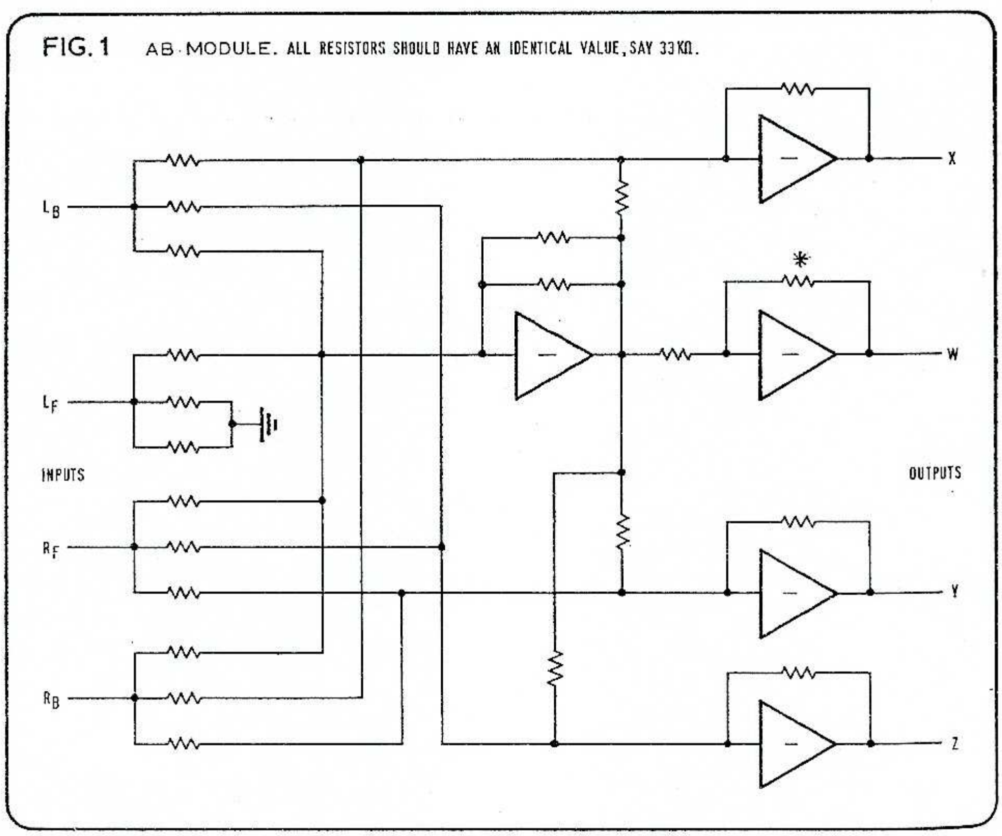
$$X = \tfrac{1}{2}(-Lb+Lf+Rf-Rb)$$
$$W=\tfrac{1}{2}(Lb+Lf+Rf+Rb)$$
$$Y=\tfrac{1}{2}(Lb+Lf-Rf-Rb)$$
$$Z=\tfrac{1}{2}(-Lb+Lf-Rf+Rb)$$

It is important to note that for horizontal only signals, the signal Z is zero, and so may be omitted, giving a three channel B-format signal in the horizontal case.

