**preprint no.** 1199 (A-5)

# THE OPTIMUM CHOICE OF SURROUND SOUND ENCODING SPECIFICATION

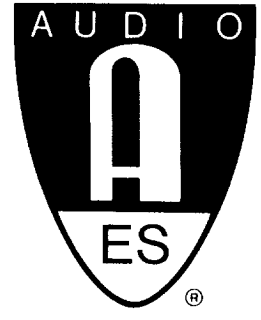
BY

M. Gerzon
Mathematical Institute
Oxford, England

**presented at the**
**56th Convention**
**March 1-4, 1977**
**Paris, France**


## AN AUDIO ENGINEERING SOCIETY PREPRINT





# The Optimum Choice of Surround Sound Encoding Specification

ABSTRACT:
The optimum choice of surround sound encoding specification with good mono and stereo compatibility is such that the best of all possible surround decoders for that system is better than the best of all possible decoders for any other system. This paper summarises psychoacoustic theories and methods that allow the discovery of this optimum system. This "System 45 J" is described, and is an improved refinement of the BBC 'matrix H' and UMX systems.

Michael A. Gerzon

Mathematical Institute, University of Oxford,

Oxford OX1 3LB, England

## I. Introduction

As well as ensuring good mono and stereo compatibility [1], it is also desirable that an encoding system for surround sound should be chosen such that the psychoacoustically best decoder for that encoding specification is better than the best decoder for any other specification. The present paper is a summary of the methods used to find this 'best-of-the-best' system , which we term System 45 J. (45 J is the development number assigned to this proposal out of the many alternative systems investigated).

To find optimum decoders and encoding systems, it is necessary to be able to describe mathematically both the properties of all encoding systems capable of being decoded with reasonable psychoacoustic results, and to have a reliable way of describing sound localisation by the ears capable of allowing the best methods of decoding to be deduced by mathematical reasoning. The full theory to do this will be the subject of a lengthy monograph in preparation, and the present paper is essentially just a sketch of some aspects of the method of deduction used.

This paper starts with a summary of various criteria involved in ensuring good sound localisation. These criteria should not be misunderstood as "theories of directional hearing", although it is true that when these criteria are satisfied, then most existing models of auditory sound localisation in the stereo literature predict correctly localised sound images. Rather, the idea is that the more of these criteria are satisfied, the better the quality and accuracy of the sound image is likely to be. Thus, although our criteria happen to involve some existing localisation models, notably Makita's theory, we make no assumption that Makita's theory itself always gives correct results.

The encoding systems considered are "kernel" systems, i.e. systems specified by giving channel gains as a function of the intended azimuth of each encoded sound. We shall only consider those horizontal-only systems whose channel gains are complex linear combinations of signals with directional

gains that are constant or equal to the sine or cosine of the azimuthal angle. This is because it may be proved that a reproduction of sound from a square or rectangle of loudspeakers in general satisfies the maximum number of the psychoacoustic criteria stated below only if the speaker feed signals are derived by a suitable matrix from just three such channels. In particular, a well-designed 3-channel rectangle decoder will outperform a rectangle of speakers fed with four non-redundant channels for all phantom image directions. Thus we do not pursue "quadraphonic" approaches, since these guarantee poor decoded results.

All the mathematical theorems and results quoted in this paper are given without proofs.

The BBC have designed an encoding system (termed "matrix H" [2]), which was soundly designed to achieve good mono and stereo compatibility. However, no comprehensive understanding of surround sound psychoacoustics was available when that system was formulated; thus its surround reproduction is suboptimal, and some designers have found a "variable matrix" decoder is required to rescue it. Although our methods permit decoder designs that do not require variable matrix "fixes" for the kernel version of matrix H (which we term system H, see Appendix B), the methods of this paper show that the best decoders for system H give results inferior to those obtainable with the optimal encoding system System 45 J. System 45 J is described in Appendix A, along with the design parameters of decoders at various levels of cost and performance. System 45 J may be decoded for surround sound via various shapes of loudspeaker layout from the two stereo "baseband" channels, from three channels, or from 2½ channels (i.e. from two channels plus a bandlimited third channel). The mono and stereo compatibility of System 45 J is good, as may be seen from the compatibility data in Appendix A. A comparison of Appendices A and B will reveal that the differences between systems H and 45 J are not large.

