1974

Ambisonic reproduction of directionality in surround-sound systems

P. B. Fellgett*

In both the technology and the aesthetics of extending high fidelity reproduction to surround-sound, reproduction of natural ambience is crucial. The 'quadraphonic' attempt to reproduce four stereo-blended tracks, derived from multi-microphone mix-down, cannot provide this. Complete spherical directionality can however be encoded on to a minimum of two audio channels to produce acoustically acceptable surround-sound systems. Limitations are set both by the number of available loudspeakers and by the number of channels.

THE history of representational art has many examples of what people of a particular age and culture were unable to perceive, and therefore to reproduce; for instance the inability of the dynastic Egyptians to perceive that the full-face eye is not seen in a profile figure. The reproduction of sound follows a similar pattern. In the early development of the phonograph and of 'wireless' the listener was so satisfied with the novelty of hearing the human voice or music issue from a machine that he demanded nothing else. Indeed there was often little conscious appreciation that anything was lacking in fidelity of reproduction; how else could Sir Arthur Conan Doyle suppose in *The Adventure of the Mazarin Stone* that Sherlock Holmes could deceive a sophisticated villain into mistaking a 'modern gramophone' of 1927 for real violin playing, even in the next room. As late as the 1930s, many listeners could see little short of perfection in the commercial 'radiogram'.

As the sense of wonder wore off, however, a more critical attitude filtered down from audio engineers to the buying public, and the search began for high fidelity reproduction.

During and immediately after the Second World War, domestic reproduction of sound depended on the 78-r.p.m.

* Address: Department of Engineering and Cybernetics, University of Reading, Whiteknights, Reading RGL 2AL.

shellac disk and a.m. broadcasting. Now with the vinyl disk and f.m. broadcasting, high fidelity reproduction has reached the stage at which the best equipment can make a chamber group sound (to me) like the real thing; but this degree of realism in the reproduction of an orchestra still eludes us. Careful introspection suggests that even this judgement is too lenient, and that we are 'listening-through' major defects which have always been inseparable from reproduced sound, and which therefore we take for granted.

One aspect that is obviously lacking is directionality. Monophonic reproduction gave no explicit directional information, even when reproduced from more than one loudspeaker. In stereophonic reproduction, sound reaches the listener only from the forward direction, whereas in the concert hall the listener is bathed in reverberant sound coming from all directions. In addition, stereo works by feeding common signal-components to a pair of loudspeakers with relative amplitude determined by the position which the illusory sound source is to occupy. It is easy to show that the perception of stereo-imagery is both physically artificial and aesthetically untrue to life; it has to be learnt and some 10% of the population are unable to do this. Moreover, although skilful recording can give some illusion of depth, the acoustic images are essentially strung out along a line joining the two loudspeakers, giving an effect that has been described as a cardboard cut-out orchestra.

Role of directionality

From the antiphonal tradition through to the school of St Mark's and the baroque, directionality played a large and explicit part in western music. In classical and romantic orchestral music this explicit component is largely abandoned, some composers even preferring a completely blended sound, but nevertheless directionality is important implicitly. A composer necessarily writes with subconscious experience of the relationship between the sections of an orchestra, or between the different departments of an organ and the relationship of these to the choir; it has been said that the works of the Bach family cannot be fully appreciated except on an organ constructed on the *Werkprinzip*.

Some conductors pay considerable attention to the disposition of the sections of the orchestra, and these relationships should be heard in reproduction. There is little need, however, to pinpoint individual instruments, which indeed is seldom possible in the concert hall. What is important is that each voice in the musical texture should be labelled with an acoustic ambience characteristic of its place of origin. The ability to follow different voices in the musical texture, despite substantial differences in level, is very much bound up with this labelling, as indeed is the appreciation of musical timbre itself.

All these aspects relate to the acoustical reverberance and ambience of the concert hall. If it is worth spending thousands of pounds on the acoustics of a concert hall, then this should be heard in reproduction.

It follows that the highest fidelity in sound reproduction requires that the directionality of sound should mimic both the direct and reverberant sound of the concert hall. This is not just a matter of a vague splash of delayed echoes, but of a relationship between directionality and time-delay which gives specific information; this information is what we shall mean strictly by ambience. Systems capable of reproducing it will be called 'ambisonic'.