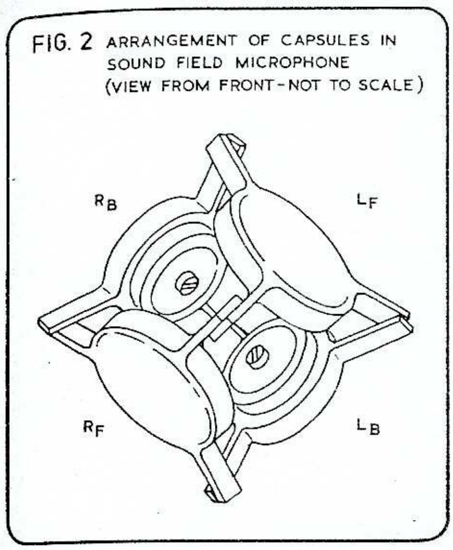
The important thing about AB modules is that exactly the same circuit converts back from B-format to A-format. Thus, if one puts X, W, Y, Z into their respective inputs, L_B, L_F, R_F, R_B comes out. If one feeds a conventional 'discrete' or pairwise mixed signal into an AB module and then discards the Z output, then the X,W,Y signals are correctly encoded B-format signals for the four corner positions, with slight deviations from the correct encoding elsewhere. (This deviation is one of the causes of conventional discrete reproduction giving poor non-corner images – the B-format signal will not make this defect worse).

There are two main ways of producing correct B-format signals. The first is the Calrec sound field microphone (which is still

Ambisonics. Part two: Studio techniques



undergoing evaluation and development). This has been developed by the present writer for the NRDC and uses (see **fig. 2**) a tetrahedral array of cardioids to

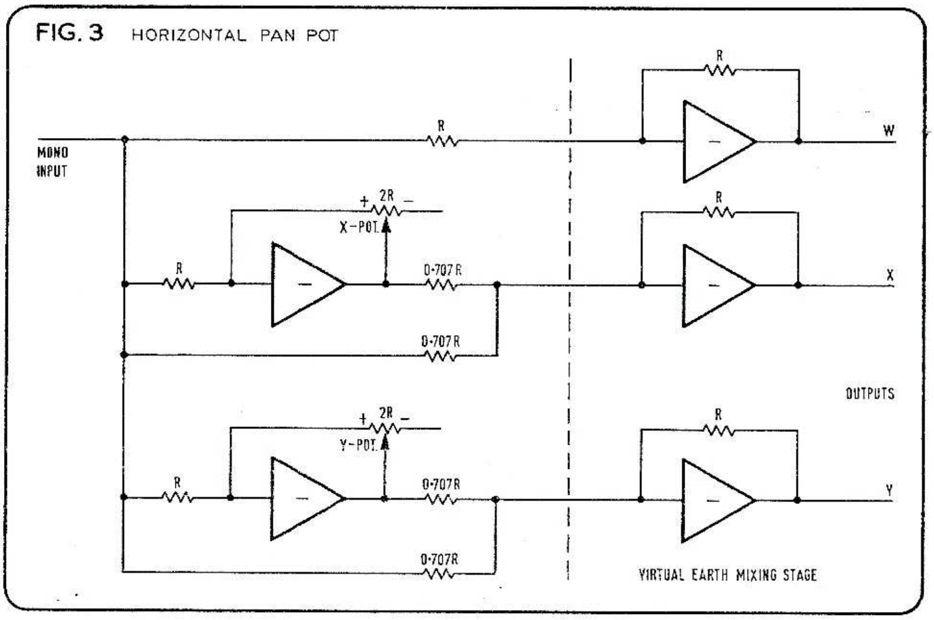


feed a frequency-dependent matrix circuit. The matrix circuit fulfils the dual function of converting to B-format and of providing electronic compensation for the spacing of the capsules, so as to give outputs that are effectively coincident and satisfy B-format encoding accurately[1]. The result is to give B-format outputs that are characterised by accurate and precisely coincident polar diagrams up to around 7.5 kHz (as compared to less than 1.5 kHz for the best existing microphone arrays). The behaviour above 7.5 kHz is arranged to be subjectively smooth, although deviating from the ideal.

In this way, the directional properties of a sound field may be captured with the minimum possible departure from the objective ideal. The microphone system should not be regarded as four microphones on a tetrahedron, but as a complete sound field transducing system. The tetrahedral configuration is purely a matter of design convenience[1] and has nothing to do with the desired form of B-format sound field encoding.