Encoding and decoding methods described in this paper have been subject to extensive experimental testing in connection with the Ambisonic surround sound project of the British NRDC.

## II. Some Basic Facts about Decoding

Consider x, y axes pointing respectively forward and to the left, and consider n loudspeakers situated on a circle around a central point in the directions $\phi_i$ (i = 1, 2, ... , n) measured anticlockwise from the x-axis. (We here ignore the effect of finite loudspeaker distance.) Then the following parameters influence psychoacoustic localisation if the loudspeaker is fed with a signal with complex gain $\alpha_i$

1. Makita localisation (low frequencies)

   Calculate $x = \mathrm{Re}\left(\frac{\Sigma\alpha_i \cos\phi_i}{\Sigma\alpha_i}\right)$

   $$y = \mathrm{Re}\left(\frac{\Sigma\alpha_i \sin\phi_i}{\Sigma\alpha_i}\right)$$

   then the apparent Makita sound direction $\theta_V$ is given

   by $\left.\begin{aligned} x &= r_V \cos\theta_V \\ y &= r_V \sin\theta_V \end{aligned}\right\}$ with $r_V > 0$

2. Velocity Magnitude condition (low frequencies)

   Given $r_V > 0$ defined as above, $r_V = 1$ is desirable for unambiguous localisation when the listener's head does not face the sound direction.

3. Phasiness condition (low frequencies)

   Calculate $q = \mathrm{Im}\left(\frac{\Sigma\alpha_i \sin\phi_i}{\Sigma\alpha_i}\right)$

   q is the "phasiness" heard by a forward facing listener, and for important sounds, it is desirable that $|q| < 0.21$, and advantageous that $|q| < 0.5$.

4. Energy Vector Direction (high frequencies)

Calculate $x_E = \frac{\Sigma|\alpha_i|^2 \cos \phi_i}{\Sigma|\alpha_i|^2}$

$$y_E = \frac{\Sigma|\alpha_i|^2 \sin \phi_i}{\Sigma|\alpha_i|^2}$$

and write $\left.\begin{aligned} x_E &= r_E \cos \theta_E \\ y_E &= r_E \sin \theta_E \end{aligned}\right\}$ with $r_E > 0$

Then $\theta_E$ is the apparent localisation of a sound at high frequencies when the listener faces the apparent source.

5. Energy Magnitude Condition (high frequencies)

Given $r_E > 0$ as above, ideally $r_E = 1$; in practice for important sound directions $r_E$ should be as large as possible, and $r_E > 0.5$ for important sound directions is desirable.

---

The basic theory of the psychoacoustic significance of the above criteria (except for "phasiness") is given in non-mathematical language in [3]. The significance of "phasiness" is discussed (for stereo reproduction) in [4]. The Makita localisation has been discussed (not always using Makita's name) repeatedly in the literature, e.g. see [5], [6].

Optimisation of the 5 localisation criteria above is not easy in terms of loudspeaker feed signals, but can be greatly simplified in the case of regular polygon loudspeaker layouts or rectangular loudspeaker layouts.

It is in practice important to ensure in the frequency region 250 Hz-1000 Hz, which represents the region in which it is not

clear whether the 'low' or 'high' frequency criteria above apply, that a suitable mixture of 'low' and 'high' frequency criteria should apply together. This is made possible by the following result:

RECTANGLE DECODER THEOREM For any rectangle decoder whose 4 speaker feed signals satisfy

$$-L_B + L_F - R_F + R_B = 0$$

the Makita localisation $\theta_V$ of a sound always coincides with the energy vector localisation $\theta_E$.

