BBC RD 1979/25

BBC

RESEARCH DEPARTMENT

REPORT

Spherical harmonic analysis and some applications to surround sound

P.S. Gaskell, M.A.

Research Department, Engineering Division
THE BRITISH BROADCASTING CORPORATION November 1979

BBC RD 1979/25

UDC 681.84:
621.317.757

SPHERICAL HARMONIC ANALYSIS AND SOME APPLICATIONS TO SURROUND SOUND

P.S. Gaskell, M.A.

Summary

Spherical harmonic analysis in three dimensions and azimuthal harmonic analysis in two dimensions are powerful methods for studying the problems associated with transmitting surround sound. This Report describes these methods and shows how they can highlight some of the fundamental factors and limitations of surround sound transmission. A number of specific examples are presented.

Issued under the authority of

Head of Research Department

Research Department, Engineering Division,
BRITISH BROADCASTING CORPORATION

November 1979

(PH-210)

SPHERICAL HARMONIC ANALYSIS AND SOME APPLICATIONS TO SURROUND SOUND

(PH-210)

SPHERICAL HARMONIC ANALYSIS AND SOME APPLICATIONS TO SURROUND SOUND

P.S. Gaskell, M.A.

1. Introduction

The two and three dimensional characteristics of the apparent disposition and resolution of sound sources in a sound field, as observed at a point in space, may be described by azimuthal and spherical harmonics. Spherical harmonic analysis furnishes a powerful method for studying surround sound and the problems of transmitting and reproducing it. A number of papers demonstrating this have appeared in the literature[1,2,3]. This Report gives a basic description of this analysis and discusses a number of applications relevant to surround sound transmission.

For largely historical reasons, symbols describing parameters in this subject have appeared in the past with a notable lack of consistency and this has led to some confusion. An attempt is made in this Report to resolve some of the discrepancies although a standard of terminology is unlikely to be achieved easily.

The problems of transmitting and reproducing surround sound are very intricate and are due largely to the sophistication of the human hearing system. It should be remembered then that when designing particular transmission systems, theoretical analysis must be complemented by a measure of pragmatism in order to arrive at a practical solution.

2. Surround sound transmission

2.1 Azimuthal harmonic analysis

At any point in a sound field, sounds arrive from an infinity of directions. The net effect may be thought of as the sum of sounds emanating from a large number of point sources. If we consider just one point source of unit magnitude at azimuth θ_i in the horizontal plane, its azimuthal spatial distribution appears to the observer as a delta function that repeats itself every 2π radians. In a similar way that repetitive time signals may be represented by a Fourier series expansion, the azimuthal distribution $g(\theta)$ of a group of sources may be expressed by

$$g(\theta) = a_o + \sum_{n=1}^{\infty} (a_n \cos n\theta + b_n \sin n\theta).$$

For a point source of unit magnitude

$$g(\theta) = \delta\,(\theta - \theta_i)$$

Evaluating the coefficients a_n and b_n for a point source in the same way as for a Fourier series expansion gives

$$g(\theta) = \frac{1}{2\pi} + \frac{1}{\pi}\sum_{n=1}^{\infty} (\cos n\theta_i \cos n\theta + \sin n\theta_i \sin n\theta)$$

Summing for all i, the complete ensemble of point sources is retrieved. Terms in $\cos n\theta$ and $\sin n\theta$ are said to be the azimuthal harmonics of order n.

Schemes for the transmission of surround sound in the horizontal plane, based on this form of harmonic analysis, have been described in earlier papers[1,2]. These have shown that, for the transmission of surround sound of order n in the horizontal plane, $(2n + 1)$ channels are needed. In this Report, mainly first order transmission will be considered. Thus for a point source, the first (and zero) order terms are given by:

$$g(\theta) = \frac{1}{2\pi} + \frac{1}{\pi}(\cos \theta_i \cos \theta + \sin \theta_i \sin \theta)$$

$$= \frac{1}{\pi}\left\{ \tfrac{1}{2} + \cos\,(\theta - \theta_i)\right\}$$

The polar diagram of $g(\theta)$ is then that of a 120° hypercardioid directed towards the source azimuth $\theta = \theta_i$ (see Fig. 1). Three transmission channels are needed, one carrying signals with an omnidirectional characteristic, and the other two carrying signals with figure-of-eight characteristics.* The magnitude of the zero order component is $1/2\pi$ and the magnitudes of the two first order components, $\cos\theta$ and $\sin\theta$ are $(1/\pi)\cos\theta_i$ and $(1/\pi)\sin\theta_i$ respectively. For the general case of an ensemble of sources with azimuthal distribution $g(\theta)$, the zero order component magnitude is

* Note that, when describing the polar characteristics of microphones or sound fields, the common practice of taking the modulus of the azimuthal variation is followed here. Thus $r = \cos\theta$ or $\sin\theta$ each has a figure-of-eight characteristic; if the modulus is not taken, these would each represent only a single circle.

$$\frac{1}{2\pi}\int_0^{2\pi} g(\theta)\, d\theta$$

and the magnitudes of the cos θ and sin θ components are respectively

$$\frac{1}{\pi}\int_0^{2\pi} g(\theta)\cos\theta\, d\theta$$

and
$$\frac{1}{\pi}\int_0^{2\pi} g(\theta)\sin\theta\, d\theta$$

The first order, horizontal, azimuthal component magnitudes at a point in a sound field may be sampled by means of first order microphones. These are widely used at present by recording and broadcasting companies and they have a general azimuthal response to a point source at azimuth θ_i of the form

$$a_o + a_1 \cos\theta_i + b_1 \sin\theta_i$$

or, if its maximum is directed towards $\theta = 0$,

$$a_o + a_1 \cos\theta_i$$

Responses of first order microphones with various ratios of a_o/a_1 are shown in Fig. 1. Clusters of three or more microphones of this type can resolve the zero and first order components of the sound field at a point by the use of a suitable matrix (see Sections 3 and 5.2). Electronically panned signals or highly directional microphones, however, may contain higher order components of the form $a_n \cos n\theta_i$ and $b_n \sin n\theta_i$ ($n > 1$) (see Section 4).

In three dimensions, the azimuthal harmonics become spherical harmonics and the expansion takes the general form

$$g(\theta, \phi) = \sum_{n=0}^{\infty} Y_n(\theta, \phi)$$

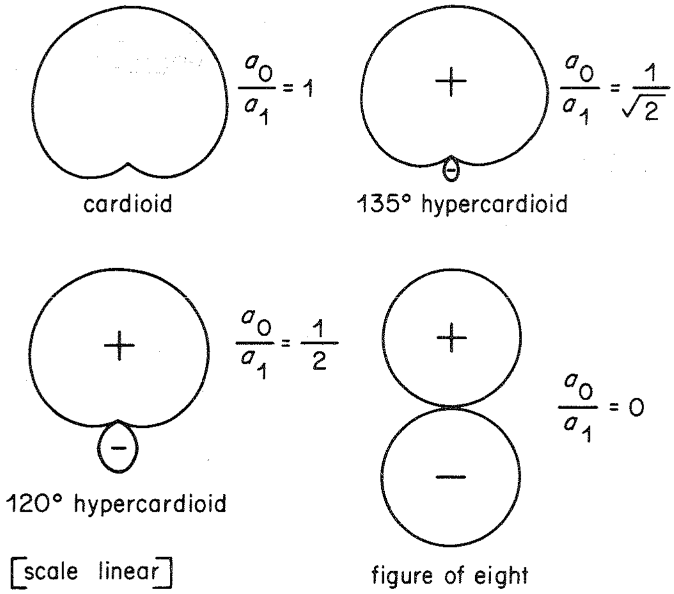


Fig. 1 - Cardioid-type characteristics

where

$$Y_n(\theta, \phi) = a_{no} P_n(\cos\phi) + 2\sum_{m=1}^{\infty} [a_{nm} \cos m\theta + b_{nm} \sin m\theta]\, P_n^{\,m}(\cos\phi)$$

θ, ϕ are the standard polar co-ordinates

and $P_n^{\,m}$ are the associated Legendre polynomials.