A second method of producing B-format signals it to use a panpot. **Fig. 3** 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' of **fig**. **3** are the potentiometers that respond to the 'up-down' and 'left-right' motions respectively of the joystick (see **fig.4**). For correct results it is vital that the travel of the joystick be restricted by a mask or cut-out to that range of X-pot and Y-pot resistances in **fig. 4** such that $x^2 + y^2 \leq 1$. In other words, the corner travel must be restricted to ±0.707 of the way from the centre to the end of the pot tracks, although the full range of each pot may be covered when the other is centred.

It is also possible to convert existing pairwise pan pots to give optimal ambisonic encoding. This is clearly a worthwhile option for use with existing equipment. The modifications involve fitting a mask to the joystick control to limit its travel, and a matrix circuit positioned after the four output channels of the mixer. The matrix circuit is exactly the same as an AB module (**fig 1**), except that the gain of the W output channel is reduced to 0.707 (by reducing the resistor marked * in **fig. 1** to one equal to 0.707 of the value of the other resistors), and the Z output is not used. This gives a B format output. The masking of the joystick controls is necessary to prevent undue exaggeration of directionality of the corners. The mask for a joystick control normally travelling in a square aperture would be as illustrated in **fig 5**. The side positions remain unaffected by the mask, but the corner positions are masked so that the corner-most joystick positions in the mask give equal outputs on the X, W and Y outputs. Because most joystick pairwise pan pots are not well designed as regards constancy of sound level with direction, no guarantee can be given that the ambisonic modification will be good in this respect either.



The pan pots described give full 'interior' effects as well as a 360° azimuth coverage with accurate encoding according to B-format specifications. A similar design of about twice the complexity prior to the mixing stage, using an additional slider pot for elevation gives full-sphere periphonic encoding. Other devices of a similar nature allow the full rotation of a whole encoded sound field ('waltz' control), a facility not possible with discrete approach. When sounds are 'circled' with a B-format pan pot, the motion is smooth rather than the jerky jumping from speaker to speaker given by existing pan pots. A design is also available for a 'width' control that alters the width of the front of a sound field relative to the back without destroying the correct encoding specification.

Having obtained a horizontal or periphonic B-format signal, either it can be converted to A-format (by the AB-module of **fig. 1**) to go through existing quadraphonic equipment, or it can be recorded in

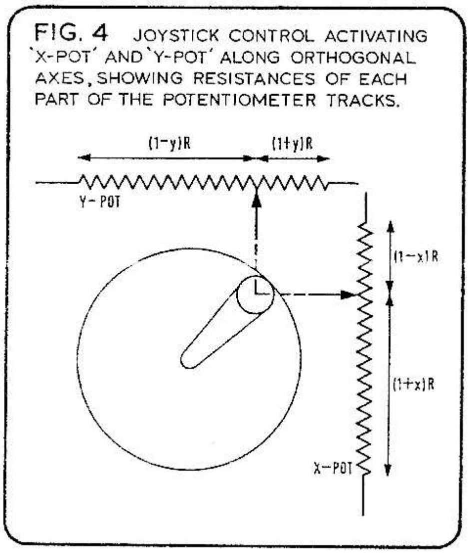
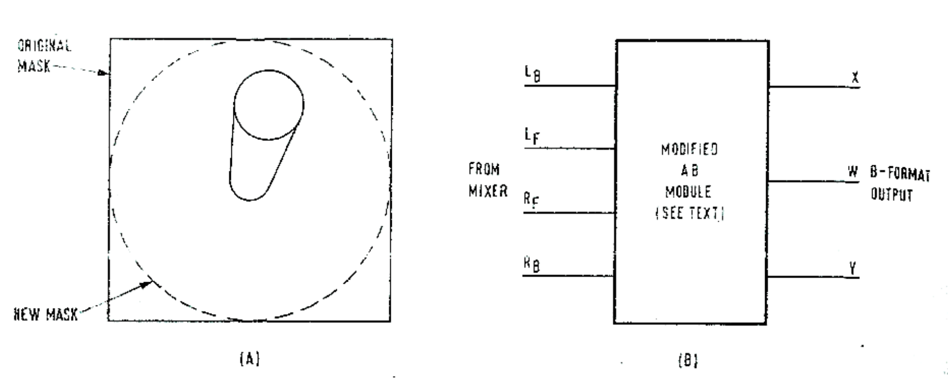