'Quadraphonic' attempt

In the recording industry it is customary to mix-down multi-microphone recordings onto four-track master tape in which signals are often stereo-blended between pairs of channels. The present upsurge of interest in surround-sound was in some measure triggered by engineers and producers playing back such four-track material directly into four amplifiers and loudspeakers distributed approximately in a square near to the corners of the monitor room.

Mixed-down tapes are unfortunately highly contrived and may have only the sketchiest relationship to natural sound. Moreover the assumption was that directionality should be encoded by pair-wise blending between channels associated with adjacent corner positions. This is a sub-optimal method of encoding, making less than full use of the information capacity of the available channels. Further, in order to achieve this coding using directional microphones it would be necessary for the polar response to have the highly improbable form of a half-cycle cosine wave over one sector of 180°, and zero response over the remaining 180°.

There are also severe restrictions in playback. If four loudspeakers are disposed about a listener, at least one pair must subtend an angle of 90° or greater at that listener. It is well established, however, that stereo blending does not work well when the angular subtense of the line joining the speakers exceeds about 60°. Moreover, stereo imagery works best over the 90° sector in front of the listener, worse at the rear, and worse still at the two sides. The attempt to extend the principle of stereo blending from the front wall to the other three sides of the listening room therefore not only retains all the limitations of stereo itself (outlined earlier) but is necessarily rather poor even judged as stereo.

The above approach is what is usually meant by the term 'quadraphony', although usage is by no means uniform. The name itself is a canard in at least two ways. First, it is a Greek–Latin hybrid, and ought to be called quadrisonic. The second is the identification, by implication, of surround-sound with the need to have four channels. In fact, as we shall see, surround-sound can be realised with as few as two channels. The false assumption that surround-sound is synonymous with four-channel then introduces the need to find distinctive names for two- and four-channel surround-sound systems, which were therefore called 'matrixed four-channel' and 'discrete four-channel', respectively. To prevent confusion I shall not use these jargon terms, and words such as four-channel and matrix will be used simply in their ordinary English and technical senses.

More systematic approach

Any system that is to reproduce specific ambience information must embody the following (Fig. 1):

(*a*) transduction of the original sound field, including its directionality and acoustic ambience;

(*b*) encoding the resultant composite signal, including information about directionality, on to the available communication channels;

(*c*) transmitting the resultant information to the listener, including recording and storage of this information where appropriate;

(*d*) decoding the channel-signals into a form suitable for feeding to a plurality of loudspeakers;

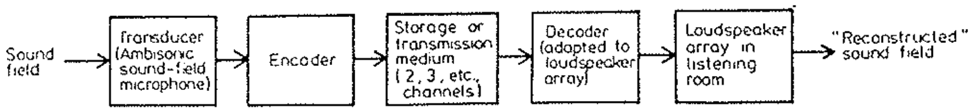

Fig. 1 Basic features of ambisonic reproduction.

(*e*) radiation from the loudspeakers of sounds which interact in the listening space so as to reproduce, as far as possible, a simulacrum of the original sound field.

Our present systematic understanding of the requirements and possibilities of these steps owes much to the work of Gerzon[2], particularly his analysis of the theory of encoding and of psycho-acoustic criteria in the reproduction of directionality, much of which is not yet published. The sections which next follow will discuss some of the salient problems in greater detail.

Sound-field microphones

It is first necessary to characterise the sound-field at a given place in the room where the performance is taking place. The sound field at a point can conveniently be characterised in terms of spherical harmonics representing the sound field pressure and its derivatives. The number of spherical harmonics up to and including various orders n is shown in Table 1. From this it is seen that the number of independent signals, each representing a spherical harmonic, is of the form $(n+1)^2$. The 0th order harmonic is equivalent to the sound-field pressure alone, and gives simply mono without directional information. The three spherical harmonics of order 1 are equivalent to the three cartesian components of particle velocity v_x, v_y and v_z. Higher orders of spherical harmonics represent non-redundant combinations of higher gradients; they are of potential interest for the future but need not be considered for the purpose of current developments.