2.2 Original, programme and reproduction sound fields.

Having resolved the azimuthal (or spherical) harmonics, their recorded component magnitudes are sent down transmission channels. Each channel is notionally labelled according to its azimuthal order, on the basis that the channel signal will be ideally decoded and reproduced according to its proper order. In the horizontal plane, "ideal" decoding implies that the "zero order, omni-directional" channel signal,

$$\frac{1}{2\pi}\int_0^{2\pi} g(\theta)\, d\theta$$

is reproduced as a true omni-directional component in the reproduced sound field, the "cos θ" channel signal,

$$\frac{1}{\pi}\int_0^{2\pi} g(\theta)\cos\theta\, d\theta$$

is correctly reproduced as a forward-facing figure-of-eight component and so on. The sum of these components may be called the "programme" sound field.

In normal programme production, it frequently occurs that the "programme" sound field differs significantly from the original sound field. In other words, it is unusual for the "programme" sound field to correspond to the sound field at any point in the space of the enclosure in which the recording was made. This results from the use of directional microphones, electronic panning, artificial reverberation and any of the other options that may be used in the course of a recording session. This is done not only to synthesise artificial sounds, but also to achieve sound balance, clarity, perspective, etc. [4] and is standard practice for normal programme production. Thus the "programme" sound field is likely to differ significantly from the original sound field even for programmes that are intended to sound "natural and life-like".

Unfortunately, the "programme" sound field remains an unrealised goal in so far as methods of achieving "ideal" decoding have yet to be discovered. Practical decoders, limited by the number of loudspeakers used and by the present incomplete understanding of the human hearing system, give "reproduction" sound fields that deviate from the ideal. This is discussed in Section 2.4.

2.3 Transmission

For transmission, the order of directional information that may be conveyed depends on the number of channels available[1,2,3]. Coding of the signals often takes place, prior to transmission, to achieve compatibility (mono/stereo, etc.) and/or if it is intended to convey high order information with only a limited number of channels. This may be done by a modulation process or by introducing interchannel phase. Coding using interchannel phase may cause complex crosstalk between the azimuthal components. Thus, when the signals are decoded and broken down into their component parts, the zero order term for example may contain traces of higher order components. Obviously, coding by multiplexing is capable of maintaining independent signal components, though at an obvious cost.

2.4 Decoding and reproduction

The ear/brain ensemble is capable of high resolution (typically 4° image width) and a very large number of loudspeakers and transmission channels may be needed to simulate the sound field exactly. Fortunately, however, the ear is interpretive. When presented with a sound field of low azimuthal order, evidence suggests that the ear interprets it as one of higher order depending to some extent on the source azimuth. Thus it perceives sound images that are sharper than those which the azimuthal harmonic content would dictate from a theoretical point of view. If a limited number of loudspeakers and transmission channels are used, the ear can be deceived into believing that a fair approximation to the intended "programme" sound field has been reproduced. This, however, requires an intimate understanding of the hearing system and this is both highly sophisticated and complex.

Over the past seventy years or more, a great deal of research on the functioning of the hearing system has been documented. It appears that the ear/brain ensemble uses many different mechanisms for judging sound quality and localising sounds. Not all of these mechanisms are clearly understood and we are still left with an incomplete understanding of the overall system. However, by satisfying as many aural conditions as possible in the reproduced sound field, the number of conflicting auditory cues that have to be resolved by the brain, when interpreting the sounds, is minimised.

Practical decoding systems using a limited number of loudspeakers, in a listening room with its own acoustic characteristics, are not able to reproduce the "programme" sound field exactly, even to a limited order. The perceived sound field is liable to be distorted, particularly if the listener turns his head or moves about the room. Each encoding format may be complemented, therefore, by a variety of different decoding systems, each with its own characteristic deviations from the ideal sound field. The choice of decoder depends to some extent on the particular requirements of the listener.

In programme production, this leads to a problem. Whenever a programme is balanced, a particular decoding system is used to monitor the programme and idiosyncrasies of the decoding system (such as tonal quality or localisation) are in many cases compensated for by the programme producer at this stage. Replaying the final programme on the same decoder gives results similar to that intended by the producer, but other decoders give different results. This difficulty is particularly evident in the case of 2-channel quadraphonic encodings where either "linear" or "logic" decoders[5] may be used.

Gerzon[6,7,8] has designed linear surround sound decoders for 2-, 3- and 4-channel transmission systems (see Appendix 2). For a transmission system designed to reproduce the azimuthal characteristics to first order, he shows that the loudspeaker feeds should only consist of first and zero order terms; the artificial generation and insertion of higher order terms (e.g. by pair-wise panning) only impairs reproduction. Decoders based on these criteria have given successful results. However, other transmission schemes (see Section 4) can be devised whereby signals containing higher order components are transmitted and decoded to give reproduction of higher order components at selected positions in the sound stage (e.g. single loudspeaker excitation). Both approaches will be discussed in the following sections.

Stereo and mono reproduction map the original sound stage to a limited arc of the reproduction sound stage. As with surround sound, the best way of decoding the recorded signals for stereo reproduction is closely linked to the hearing system. This has been discussed in other papers[9] and will only be mentioned briefly here (see Appendix 2).

3. M, S and T components of a sound field

Symbols widely used in discussions on sound transmission are M, S and T*. These may be defined as the magnitudes of the first order, horizontal, azimuthal components of a sound field at a point within it; M is the magnitude of the omni-directional (pressure) component and S and T are the magnitudes of two figure-of-eight (velocity) components, one directed towards centre-left and the other towards centre-front respectively. These symbols are also used to represent signals obtained from microphones whose magnitudes are proportional to the magnitudes of the M, S and T components of a sound field. Thus, as was seen in Section 2.1, the normalised values of M, S and T for a sound field generated by a point source at azimuth θ_i are given by

$$M = \frac{1}{2\pi}$$

$$S = \frac{1}{\pi}\sin\theta_i$$

$$T = \frac{1}{\pi}\cos\theta_i$$

* Unfortunately, the symbols M and S are also widely used to describe the Sum and Difference signals of F.M. stereophonic transmissions where they have a different meaning to that in the present context.

Given an ensemble of sources with azimuthal distribution $g(\theta)$,

$$M = \frac{1}{2\pi}\int_0^{2\pi} g(\theta)\,d\theta$$

$$S = \frac{1}{\pi}\int_0^{2\pi} g(\theta)\sin\theta\,d\theta$$

and

$$T = \frac{1}{\pi}\int_0^{2\pi} g(\theta)\cos\theta\,d\theta$$

For a single source, the ratio of M to the peak values of S and T, S_{peak} and T_{peak} is ½. This ratio changes, however, with more than one source according to the source distribution $g(\theta)$. It is often convenient therefore to introduce a constant gain factor into the M channel for the purposes of signal handling and transmission (see Section 7).

The values of M, S and T may be derived from a minimum of three first order microphones suitably disposed. A convenient, but less efficient method is to use four coincident, orthogonal cardioid or hypercardioid microphones directed towards the four corners of a square, L_F, R_F, L_B and R_B, using conventional notation[5] (L = left R = right, F = front, B = back). The use of four such cardioid microphones will now be discussed.