FIG. 5 TYPICAL MASKING OF PAIRWISE PAN POT (A) AND THE PROCESSING AT THE MIXER OUTPUT (B)



B-format. There are considerable advantages gained by staying in B-format in the tape recording stage. If one records in A-format then the effects of noise reduction systems are quite audible when the signal is played back. Existing stereo covers only a 60° stage and the small 'pumping' effects inevitable even with a well adjusted noise reduction system do not cause noticeable movements in the stereo image. Similarly, while conventional discrete material covers 360° its very poor image quality means that pumping effects are not noticed. However, once the image sharpness is improved by ambisonic encoding, the varying image shifts (which subjectively seem to be around 15°) start becoming comparatively objectionable. It is found that recording in B-format renders the signal much less susceptible to all forms of image degradation whether due to channel imbalances, phase errors, or noise reduction pumping. In fact, it seems likely that B-format recording should reduce signal degradation significantly even for conventional 'discrete' pan pot recordings.

Ambisonic B-format signals may be converted to A-format by an AB module, and may be used, if necessary, for feeding existing 'quadraphonic' systems via existing commercial encoding equipment (with the exception of the SQ system*), provided that the Z signal of the B-format is omitted. While the results should generally be better than with existing mastertape encoding methods, they will not be as good as they might be, except for the UD-4 system which will give correct TMX encoding.

* However, an SQ encoder may be used if it is set to 'interior' encoding mode. The results with SQ cannot be optimal.

Table 1 Gains of shelf filters in three channel decoder of **fig. 6**.

	Shelf filter 1	Shelf filter 2
low frequencies	0 dB	0 dB
high frequencies	+1.76 dB	-1.25 dB

For each of the following systems: CD4, UD4, RM and the BBC matrix systems, there is an optimal encoder design available working straight from B-format. It is understood that the BBC has evolved an encoding technique similar to B-format. The use of an optimal encoder ensures that the consumer-encoded format (C-format) accurately follows the correct specification. At the time of writing, no existing system is considered by broadcasting organisations and independent record companies to fulfil the necessary compatibility requirements adequately along with good four speaker results, and it is expected that industry discussions will be held regarding the choice of a generally acceptable system. However, ambisonic studio technology is compatible with all systems capable of good localisation for all directions from four speakers.

It is, of course, necessary to monitor the results of an encoded, A-, B- or C-format signals, and no existing decoder design is capable of optimal results. It must be remembered that it is no more desirable that the four speakers should be heard as direct sound sources than one would wish to see the individual phosphor dots on a colour tv screen. The speakers are purely a means of feeding information into the room to create a convincing (or otherwise) illusion of sounds from all directions around the listener. If the shape of the speaker layout deviates from a perfect square, the image stability and directional effect from the loudspeakers would alter unless the decoder is modified to compensate for a non-square speaker layout.

Other problems in designing decoders arise from the fact that the ears localise sounds by different mechanisms at low frequencies (< 700 Hz) and high (> 700 Hz). This means that the optimum design for a decoder is different at low and high frequencies. Moreover, when speakers are fed with signals from a matrix with no frequency variation, it is found that the tone quality of reproduction is very coloured and 'thumpy' in the bass (below 350 Hz). This is because at low frequencies the intensity at the listener is the sum of the pressures due to the four speakers, while at high frequencies it is the sum of the energies.

A design of three channel horizontal studio decoder, working off a B-format input is shown in **fig. 6**. This is intended to feed a rectangular speaker layout, and the 'layout control' adjusts the decoder to compensate for the actual rectangular shape used. The two types of 'shelf filter' used have gains as in Table 1 at low and high frequencies, and are 'phase compensated' to have identical phase response. The transition between low and high frequency gains in the step filters is gradual (using simple RC type circuits) to avoid coloration, and is centred on 350 Hz approximately.