In order to deal properly with indirect sound, which may of course arrive from any direction, a sound-field microphone must encode in a deliberately designed manner any sound reaching it, including sounds having a vertical component of travel. It can be shown that microphones which violate this condition necessarily introduce undesirable coloration into the reproduction of reverberant sounds, even when the vertical information is deliberately suppressed in subsequent processing. From this requirement, having regard to Table 1, it follows that the number of independent signals generated by a microphone should correspond to the square of the natural numbers; $(n+1)^2 = 1$ giving monophonic reproduction, and $(n+1)^2 = 4$ the minimum uncoloured reproduction which includes directionality.

Table 1 Spherical harmonics

Order	No. of harmonics	Cumulative total
0	1	1
1	3	4
2	5	9
n	$2n+1$	$(n+1)^2$

Gerzon has shown that particular polyhedral arrangements of microphone capsules can fulfil the relevant requirements. The simplest of these places the capsules at the face centres of the regular tetrahedron. The spacing of the capsules should be kept as small as possible, but as it cannot practicably be small compared with the wavelength at higher audio frequencies, correction circuits are needed to compensate for the resultant disturbance between the pressure and velocity responses of the composite microphone. There are some advantages in using a larger polyhedral array to generate a $(n+1)^2 = 4$ signal, and these are likely to be explored in the future.

The earliest ambisonic experiments were made using an improvised tetrahedral array of ordinary cardioid-response microphones. This arrangement, however, gives rise to considerable mutual interference, and it is impossible to obtain a sufficiently close spacing. Use of microphones in undesigned non-polyhedral arrays, or with spacings of 30 cm or more, or using replay direct into loudspeakers without pressure-velocity compensation, all give markedly sub-optimal results.

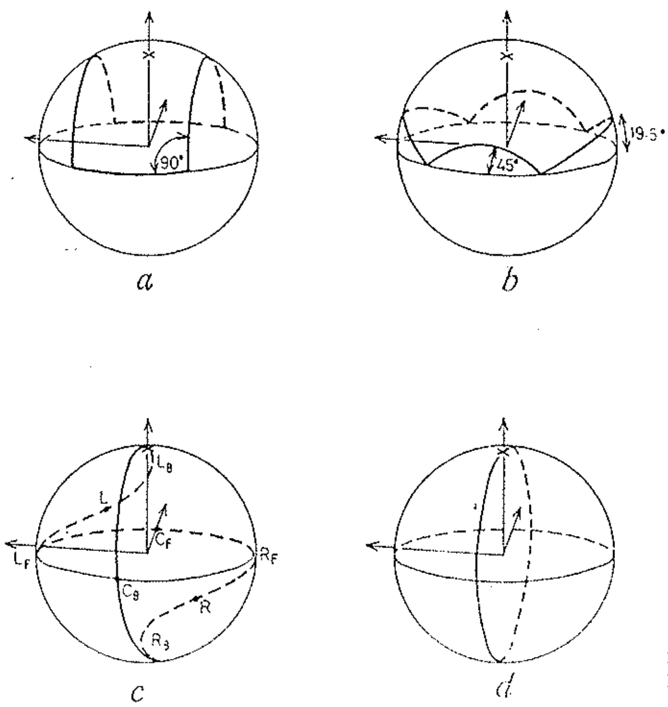

Fig. 2 Some representative horizontal pan-loci. *a*, Sansui 'QS'; *b*, closer approximation to great-circle locus consistent with four-channel master tapes (not in commercial use); *c*, CBS 'SQ'. Note cusps and left-right asymmetry due to choice of front-sector mapping and limitations of encoding from four pair-wise blended channels. C_F, Centre front; C_B, centre back; L, full left; R, full right; L_F, left front; and so on; *d*, a great-circle locus consistent with ambisonic encoding.

It is desirable to subject the raw signals from the sound-field microphone to a 4×4 matrix transformation into a form, call B-format, in which they have reduced sensitivity to small phase or amplitude errors in transmission or recording.

Encoding

The problem of encoding of directional information may be stated either in the form of how to make best use of a given number of available channels of communication, or of asking how many channels are necessary and sufficient for a given purpose.