The general response of a cardioid-type microphone, with its maximum directed towards $\theta = 0$, to a point source at azimuth θ_i is

$$V(\theta_i) = A + \cos\theta_i$$

where A = constant

If θ_i is measured anti-clockwise from centre-front, the responses of four cardioid microphones, directed towards the corner of a square, are given by:

$$L_F = A + \cos(\theta_i - 45^\circ)$$

$$R_F = A + \cos(\theta_i + 45^\circ)$$

$$L_B = A + \cos(\theta_i - 135°)$$

and $$R_B = A + \cos(\theta_i + 135°)$$

Signals corresponding to the values of M, S and T can be derived from these microphone responses according to

$$M = ½\lambda(L_F + R_F + L_B + R_B) = \frac{1}{\sqrt{2}}$$

$$S = ½\ (L_F - R_F + L_B - R_B) = \sqrt{2}\sin\theta_i$$

and $$T = ½\ (L_F + R_F - L_B - R_B) = \sqrt{2}\cos\theta_i$$

where $$\lambda = \frac{1}{2\sqrt{2}A}$$

If λ is set to unity, the often quoted equations for M, S and T are given as follows

$$M = ½(L_F + R_F + L_B + R_B)$$

$$S = ½(L_F - R_F + L_B - R_B)$$

and $$T = ½(L_F + R_F - L_B - R_B)$$

However, the correct ratio of M to S and T is then no longer preserved, except in the special case of

$$A = 1/2\sqrt{2}$$

The ratio of ideal M, given when

$$\lambda = \frac{1}{2\sqrt{2}A},$$

to the value of M when $\lambda = 1$ is shown in Table 1 for various microphone types. The parameter quoted is the factor by which M is in error relative to S and T. The results hold for both the single source case and for an ensemble of sources.

Thus, although all the cardioid-type microphones (except figure-of-eight) give the correct form of the M, S and T responses, the ratio of M to S_{peak} and T_{peak} changes. It is necessary, therefore, to standardise to a particular format (i.e. the ratio of M to M_{ideal}) for transmission for the sake of decoding. Criteria of signal handling dictate the choice as will be discussed in Section 7.

The symbols M, S and T are often used to describe parameters other than those just given. In the present context, therefore, some authors[8] use a different notation as follows:

$$W = M$$

$$X = T$$

$$Y = S$$

with the addition of a fourth term Z, corresponding to a figure-of-eight characteristic directed vertically upwards. These symbols reflect conveniently the directions of the right-handed set of x, y, z axes of cartesian geometry.

4. Microphone mixing techniques

4.1 Pair-wise panning

One of the simplest forms of surround sound reproduction is effected by using four separate transmission channels to each of four loudspeakers in a square or rectangular array; this transmission system is often referred to as "discrete" quadraphony. One widely used method of panning mono sound sources that may be readily implemented with this system consists of panning a mono source according to a sin/cos law between two adjacent channels – so called "pair-wise" panning. To gain further insight into its implications on surround sound transmission, it is profitable to treat the technique with azimuthal harmonic analysis.

4.1.1 Decoding

In a first order transmission scheme, one

Table 1 – M, S, T response for various cardioid-type microphones

Microphone Type	A	Ratio of M to M_{ideal}
Cardioid	1	$2\sqrt{2}$
135° hypercardioid	$1/\sqrt{2}$	2
120° hypercardioid	½	$\sqrt{2}$
Figure-of-eight	0	0

method of decoding[7] dictates that the signal feeds to four loudspeakers in a rectangle should be

$$L'_F = \tfrac{1}{2}\left(kM + \frac{S}{\sqrt{2}\sin\psi} + \frac{T}{\sqrt{2}\cos\psi}\right)$$

$$R'_F = \tfrac{1}{2}\left(kM - \frac{S}{\sqrt{2}\sin\psi} + \frac{T}{\sqrt{2}\cos\psi}\right)$$

$$L'_B = \tfrac{1}{2}\left(kM + \frac{S}{\sqrt{2}\sin\psi} - \frac{T}{\sqrt{2}\cos\psi}\right)$$

$$R'_B = \tfrac{1}{2}\left(kM - \frac{S}{\sqrt{2}\sin\psi} - \frac{T}{\sqrt{2}\cos\psi}\right)$$

where the loudspeakers are at azimuths $\pm\,\psi$ and $180^\circ \pm \psi$ and k is a constant.

If M, S and T are such that the ratio of M to M_{ideal} is $\sqrt{2}$, as has been suggested for some transmission systems, then k should assume values of

$$1 \text{ or } \frac{1}{\sqrt{2}} \text{(see reference 7).}$$

If $k = 1$ and the speakers are in a square with $\psi = 45^\circ$, then

$$\left.\begin{aligned} L'_F &= \tfrac{1}{2}(M+S+T) \\ R'_F &= \tfrac{1}{2}(M-S+T) \\ L'_B &= \tfrac{1}{2}(M+S-T) \\ \text{and} \quad R'_B &= \tfrac{1}{2}(M-S-T) \end{aligned}\right\} \quad (1)$$

This is one of the simplest forms of decoding and forms the basis for the pair-wise panning scheme.

4.1.2 Recording and transmission

With four loudspeakers in a square, by modifying the decoding Equations (1), it is possible to send M, S and T like signals down the "M", "S" and "T" channels, together with a fourth signal Q in order to give single speaker excitation. This, of course, represents a high azimuthal order in the reproduced sound field and is not in keeping with the first order scheme. The four channel signals are formulated as follows:

$$\left.\begin{aligned} M_P &= \tfrac{1}{2}(L_F + R_F + L_B + R_B) \\ S_P &= \tfrac{1}{2}(L_F - R_F + L_B - R_B) \\ T_P &= \tfrac{1}{2}(L_F + R_F - L_B - R_B) \\ Q_P &= \tfrac{1}{2}(L_F - R_F - L_B + R_B) \end{aligned}\right\} \quad (2)$$

where the L_F, R_F, L_B and R_B signals are derived from a mono signal panned according to a sin/cos law between pairs of adjacent channels, so called pair-wise panning. If the source is to be panned across a corner, switching of the signal pairs takes place. The azimuthal characteristics are similar to those of pairs of orthogonal figure-of-eight microphones under free-field conditions. If we take as an example the front pair of corners, where pair-wise panning is achieved by arranging that

$$\begin{aligned} L_F &= \cos(\theta_i - 45^\circ) \\ R_F &= \cos(\theta_i + 45^\circ) \\ L_B &= 0 \\ R_B &= 0 \end{aligned}$$

we have, by substitution in Equations (2),

$$M_P = \frac{1}{\sqrt{2}}\cos\theta_i$$

$$S_P = \frac{1}{\sqrt{2}}\sin\theta_i$$

$$T_P = \frac{1}{\sqrt{2}}\cos\theta_i$$

$$Q_P = \frac{1}{\sqrt{2}}\sin\theta_i$$

The corresponding decoding (for a square layout) is:

$$\left.\begin{aligned} L'_F &= \tfrac{1}{2}(M_P + S_P + T_P + Q_P) \\ R'_F &= \tfrac{1}{2}(M_P - S_P + T_P - Q_P) \\ L'_B &= \tfrac{1}{2}(M_P + S_P - T_P - Q_P) \\ L'_B &= \tfrac{1}{2}(M_P - S_P - T_P + Q_P) \end{aligned}\right\} \quad (3)$$

Substituting for M_P, S_P, T_P and Q_P gives

$$L'_F = \cos(\theta_i - 45°)$$

$$R'_F = \cos(\theta_i + 45°)$$

$$L'_B = 0$$

$$R'_B = 0$$

as before. Thus, if a source is panned to L_F, i.e. $\theta_i = 45°$, only the L'_F speaker is excited. It is seen that this decoding is compatible with Equations (1).