The above theorem has the following useful corollary.

COROLLARY 1 Let the 4 speaker feed signals $L_B$, $L_F$, $R_F$, $R_B$ of a rectangle decoder be related to 3 signals W, X, Y via

$$L_B = \tfrac{1}{2}(-X + W + Y)$$

$$L_F = \tfrac{1}{2}(X + W + Y)$$

$$R_F = \tfrac{1}{2}(X + W - Y)$$

$$R_B = \tfrac{1}{2}(-X + W - Y) ,$$

so that

$$X = \tfrac{1}{2}(-L_B + L_F + R_F - R_B)$$

$$W = \tfrac{1}{2}(L_B + L_F + R_F + R_B)$$

$$Y = \tfrac{1}{2}(L_B + L_F - R_F - R_B)$$

$$0 = \tfrac{1}{2}(-L_B + L_F - R_F + R_B) .$$

Then the Makita and energy vector localisations coincide, and lie in a direction $\theta$ given by

$$\left.\begin{aligned} \mathrm{Re}\,(XW^*) &= r\cos\theta \\ \mathrm{Re}\,(YW^*) &= r\sin\theta \end{aligned}\right\} \quad r > 0$$

where X, Y, W are also used to represent the complex channel gains of an encoded sound.

COROLLARY 2 ("Preference Theorem") For a rectangle decoder given as in Corollary 1, a new decoder with

$$\begin{pmatrix} W \\ X \\ Y \end{pmatrix} \quad \text{replaced by} \quad \begin{pmatrix} W \\ X \\ Y + jk\,W \end{pmatrix}$$

shares the same Makita and energy vector localisations as the original decoder, when k is real.

---

Comment The addition of the signal kjW to Y allows the phasiness of some directions to be reduced at the expense of increasing the phasiness of other encoded directions. If the phasiness of front is reduced at the expense of back, we term the modification of corollary 2 "forward preference". Note that forward preference does not alter the Makita or energy vector localisation or the velocity magnitude condition. Forward preference does in general modify the energy magnitude condition, and may be used to increase $r_E$ in the front at the expense of reducing $r_E$ at the back.

Forward preference also has the effect of changing the decoded energy gain in a manner varying with encoded direction. In particular, preference may be used to make the decoded directional gain more uniform in some encoding systems.

III. Limitations in 2-channel encoding

For an encoded azimuth $\theta$ (measured anticlockwise from front), a left/right symmetric 2-channel encoding system has left and right channel gains of the form

$$L = (a + jb) + (c + jd)\cos\theta + (f + je)\sin\theta$$

$$R = (a - jb) + (c - jd)\cos\theta + (-f + je)\sin\theta$$

with a, b, c, d, e, f real coefficients. The following facts may be proved, (provided a to f have not too "unreasonable" values).

(i) Decoders may be designed for such an encoding system satisfying to a high degree of approximation the Makita and energy vector localisation conditions with correct decoded azimuth and also satisfying (to a good accuracy) the velocity magnitude condition.

(ii) In general, such a decoder will not have uniform overall energy gain with direction.

(iii) It is not possible for such a decoder to have low phasiness for all decoded directions; in general a decoder satisfying Makita and energy vector localisation conditions and having a reasonable $r_E$ (condition (5)) will tend to have mean phasiness magnitude around $|q| = 0.5$, which is unacceptable for front centre sounds.

For a "symmetric circle encoding system" (i.e. one transformable to BMX [6] via a 2 x 2 matrix), the use of forward preference to reduce phasiness for front-encoded sounds has the unavoidable inevitable side-effect of increasing the energy gain of rear-encoded sounds (typically by 3 dB relative to front-encoded sounds). This is particularly objectionable since rear-encoded sounds are reproduced with increased phasiness.