Direction of travel can be characterised by the azimuth angle θ and the altitude angle $(\pi/2-\varphi)$. Both parameters can be encoded on to only two channels carrying signals A and B for example as

$$|B|/|A| = \tan(\varphi/2) \quad (1)$$

$$\angle B - \angle A = \theta \quad (2)$$

There are of course many other informationally-equivalent forms of these equations. In equation (1), $\tan(\varphi/2)$ is taken in order to avoid the 180° ambiguity that would result from using $\tan\varphi$.

Equations (1) and (2) can equivalently be regarded as defining polar angles θ and φ (not necessarily identical with azimuth and altitude) which characterise the relative amplitude and relative phase of signals encoded on to two channels. Any given encoding can then be represented graphically by a point on a sphere, known as the Poincaré or energy sphere[3]. This graphical representation is extremely useful in enabling the properties of proposed coding schemes to be understood, and it possesses a number of useful mathematical properties. The path on the energy sphere described by the representative point as the azimuth explores the horizontal circle is known as the horizontal pan locus; some representative examples are illustrated in Fig. 2.

There are several advantages in making this locus a great circle, and little or nothing to be gained by deviating from this form. Gerzon has shown that matrixing four pairwise-blended channels onto two channels, as in so-called quadraphony, necessarily gives a pan-locus consisting of four circular arcs intersecting so as to include a total angle of π rad at their corners; it therefore cannot give a great circle locus.

Encodings on to three or more channels can be represented similarly by points on hyper-spheres of appropriate dimensionality, but this is somewhat less useful as an aid to intuition because of the difficulty of graphical representation.

Equations (1) and (2) show that directionality can be completely encoded on to only two channels in the 'direction finding' sense, that is, any direction of arrival can be represented unambiguously. This is evidently necessary, but it is not sufficient, to ensure satisfactory performance of the overall ambisonic system, because it does not of itself ensure that the signal channels can be decoded in the required manner.

Decoding and psycho-acoustic criteria

An exact replica of the original sound field, in the whole region occupied by the listener's head and up to the highest audible frequencies, would require many thousands of audio channels of communication and is quite impracticable. The closeness of approximation that can be achieved depends both on the number of channels and on the number of loudspeakers that are available.

At frequencies low enough for the human head to be small compared with the wavelength of sound, it is indeed possible to reproduce physically the original sound field. At higher frequencies this is no longer possible, and the best that can be done is to fulfil as many as possible of the psycho-acoustic criteria whereby sound directions are localised. These criteria, in relation to practicable systems, have been established particularly by the work of Gerzon already referred to.

Ambisonic reproduction in which vertical directional information is preserved may be called periphonic, and ambisonic reproduction with only horizontal information pantophonic. Pantophonic reproduction geometrically requires a minimum of three loudspeakers, since the triangle is the minimal figure able to enclose space in a plane. Similarly a minimum of four loudspeakers are geometrically necessary to surround the listener in three dimensions and give periphonic reproduction.

To satisfy the psycho-acoustic criteria sufficiently well, however, the practical minimum is four loudspeakers for pantophonic reproduction and six for periphony. One of the effects of using too few loudspeakers is that the loudspeakers themselves can be heard obtrusively as separate sources in addition to any ambisonic impression. This is closely associated with the 'detent effect' whereby as an actual source of sound moves its reproduced image seems to cling to each loudspeaker before jumping relatively rapidly to the next.

The limitations set by the number of loudspeakers interact with those set by the number of available channels. Qualitatively this is as expected, since the available number of loudspeakers limits the number of constraints available to control the sound field in the listening room, while the number of channels places restrictions on the extent to which these control constraints are independent. If four loudspeakers is taken as the practicable minimum for pantophonic reproduction, this is also in most people's minds at present the economic maximum. It can be shown that the capabilities of four loudspeakers in a planar array can be fully exploited using just three independent channels of audio information. By a happy coincidence, the current system of f.m. stereo broadcasting has sufficient bandwidth to transmit three audio channels. Interestingly, if the attempt is made to feed four loudspeakers from four channels in a non-redundant manner, the result is to enhance the detent effect. It can be shown that similarly four channels are sufficient to feed the minimal number of six loudspeakers required for periphonic reproduction. In general, the number of loudspeakers should exceed the number of channels, so that it always pays to interpose a decoder between the communication channels and loudspeakers.