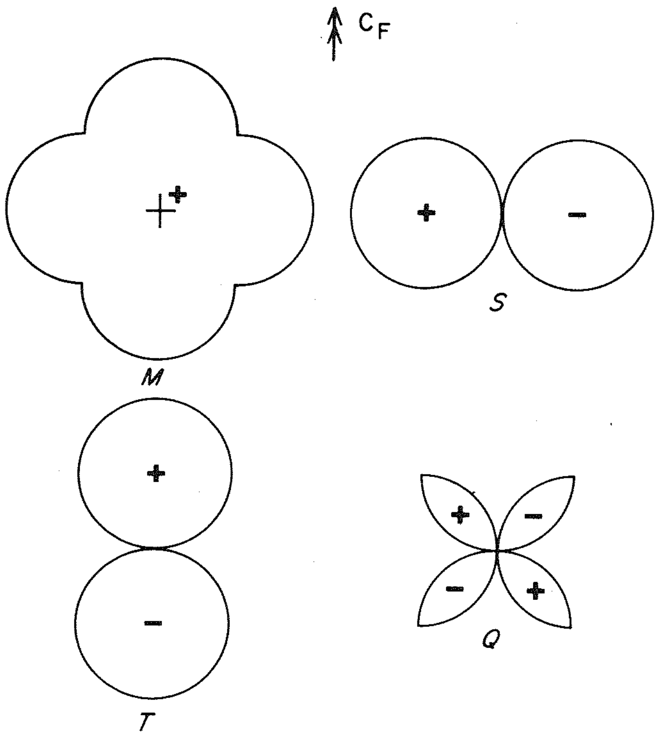


Fig. 2 - Pair-wise panning – M, S, T, Q.

The manner in which the M_P, S_P, T_P and Q_P signals vary with θ_i for all the pairs of corners is shown in Table 2 and Fig. 2. S_P and T_P have the correct sin θ_i and cos θ_i form respectively for all values of θ_i from 0 to 360°, but the M_P signal only approximates to a constant and drops by 3 dB at the corners; the Q_P signal has a X-shaped characteristic. The M_P and Q_P signals suffer discontinuities at the corner positions and therefore contain an infinite series of high order azimuthal components.

The series expansion of Q_P may be shown to be

$$\frac{1}{\pi}\sum_{n=1}^{\infty}\left(\frac{1}{2n-1}\sin\frac{n\pi}{2}+\frac{1}{2n+1}\cos\frac{n\pi}{2}\right)\sin 2n\theta_i$$

$$=\frac{1}{\pi}\left(\sin 2\theta_i - \frac{1}{5}\sin 4\theta_i - \frac{1}{5}\sin 6\theta_i + \frac{1}{9}\sin 8\theta_i + \ldots\right)$$

4.1.3 Discussion

Pair-wise panning and the associated encoding and decoding are an artificial means of achieving high order transmission and reproduction. Advantage is taken of the idiosyncracies of the decoding (due primarily to the limited number of loudspeakers) to obtain sound images of anomalously high resolution at selected positions in the reproduced sound field. To achieve this, special high order signals, similar to *M*, *S* and *T*, together with a fourth signal *Q* are transmitted down the "*M*", "*S*", "*T*" and "*Q*" channels. These signals (M_P, S_P, T_P and Q_P) do not correspond to the azimuthal components of a real sound field.

To realise this high degree of resolution, the encoding must correspond closely to the decoding format. The coding used for pair-wise panning

Table 2 — Pair-wise panning for a square layout

	Pair-wise panning between corners			
	L_F and R_F	L_B and L_F	R_B and L_B	R_B and R_F
M_P	cos θ_i	sin θ_i	–cos θ_i	–sin θ_i
S_P	sin θ_i	sin θ_i	sin θ_i	sin θ_i
T_P	cos θ_i	cos θ_i	cos θ_i	cos θ_i
Q_P	sin θ_i	cos θ_i	–sin θ_i	–cos θ_i

differs therefore from correct M, S and T coding which is an entirely general scheme and is much less dependent on the decoding adopted. However, although pair-wise panning does not conform to the strict characteristics of correct M, S and T transmission, it does make best use of the limitations of the rectangular array of loudspeakers as a decoding method. In so doing, it provides an effective and practicable technique that is an important facility in the production repertoire.

In view of the discontinuous nature of pair-wise panning signals, it is not easy to convert them to correct M, S and T signals. However, within the context of 2-channel, System HJ (13LP2) encoding, both types of signal may be fed to an HJ encoder with L_F, R_F, L_B and R_B inputs with satisfactory results. Pair-wise panning signals appear as four corner signals L_F, R_F, L_B and R_B and are fed directly to the encoder inputs. With M, S and T signals, where the ratio of M to M_{ideal} is 2, corner signals corresponding to those from a cluster of four 135° hypercardioid microphones may be derived using Equations (1). These signals, or those from an actual 135° hypercardioid microphone cluster, may then also be fed directly to the HJ encoder. Two encode loci are given with the two sets of signals and both give satisfactory results. These are shown on the Scheiber sphere in Fig. 3 and they touch at the centre quadrant positions. The compatibility between the two types of source signal is an important and valuable feature of the 13LP2 encode option of System HJ.

4.2 Coincident pairs of microphones

Pairs of cardioid microphones, subtending angles typically of 90° or 180° are very frequently used in quad and stereo[4]. These normally give the same result as an orthogonal pair of figure-of-eight microphones, but with the addition of a constant omni-directional component to the M and T signals. Thus, an orthogonal pair of cardioid microphones, directed towards C_F, with the "left" microphone signal

$$L = 1 + \cos(\theta_i - 45°)$$

fed to the L_F input and the "right" microphone signal

$$R = 1 + \cos(\theta_i + 45°)$$

fed to the R_F input, gives, using Equations (2),

$$M = 1 + \frac{1}{\sqrt{2}}\cos\theta_i$$

$$S = \frac{1}{\sqrt{2}}\sin\theta_i$$

$$T = 1 + \frac{1}{\sqrt{2}}\cos\theta_i$$

and

$$Q = \frac{1}{\sqrt{2}}\sin\theta_i$$

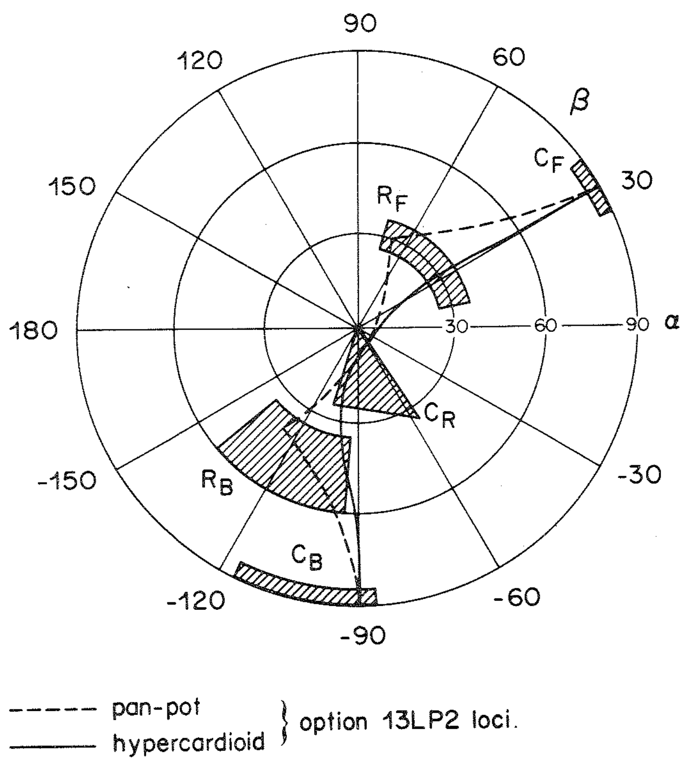


Fig. 3 - System HJ tolerance zones and option 13LP2 loci

As with pair-wise panning of mono sources, which is equivalent to pairs of figure-of-eight microphones, pairs of other cardioid-type microphones have azimuthal responses that are only approximations to correct *M, S, T* signals, i.e. of the form *A*, $\sin\theta_i$, $\cos\theta_i$ respectively. Nevertheless, when decoded and reproduced as described by Equations (2) and (3), satisfactory results are given with this technique.