It is found that a correctly encoded B-format signal fed to the three channel studio monitor gives stable and sharp images even at the sides of the listener and for listeners well away from the centre of the listening area. The four loudspeakers used should be matched both in frequency and phase response, and should be reasonably 'omnidirectional' in their polar diagrams over $\pm 45^{o}$ off their respective axes. The layout control requires careful adjustment, but once set need not be changed.

The three channel/four speaker studio decoder fed by B-format gives what we believe is the most accurate reproduction of sounds from any desired direction around the listener possible with existing technology via four speakers. It is not perfect. In particular, it was predicted (in advance of construction) by a new 'bispectral' model for human hearing that sound waveforms with a very high degree of asymmetry would still tend to be 'pulled' to the nearest speaker position, and this is quite noticeable on clapping. Experiment and theory agree here, and theory shows that there would be no ways of overcoming this fault other than going to a five speaker decoder. Many critical listeners would possibly regard this as 'hair-splitting' by comparison with many existing four speaker systems, such as 'discrete' encoding.

It is not claimed that there is any decoder capable of giving really accurate decoding in a very large room or auditorium via four speakers. Ambisonics is essentially designed for domestic or studio listening conditions, although results in a large auditorium can be quite reasonable.

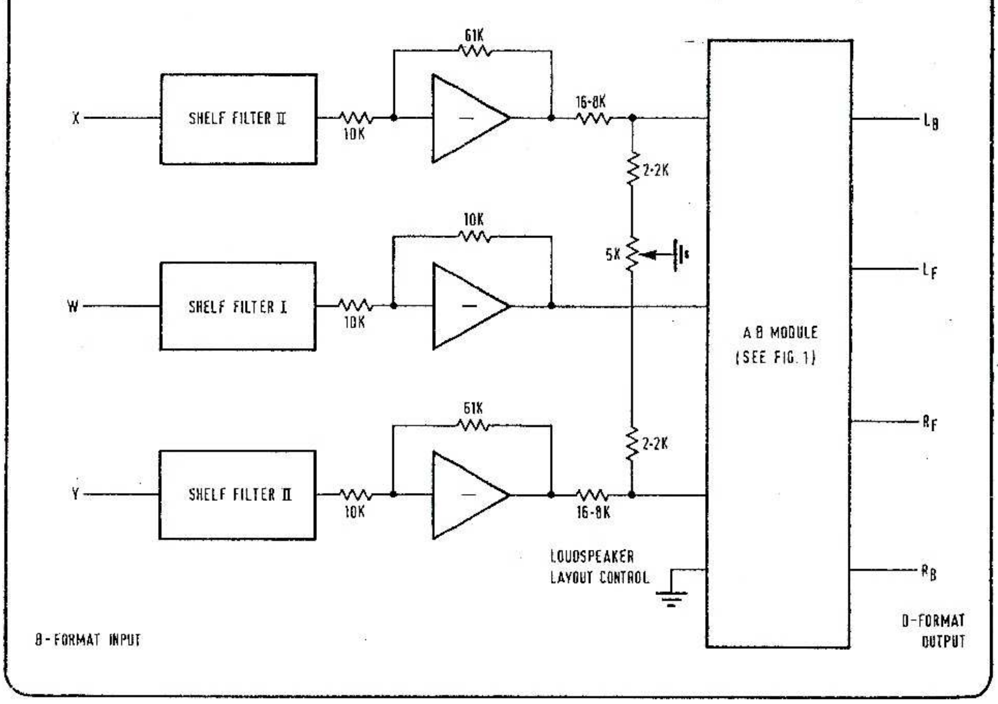


FIG. 6 THREE-CHANNEL, FOUR LOUDSPEAKER DECODER FOR HORIZONTAL STUDIO MONITORING. THE SHELF FILTERS HAVE MATCHED PHASE RESPONSES AND GAINS AT LOW AND HIGH FREQUENCIES AS IN TABLE ONE. RESISTOR VALUES GIVEN ASSUME AN 11KΩ INPUT IMPEDANCE FOR THE AB MODULE.