The practical outcome is that a two-channel four-loudspeaker pantophonic system can give good directional localisation with fulfilment of the salient psycho-acoustic criteria over a larger area than two-speaker stereo. Under favourable conditions, in surroundings that are acoustically neither too live nor too dead, the area of satisfactory and relaxed listening can fill most of an ordinary domestic room. The chief defect is a relative phase shift of 180° around the azimuth circle, unavoidable in any two-channel system, and the best that can be done is to distribute it in an optimally compromised manner. Current ambisonic decoders provide adjustments to compensate for spherical-wave effects of finite loudspeaker distance, and for non-square loudspeaker layouts.

The use of three channels enables the phase difficulty to be removed, and gives some improvement in the accuracy and stability of directional location. The next step in improving pantophonic reproduction is, as already indicated, to add additional loudspeakers rather than additional channels (which can actually make the results worse). Periphonic reproduction is, in the opinion of some who have heard it, at least as great an advance on pantophonic as the latter is on stereo, or stereo on mono. Three-channel periphonic systems may be possible with compromises not dissimilar to those involved in two-channel pantophonic reproduction. The minimum requirement of six loudspeakers in a non-planar array inhibits commercial interest at present, although periphony may have considerable interest for the future.

Beyond the $(n+1)^2 = 4$ periphonic system, the next basis is $(n+1)^2 = 9$, which is also very much a matter for the future.

Monophonic and stereophonic compatibility

The introduction of a new system always involves problems of compatibility with what went before. Ambisonic compatibility is therefore essentially a matter of the compatibility of its two-channel realisation with the existing two-channel mono and stereo formats.

Complete aesthetic compatibility is never possible between an old system and a new one providing additional information, as for example between stereo and mono or between colour and monochrome television. Technical compatibility depends on which out of a gamut of informationally equivalent ambisonic codings is employed. For ambisonic playback, these codings are equivalent because there is freedom to design the decoder to match the encoding. Ambisonic encoding can easily be made compatible with either stereo or mono, but not both simultaneously. A compromise is necessary, and the basic cause of this is not any fault in ambisonic encoding, but in the historical accident that the apparently obvious stereo and mono encodings concealed an inherent mutual incompatibility. Essentially, a balance must be struck between some attenuation of rear sources and some phase shift between the two loudspeakers in stereo playback; fortunately this can be done in a fairly harmless way.

Ambisonics is still in the experimental stage. It can be realised using two, three or more channels. The two-channel version has particular interest because of the wide availability of two-channel recording and broadcasting media developed originally for stereo use, and the three-channel version has particular interest in relation to f.m. broadcasting. Ambisonics is a complete system from the pick-up of the original sound field through encoding and decoding to the reconstruction of an approximation to the original sound-field in the listener's home. It therefore requires the simultaneous availability of both hardware and software; that is, of suitable decoders and of recordings or broadcasts on which to use them. None of these are commercially available at the present time, but may become so before very long.

Within the limitations of this article the attempt has been made to outline the basic principles of ambisonics with a minimum of purely technical detail. Ambisonics is not so much a single system as a family of realisations having in common the attempt to state clear aesthetic and technological aims, to adjust these aims to what is physically and mathematically possible, and then to implement a properly engineered system to fulfil the objectives.

The development has been performed principally by M. Gerzon, J. S. Wright and the author with the help and support of the National Research Development Corporation. The experimental equipment was built mainly in the Research Unit for Instrument Physics of the University of Reading. The help of numerous academic and industrial colleagues is also gratefully acknowledged.

1 Shannon, C. E., *Bell Syst. tech. J.*, **27**, 379, 623 (1948).
2 Gerzon, M., *Wireless World*, Dec. 1974 (in the press).
3 Poincare, H., Theorie Mathematique de la Lumiere II, Georges Carré, Paris **189**, 282–285; 302.
4 Scheiber, P., *J. Audio Engng Soc.*, **19**, 835–9 (1971).