In the course of a normal sound balance, the particular tonal and azimuthal characteristics of the various cardioid pairs are often exploited to achieve certain effects[4]. Tonal differences between the various cardioid pairs can be traced to the anomalous tonal characteristics and the non-coincidence of the microphones at high frequencies. The recently developed "Sound field" microphone[10] goes some way towards overcoming these deficiencies, as well as giving correct *M, S* and *T* signals. This will be described in Section 5.2.

5. Spherical harmonic analysis

5.1 First order characteristics in three dimensions

Using polar co-ordinates (r, θ, ϕ) (see Fig. 4), the figure-of-eight characteristics along the *x*-, *y*- and *z*-axes are respectively

$$\cos\theta\sin\phi$$

$$\sin\theta\sin\phi$$

and $$\cos\phi$$

These form the first order azimuthal components in three dimensions, i.e. the first order spherical harmonics (the omni-directional component is the zero order term). Adopting the notation quoted at the end of Section 3, we have that the first order component magnitudes for a point source located in the direction (r, θ_i, ϕ_i) are

$$W = 1/3$$

$$X = \cos\theta_i \sin\phi_i$$

$$Y = \sin\theta_i \sin\phi_i$$

$$Z = \cos\phi_i$$

Thus, if the source distribution function $g(\theta, \phi)$ corresponds to a single point source at (r, θ_i, ϕ_i),

$$g(\theta, \phi) = 1/3 + \cos\theta_i \sin\phi_i \cos\theta \sin\phi + \sin\theta_i \sin\phi_i \sin\theta\sin\phi + \cos\phi_i\cos\phi + \ldots$$

For an ensemble of sources with general distribu-

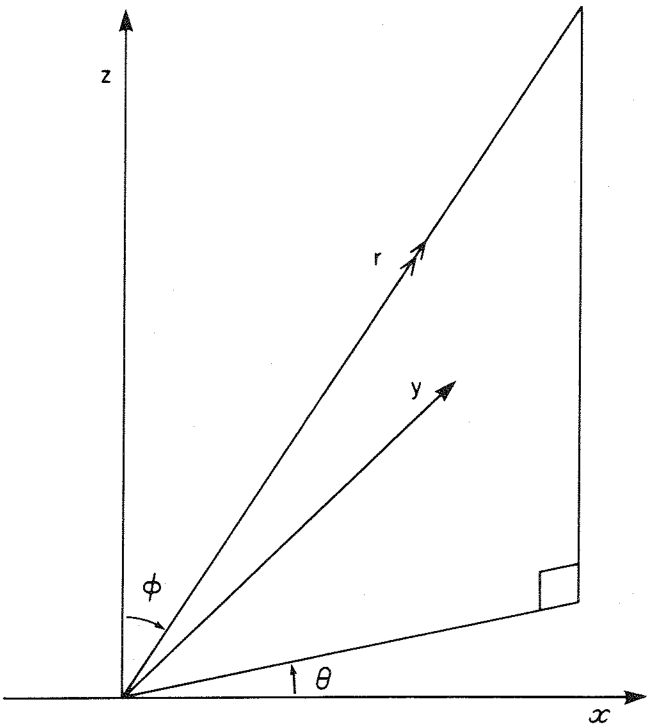


Fig. 4 - Polar co-ordinates

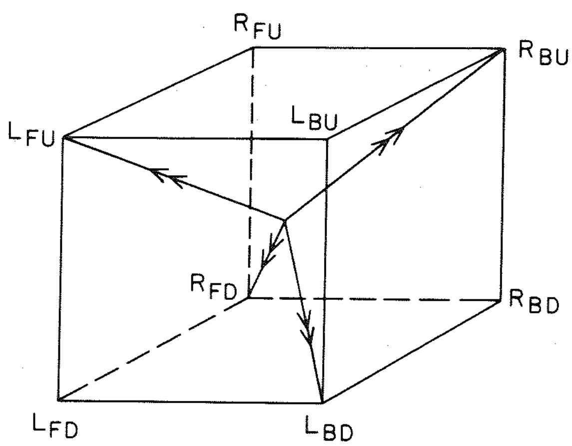


Fig. 5 - Sound-field microphones — capsule axes

tion function $g(\theta, \phi)$,

$$W = 1/3 \int_{\theta=0}^{2\pi} \int_{\phi=0}^{\pi} g(\theta, \phi) \sin\phi \, d\theta \, d\phi$$

$$X = \int_{0}^{2\pi} \int_{0}^{\pi} g(\theta, \phi) \cos\theta \sin^2\phi \, d\theta \, d\phi$$

$$Y = \int_{0}^{2\pi} \int_{0}^{\pi} g(\theta, \phi) \sin\theta \sin^2\phi \, d\theta \, d\phi$$

$$Z = \int_{0}^{2\pi} \int_{0}^{\pi} g(\theta, \phi) \cos\phi \sin\phi \, d\theta \, d\phi$$

As in the 2-D case, it is often convenient to change the gain of W relative to X, Y and Z for optimum signal handling and transmission.

Spherical harmonic analysis forms a powerful technique for studying surround sound in three dimensions[3] and two applications are now described.

5.2 Sound field microphone

The "Sound-field" microphone[10] samples in three dimensions the sound field at a point in space. If offers unprecendented operational versatility and allows one effectively to tilt and rotate the microphone and to change its directional characteristics remotely by electronic processing of output signals. It was developed by P.G. Craven and M.A. Gerzon under the auspices of the NRDC*. So far as is known, no analytical description of its functioning has been publicised and the principles of operation are presented here.

The "microphone" in fact consists of four capsules that are directed symmetrically in space, that is, towards the corners of a tetrahedron (see Fig. 5). It is further necessary that the microphones are effectively coincident at all audio frequencies to prevent phase anomalies and this is achieved at all but the highest frequencies.

The response of a figure-of-eight microphone in the direction (r, θ_1, ϕ_1) to a source at (r, θ_i, ϕ_i) is

$$\sin\phi_i \sin\phi_1 \cos(\theta_i - \theta_1) + \cos\phi_i \cos\phi_1$$

It may be recalled from Section 3 that the general cardioid response is given by

$$A + \text{figure-of-eight response}$$

This gives the general cardioid response in three dimensions as

$$A + \sin\phi_i \sin\phi_1 \cos(\theta_i - \theta_1) + \cos\phi_i \cos\phi_1$$

or

$$A + X\cos\theta_1 \sin\phi_1 + Y\sin\theta_1 \sin\phi_1 + Z\cos\phi_1 \quad (4)$$

Using Equation (4), we find that the responses of the four capsules of the sound field microphone, assuming each has a cardioid characteristic (i.e. $A = 1$) are

$$L_{FU} = 1 + \sqrt{1/3}\,[\sin\phi_i(\cos\theta_i + \sin\theta_i) + \cos\phi_i]$$

$$R_{FD} = 1 + \sqrt{1/3}\,[\sin\phi_i(\cos\theta_i - \sin\theta_i) - \cos\phi_i]$$

$$L_{BD} = 1 + \sqrt{1/3}\,[\sin\phi_i(-\cos\theta_i + \sin\theta_i) - \cos\phi_i]$$

and

$$R_{BU} = 1 + \sqrt{1/3}\,[\sin\phi_i(-\cos\theta_i - \sin\theta_i) - \cos\phi_i]$$