Table 2 Gains of shelf filters in two channel decoder of fig. 7.

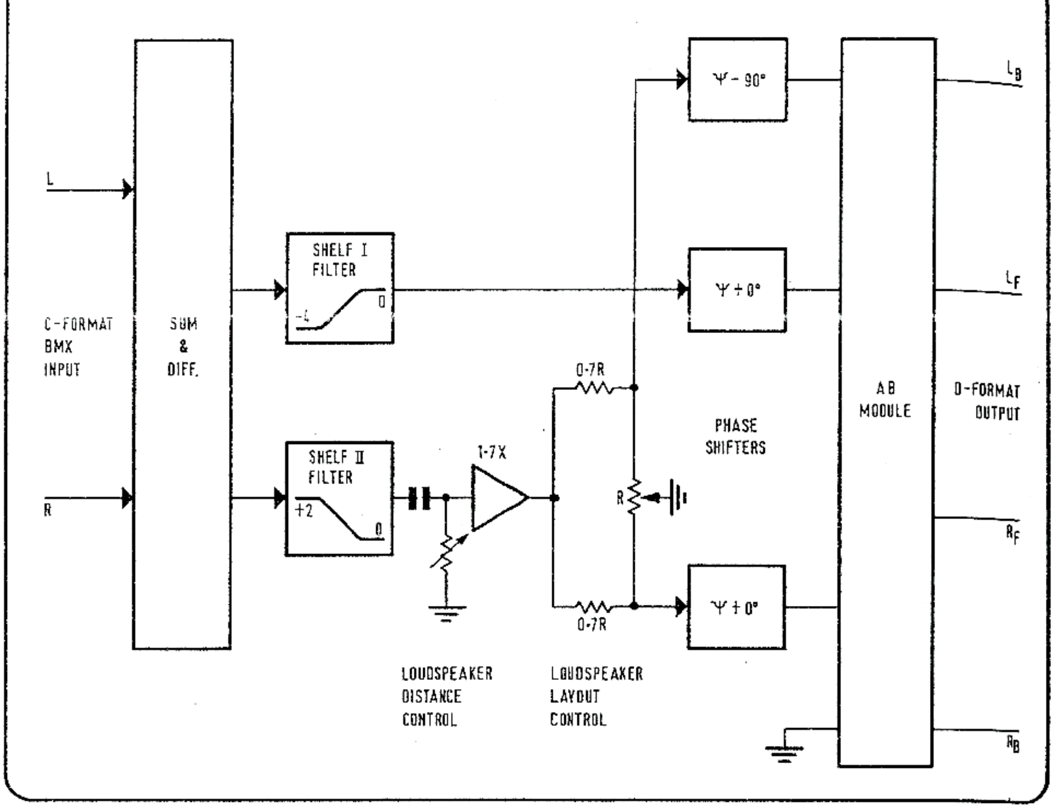
	Shelf filter 1	Shelf filter 2
low frequencies	-3.98 dB	+2.04 dB
high frequencies	0 dB	0 dB

Domestic decoders have been designed according to a range of psychoacoustic theories[3] and at various levels of cost and complexity for all the major existing or proposed surround sound systems other than SQ. It is not possible to design an SQ decoder to satisfy the psycho-acoustic criteria established by the author in ref. 3.

Decoders including layout controls and frequency dependence are rather similar to that of figure 6 and have been designed for two channel C-format decoding, 2½ channel (as used in UD-4) and three channel decoding. As an example, **fig. 7** shows a basic decoder for the BMX system, with phase-matched filters centred on 350 Hz with gains as in Table 2. This is not the most refined version under development; improvements to reduce ‘phasiness’ will be announced shortly.