Applying matrix addition similar to that used for M, S, T and Q (see Equation (2)), we have

$$W = ½\,(L_{FU} + R_{FD} + L_{BD} + R_{BU}) = 2$$

$$X = ½\,(L_{FU} + R_{FD} - L_{BD} - R_{BU}) = 2\sqrt{1/3}\cos\theta_i \sin\phi_i$$

$$Y = ½\,(L_{FU} - R_{FD} + L_{BD} - R_{BU}) = 2\sqrt{1/3}\sin\theta_i \sin\phi_i$$

$$Z = ½\,(L_{FU} - R_{FD} - L_{BD} + R_{BU}) = 2\sqrt{1/3}\cos\phi_i$$

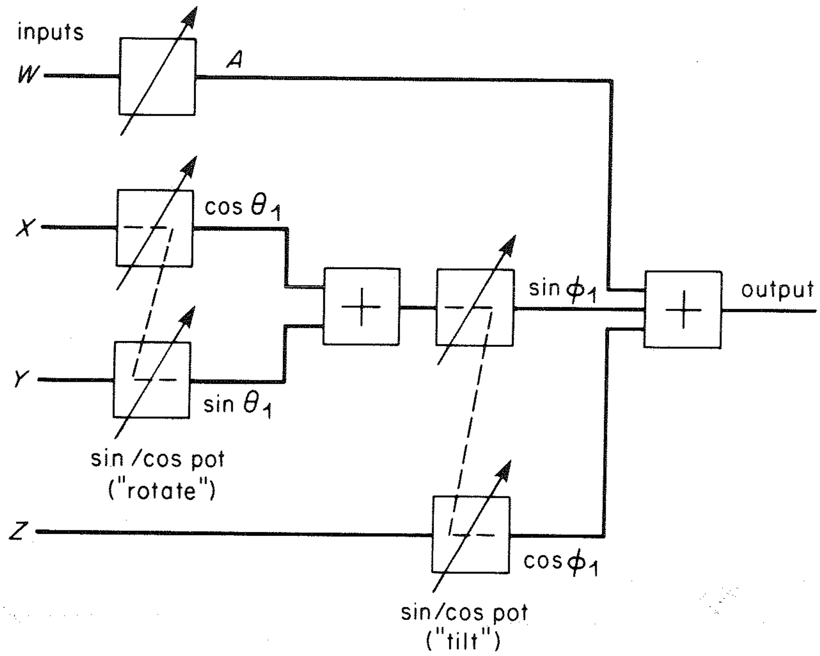
These signals form the basis for all subsequent processing. From W, X, Y and Z, any cardioid characteristic directed along any direction (r, θ_1, ϕ_1) may be synthesised (see Equation (4)) according to

$$A + \sin\phi_1 [X\cos\theta_1 + Y\sin\theta_1] + Z\cos\phi_1$$

where A is proportional to W.

* National Research and Development Corporation

Fig. 6
Cardioid microphone in the direction (r, θ_1, ϕ_1)



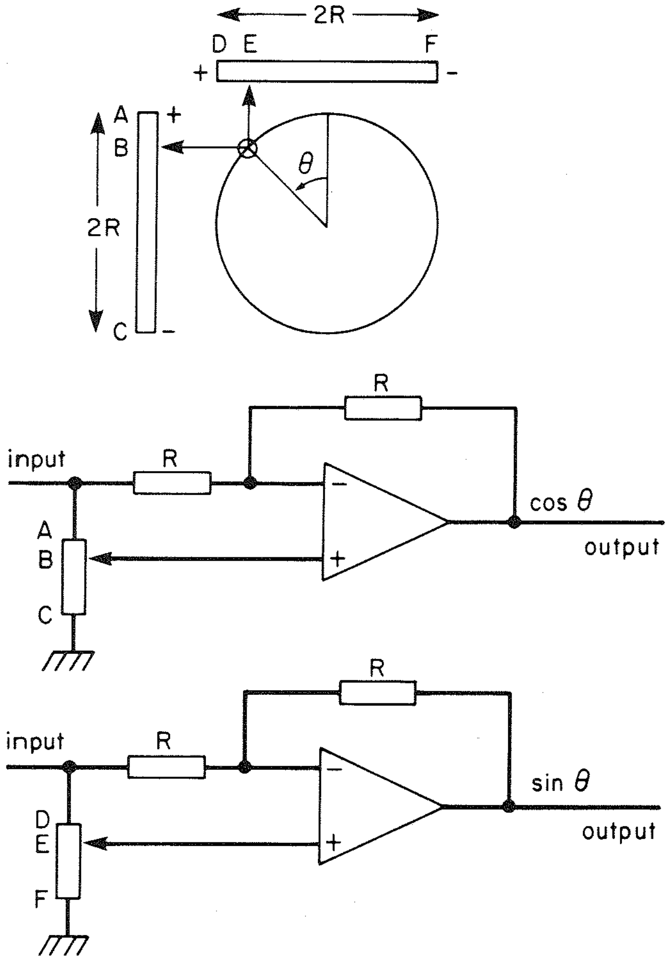


Fig. 7 - Sine/cosine potentiometer

Fig. 6 shows a circuit for generating this response from the signals *W, X, Y* and *Z*. Sin/cos potentiometers (see Fig. 7) are used and are described in more detail in Reference 11. To synthesise a pair of orthogonal cardioid microphones in the horizontal plane, directed along (*r*, ± 45°, 90°) for example,

$$A = 1$$

$$\theta_1 = \pm 45^\circ$$

and

$$\phi_1 = 90^\circ$$

The result is independent of the *Z* component.

Having generated *W, X, Y* and *Z* from the sound field microphone, one may effectively tilt and rotate the entire array using the following matrix

$$\begin{bmatrix} W' \\ X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta_1 \cos\phi'_1 & \sin\theta_1 \cos\phi'_1 & \sin\phi'_1 \\ 0 & -\sin\theta_1 & \cos\theta_1 & 0 \\ 0 & -\cos\theta_1 \sin\phi'_1 & -\sin\theta_1 \sin\phi'_1 & \cos\phi'_1 \end{bmatrix} \begin{bmatrix} W \\ X \\ Y \\ Z \end{bmatrix}$$

where θ_1 = angle by which the array is rotated anticlockwise about the *z*-axis and ϕ'_1 = angle by which the array is elevated about the current y' (or left/right) axis.

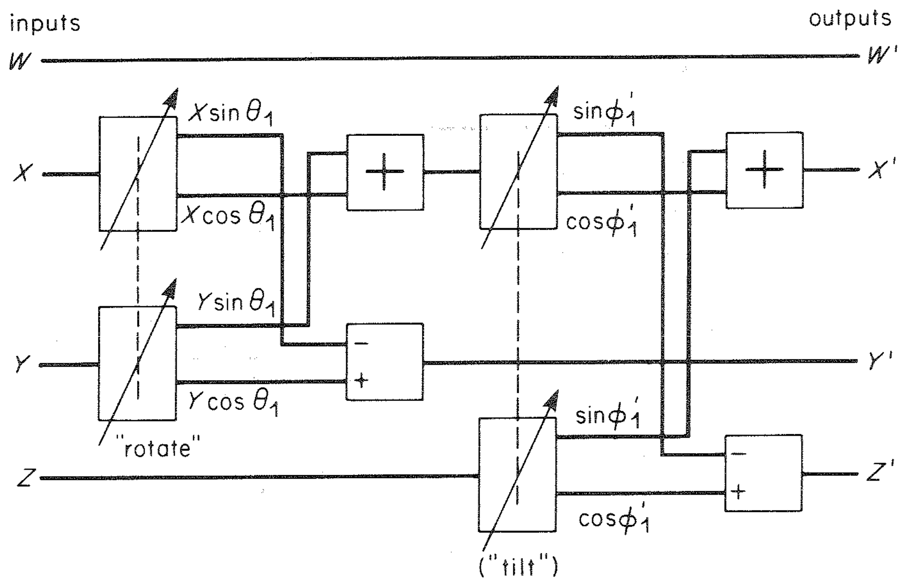


Fig. 8
"Rotate" and "tilt" control of W, X, Y, Z signals.

Fig. 8 shows means of realising this matrix. In addition a variable gain of the W signal would allow remote selection of the directional characteristics of the "microphones". Further processing of the W, X, Y and Z signals may be carried out readily to vary the stereo width and other parameters[11,12].

5.3 W, X, Y, Z pan-pots

Based on principles similar to those just discussed, two dimensional and three dimensional pan-pots that generate W, X, Y and Z signals may be formulated[11,12]. The magnitudes of the first order azimuthal components due to a source at (r, θ_i, ϕ_i) are given by

$$W = \text{constant}, B$$

$$X = \cos\theta \sin\phi$$

$$Y = \sin\theta \sin\phi$$

and

$$Z = \cos\phi$$

To simulate a source at this position, given a mono signal, the same signals should be generated. In the horizontal plane $\phi = 90°$ and Fig. 9 shows one method of realising a horizontal pan-pot. Fig. 10 shows a realisation of a three dimensional pan-pot.

When $W = 1/3$, these pan-pots give notionally "exterior panning", i.e. the source is panned along the perimeter of the stage (this depends to some extent on the decoding method adopted). In order to pan inwards, W is increased, although it is more practical to reduce the level of the X, Y and Z signals. When X, Y and Z equal zero, the centre position is simulated.

6. Routing module

In a mixing desk with quadraphonic capability, each input channel (routing module) has panning arrangements that allow the source to be panned to any azimuth in the horizontal plane. Conventionally, these give L_F, R_F, L_B and R_B outputs,

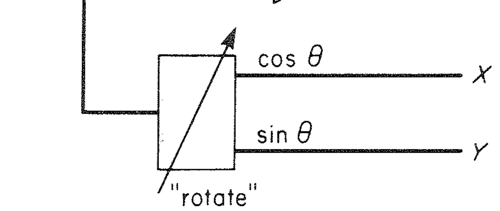


Fig. 9 - Horizontal W, X, Y pan-pot

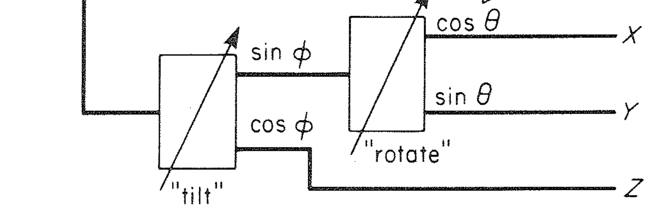


Fig. 10 - Three dimensional W, X, Y, Z pan-pot

but if only first order components are required, the surround sound signals may be represented by W, X and Y (see Sections 2 and 3) thereby saving a channel. Panning arrangements based on those described in Sections 5.3 and Fig. 9 may be used for this purpose.

A mixing desk based on these W, X, Y routing modules has W, X and Y group outputs. This may be followed by a 2-channel System HJ encoder which accepts W, X and Y input signals. Under certain circumstances, it may be desirable to encode points corresponding to positions given by pair-wise panning. As was seen in Section 4, three signals, W, X and Y, are not normally sufficient to describe such positions. However, given a particular 2-channel encode matrix, it is possible to simulate them in the following way. Since the absolute levels and relative phase fully describe the two encoded output signals, three variables are sufficient to describe them. Thus, the three variables W, X and Y (which are real) are also sufficient to describe and simulate any position on the Scheiber sphere. Analytic equations have been derived (see Appendix 1) to determine these values. Values of W, X and Y that simulate the corner positions of pair-wise panning for the 13LP2 locus (within the System HJ specification) have been determined. These are given in Appendix 1 together with the W, X, Y, 13LP2 encode equations.

One result of this process is that the absolute phase of the encoded output signals is normally different to that of the standard 13LP2 locus given by pair-wise panning. This implies a W, X, Y "pair-wise" pan-pot with switchable positions because a continuous version would be inordinately complex.

7. W, X, Y, Z signal handling.

It is often convenient to transfer and record signals in W, X, Y, Z form and since the gain of the W signal may be chosen arbitrarily (except for zero), standardisation is needed. The main criterion is signal level. If, for a single source, W, X, Y and Z take the form B, $\cos\theta_i \sin\phi_i$, $\sin\theta_i \sin\phi_i$ and $\cos\phi_i$ respectively, then X, Y and Z peak to unity and B* may be determined by one of the following criteria:

(1) If the criterion is that, for a single panned source, W, X, Y and Z should all peak to the same figure, then B should be set to unity.

(2) If there is an equal probability that a source may be at any position in the horizontal plane (neglecting interior panning) and if there are no sources off the horizontal plane, then, taking a long term or ensemble average,

$$B \text{ should be } \frac{1}{\sqrt{2}}$$

(3) If there is an equal probability that a source will be placed at any position in three dimensions (neglecting interior panning), then

$$B \text{ should be } \frac{1}{\sqrt{3}}$$

It is not clear which of these options is the most appropriate because of the complex nature of real programme signals. At worst, a loss of 7 dB of signal-to-noise may result.

* Note that the constant B need not be equal to the constant A mentioned in Sections 3 and 5.

8. Transmission

Up to this point, the handling and processing of signals closely related to the original sound-field have been discussed. These original signals may under modifications by being mixed with others (e.g. from other microphone inputs or reverberation plate returns) and the resultant "mix" represents the "programme" sound field. Additionally, the signals may be encoded by a complex matrix from four to two channels whereby phase information is introduced.

It has been common practice, particularly before the days of quadraphonic matrix encoding, to use the symbols M and S to describe the sum and difference signals of the left and right stereo signals, thus:

$$M = \tfrac{1}{2}(L + R)$$

$$S = \tfrac{1}{2}(L - R)$$

These are the signals that are multiplexed for stereo transmission. With normal stereo they relate closely to the M and S channel signals of the "programme" sound field. In the case of matrix quadraphony, however, this is only true in the simplest sense because of the phase-difference that is introduced between the signals. It is convenient therefore to define new symbols[1] to describe the transmission sum and difference signals:

$$\Sigma = \text{sum}$$

and $$\Delta = \text{difference}$$

Third and fourth terms $\widetilde{T}$ and $\widetilde{Q}$ are also sometimes used to describe third and fourth signals to be multiplexed[13].

9. Conclusions

This Report has shown how spherical (and azimuthal) harmonic analysis can help to delineate the essential elements of the recording, transmission and reproduction of surround sound. Theoretically ideal systems have been described based on results derived using this form of analysis. Means of recording that conform to these ideal systems have been developed, namely the "Soundfield" microphone, which samples the first order characteristics at a point in a sound field, and the *W, X, Y* pan-pots.

It is argued that the familiar and widespread studio technique of pair-wise panning is an anomalous means of achieving high order resolution in the reproduced sound stage, given a limited number of transmission channels and loudspeakers. Sharp images are only achieved by taking advantage of the deficiencies of the decoding. Nevertheless, in quadraphonic production, the technique has been shown empirically to be successful and valuable.

With current developments of surround sound transmission, the weakest link in the chain is probably that of decoding. With a limited number of loudspeakers present, only approximations to "ideal" decoding have been achieved and these rely on the inadequately understood properties of hearing. Considerable scope for research presents itself in this field.

10. References

1. COOPER, D.H. *and* SHIGA, T. Discrete-matrix multichannel stereo. *JAES,* **20**, 5, June 1972.

2. GIBSON, J.J., CHRISTENSEN, R.M. *and* LIMBERG, A.L.R. Compatible FM broadcasting of panoramic sound. *JAES,* **20**, 10, December 1972.