Other decoders have been designed adjustable for cuboid (box-shaped)

FIG. 7 DECODER FOR BMX ENCODED SIGNALS



with-height speaker layouts for periphonic reproduction. At the present time, this would be confined to experimental and in-studio use, where producers might find it worthwhile to explore the artistic possibilities of full-sphere directional effects well before they are pushed into premature commercial exploitation. Three and four channel C-format encoding has been designed for periphony compatible with existing or proposed horizontal C-format encoding, and it would even be possible to release periphonic material on disc (without announcement) to avoid inventory troubles at a future time.

Ambisonic technology also offers new opportunities for existing mono and stereo recording. The sound field information from a sound field microphone may be recorded in B-format on four channel tape, and be mixed down later to any coincident stereo microphone technique that may be required. Any image width and microphone polar diagrams may be selected off tape, along with any vertical angle of tilt. A control unit to perform these functions in an intuitive and easy to grasp fashion has been designed, and also gives adjustable 'quadraphonic' outputs for four speaker use. Due to their very good polar diagrams in the mid-treble frequency region, the sound field microphone technology also gives stereo with a particularly 'clean' and uncoloured quality of sound.

We have here only been able to skim over a few aspects of ambisonics studio technology. Other devices include an apparatus to convert certain existing types of artificial stereo reverb units to full surround reverb, and devices for compensating for and improving non-ideal microphone techniques that may have been used on any existing surround sound recordings.

Details of these and other devices will be released at a future time.

Ambisonic technology can, of course, be used as part of existing 'quadraphonic' systems (excluding SQ) by treating the A-format horizontal-only signals as if they were 'discrete' signals. Similarly ambisonic decoders may be used with existing quadraphonic systems (again excluding SQ). However, it will be appreciated that the inherent faults of the quadraphonic systems will be apparent in such cases, as the whole system, from microphones or pan pots to decoders and loudspeakers has to be designed correctly for correct results. When parts of the system are incorrect, ambisonic technology will not make the system any worse, and so a compatibility exists between ambisonics and quadraphonics – ie ambisonic material can be used for quadraphonic results, although the converse is not true – just as a poor colour film cannot be made good by good projector optics.
In conclusion, the ambisonic technology developed for the NRDC, and in particular by Professor Peter Fellgett, the present author and John Wright of IMF, is compatible with several existing and proposed encoding systems. It gives enhanced creative possibilities to the producer both by ensuring that what he hears will be substantially passed on to the consumer despite differing loudspeaker layouts and seating positions, and by giving him convincing side localisation and smooth 'circling' effects. In addition to existing 'interior' or 'in the head' effects, new 'waltz' (rotation) and 'width' effects are available as well as the first practical control over with-height periphonic effects, if required.
When required, ambisonic equipment can cope with existing

recordings and equipment, but of course cannot remove the faults in existing material. Besides these 'creative' possibilities, the sound field microphone allows uniquely accurate recording, storage and playback of natural sound fields, with all their attendant advantages. The 'ambient labelling[4]' given by the sound field microphone in particular permits remarkably good sound localisation as well as the ability to separate by ear musical lines that would be completely masked by other lines in a pan pot recording.

References

1. Michael Gerzon 'The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound' Preprint of the 50th Audio Engineering Society Convention, London, 4th March 1975.
2. Michael Gerzon 'What's Wrong with Quadraphonics?' Studio Sound May 1974.
3. Michael Gerzon 'Surround Sound Psychoacoustics' Wireless World, December 1974, pp 483-6
4. P. B. Fellgett, 'Ambisonic Reproduction of Directionality in Surround-Sound Systems'. Nature, vol. 252, pp 534-8. December 13, 1974.
5. Michael Gerzon 'Principles of Quadraphonic Recording Part Two: The Vertical Element'. Studio Sound, September 1970.