3. GERZON, M.A. Periphony: with-height sound reproduction. *JAES*, **21**, 1, January/February 1973.

4. Stereophonic broadcasting. BBC Technical Instruction p.4, February 1969.

5. GASKELL, P.S., *and* RATLIFF, P.A. Quadraphony: developments in Matrix H decoding. BBC Research Department Report No. 1977/2.

6. GERZON, M.A. Psychoacoustic criteria for the design of matrix and discrete quadraphonic systems (private communication).

7. GERZON, M.A. Design of Ambisonic decoders for multispeaker surround sound. Presented at the 58th AES Convention, New York, November 1977.

8. GERZON, M.A. Encoding standards for NRDC Universal HJ surround-sound encoding system: "System UHJ". Submitted as document NRDC/FCC2 to the FCC Notice of Inquiry on FM Quadraphonic Broadcasting, Docket No. 21310, 8th December 1977.

9. HARWOOD, H.D. An investigation of the law of displacement and of width of a stereophonic image for male speech. BBC Research Department Report No. 1967/24.

10. GERZON, M.A. The design of precisely coincident microphone arrays for stereo and surround sound. Presented to the 50th AES Convention, London, March 1975.

11. GERZON, M.A. NRDC Ambisonic Technology. Report No. 3. Panpot and Sound Field Controls. August 1975 (Revised June 1976).

12. GERZON, M.A. Ambisonics Part Two: Studio Techniques. Studio Sound **17**, 8, August 1975.

13. GERZON, M.A. FM Multiplex broadcasting of 2/2½/3 channel surround sound. Submitted as document NRDC/FCC7 to the FCC Notice of Inquiry on FM Quadraphonic Broadcasting, Docket No. 21310, 8th December, 1977

Appendix 1
W, X, Y 13LP2 (HJ) Encoding and panning

The 13LP2 matrix encoding equation for *W, X, Y* inputs is

$$\begin{pmatrix} L_T \\ R_T \end{pmatrix} = \begin{pmatrix} 0.6666\underline{/-14.6^\circ} & 0.3852\underline{/\ 73.8^\circ} & 0.6370\underline{/\ -9.1^\circ} \\ 0.6666\underline{/\ 14.6^\circ} & 0.3852\underline{/-73.8^\circ} & 0.6370\underline{/-170.9^\circ} \end{pmatrix} \begin{pmatrix} W \\ X \\ Y \end{pmatrix} \quad \text{(A1)}$$

where

$$W = 1$$

$$X = \cos\theta_i$$

$$Y = \sin\theta_i$$

and where θ_i = source azimuth, measured anti-clockwise from C_F.

With this encoding, if we wish to simulate a particular microphone and source combination that gives a known combination of L_T and R_T (as for example for pair-wise panning), then the values of *W, X* and *Y* may be found in the following way. We will make use of the fact that, for matrix encoding, it is necessary to maintain interchannel phase integrity, but not that of absolute phase. The absolute levels of L_T and R_T must, however, be correctly reproduced.

Equation (A1) is of the form:

$$\begin{pmatrix} L_T \\ R_T \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} W \\ X \\ Y \end{pmatrix}$$

or

$$L_r = a_{11r}\, W + a_{12r}\, X + a_{13r}\, Y$$

$$L_i = -a_{11i} W + a_{12i} X - a_{13i}\ Y$$

$$R_r = a_{11r}\, W + a_{12r}\, X - a_{13r}\, Y$$

$$R_i = a_{11i}\, W - a_{12i}\, X - a_{13i}\ Y$$

where L_r, L_i and R_r, R_i are the real and imaginary parts of L_T and R_T respectively and a_{pqr}, a_{pqi} are the moduli of the real and imaginary parts of a_{pq}.

If we advance the phases of L_T and R_T by an angle η whereby

$$L'_r = L_r \cos\eta - L_i \sin\eta$$

$$L'_i = L_i \cos\eta + L_r \sin\eta$$

and similarly for R'_r and R'_i, we find that in order to satisfy the given equations,

$$\tan\eta = \frac{a_{13i}\,(L_r - R_r) + a_{13r}\,(L_i + R_i)}{-a_{13r}\,(L_r + R_r) + a_{13i}\,(L_i - R_i)}$$

W, X and *Y* are then given by

$$W = \frac{a_{12i}\,(L'_r + R'_r) - a_{12r}\,(L'_i - R'_i)}{2(a_{11r}\,a_{12i} + a_{12r}\,a_{11i})}$$

$$X = \frac{a_{11i}\,(L'_r + R'_r) + a_{11r}\,(L'_i - R'_i)}{(a_{11r}\,a_{12i} + a_{12r}\,a_{11i})}$$

$$Y = \frac{L'_r - R'_r}{2a_{13r}}$$

The values of *W, X* and *Y* that, when encoded according to Equation (A1), simulate the corner positions given by pair-wise panning, are shown in Table A1.

Table A1 – 13LP2 Pair-wise panning

Source Position	W	X	Y
L_F	0.7039	0.7660	0.6750
R_F	0.7039	0.7660	−0.6750
L_B	0.7135	−0.7765	0.6889
R_B	0.7135	−0.7765	−0.6889

Appendix 2
Decoding and reproduction

A number of important aspects of decoding are discussed in Sections 2 and 4. This Appendix briefly quotes results that are developments of the decoding methods already described.

As stated in Section 4, one philosophy of decoding[7] dictates that the signal feeds to loudspeakers in a rectangular layout should be

$$L'_F = \tfrac{1}{2}\left(kW + \frac{X}{\sqrt{2}\cos\psi} + \frac{Y}{\sqrt{2}\sin\psi}\right)$$

$$R'_F = \tfrac{1}{2}\left(kW + \frac{X}{\sqrt{2}\cos\psi} - \frac{Y}{\sqrt{2}\sin\psi}\right)$$

$$L'_B = \tfrac{1}{2}\left(kW - \frac{X}{\sqrt{2}\cos\psi} + \frac{Y}{\sqrt{2}\sin\psi}\right)$$

$$R'_B = \tfrac{1}{2}\left(kW - \frac{X}{\sqrt{2}\cos\psi} - \frac{Y}{\sqrt{2}\sin\psi}\right)$$

where the loudspeakers are at azimuths $\pm\psi$ and $180° \pm\psi$ and k is a constant. If W, X and Y are such that the ratio of W to W_{ideal} is $\sqrt{2}$, one simple decoding method is given when $k = 1$, although $k = \sqrt{2}$ also gives good results. A more sophisticated method that takes some account of the way in which the ear changes its method of localisation with frequency, is effected by putting $k = 1$ at frequencies less than about 400 Hz and $k = \sqrt{2}$ at higher frequencies. An overall gain of 0.866 at the higher frequencies should also be included to preserve equal energy in the two frequency bands. This is referred to as "psycho-acoustic compensation"[6].

For the more general case of $n(\geqslant 4)$ loudspeakers equally distributed on the circumference of a circle, with the i'th loudspeaker at azimuth

$$360° \times \frac{i}{n},$$

the speaker feed to the i'th loudspeaker, omitting psychoacoustic compensation, should be

$$P_i = W + \sqrt{2}\,X\cos\psi_i + \sqrt{2}\,Y\sin\psi_i$$

This decoding approach is discussed in Ref. 7.

Pair-wise panning and the associated decoding has already been discussed in Section 4.1. An account of programme-dependent decoding is given in Ref. 2.

The simple and familiar stereo decoding is given by feeding to the left and right loudspeakers, signals

$$L' = \tfrac{1}{2}(W + Y) = \tfrac{1}{2}(M + S)$$

and $$R' = \tfrac{1}{2}(W - Y) = \tfrac{1}{2}(M - S)$$

It is possible that stereo decoding may be improved by employing frequency compensation as for surround reproduction. Known practical demonstrations of this technique, however, have not been successful[9].

Printed by BBC RESEARCH DEPARTMENT, Kingswood Warren, Tadworth, Surrey, KT20 6NP