There are, however circle-locus systems which are <u>not</u> transformable to BMX via a 2 x 2 matrix, and which are thus not rotationally symmetric. Such systems may incorporate a built-in reduction of 3 dB for rear encoded sounds to compensate for the boost caused by the use of forward preference. This 3 dB reduction may be hidden during stereo reproduction by means of a 2 x 2 matrix which has the effect of boosting back sounds by 3 dB. Such systems replace the 2 signals with directional gains 1 and $\cos\theta - j \sin\theta$ of BMX by signals with gains $1 + \frac{1}{2\sqrt{2}} j \sin\theta$ and $\cos\theta - \frac{3}{2\sqrt{2}} j \sin\theta$; the modified system has the same circle locus as BMX and a stereo gain almost uniform (within 0.51 dB) with direction, but has a widened reproduced stereo image. Similar modifications may be made to other symmetric circle encoding systems.

IV. System design philosophy

Any 2 channel encoding system should satisfy the following requirements (see [1] for a more detailed discussion):

(i) The front sector of the sound stage should reproduce in stereo with acceptable phasiness $Q < 0.45$ and with adequate stereo width.

(ii) The rear sector of the sound stage will reproduce in stereo with more phasiness, and thus on no account should it also be raised in level relative to the front sector.

(iii) It should be possible to add a bandlimited third channel permitting decoding with a flat frequency response and satisfying Makita and energy vector localisation conditions accurately (say with maximum error 2°) throughout the frequency range.

(iv) The 2-channel decoder should be of particularly good quality in the front sector; certainly not significantly inferior to the quality of good stereo, with which it is in competition.

(v) Localisation in 2- or 3-channel decode modes should be acceptable for all encoded directions and all listener orientations.

Note that condition (v) does not demand that the system be rotationally invariant, only that Makita and energy vector localisation conditions be satisfied. Indeed condition (iv) cannot be well satisfied along with condition (v) for a circle-symmetric system. Also conditions (i) and (ii) cannot all be satisfied for a circle symmetric system, and the front stage reproduction of a $2\frac{1}{2}$ channel circle symmetric system (e.g. TMX) is very poor at high frequencies due to phasiness effects.

The emphasis on getting front-stage sounds to be particularly good is not unreasonable, since the majority of musical uses of surround sound (as distinct from quadraphonic gimmickry) tend to have most important musical sources in this sector. This is true both

of classical and pop. For this reason, front-centre phasiness in stereo is much more objectionable than back-centre phasiness, and dictates an interchannel phase not much exceeding 45° for front sounds. Sounds outside the front sector are still adequately handled provided (v) holds, but it is felt that less weight should be given to stereo phasiness in the rear sector than in the front sector.

We again emphasise that decode possibilities are limited by the choice of encode equations, and that all possible decode options for given 2-channel encode equations have been mathematically classified in terms of the psychoacoustic theories given earlier. It is not possible to decode circle-symmetric 2-channel encodings (e.g. BMX) to give the optimum results from 2 channels ; a non-symmetric circle system is necessary for this.

v. Decode options for an encoding system

Let $\Sigma = L + R$, $\Delta = L - R$ be the sum and difference of left and right encoding channels, and let T be a third channel. The encoding kernel equations for azimuth $\theta$ anticlockwise from due front may be of the form

$$\begin{pmatrix} \Sigma \\ \Delta \\ T \end{pmatrix}_{\text{gain}} = \begin{pmatrix} a & c & je \\ jb & jd & f \\ jg & jh & i \end{pmatrix} \begin{pmatrix} 1 \\ \cos\theta \\ \sin\theta \end{pmatrix}$$

with real coefficients a , b, c, d, e, f, g, h, i; j represents a 90° phase shift.

The basic kernel decode equations are

$$\begin{pmatrix} W \\ X \\ Y \end{pmatrix} = \begin{pmatrix} \alpha & j\beta & j\gamma \\ \delta & j\varepsilon & j\zeta \\ j\chi & \psi & \omega \end{pmatrix} \begin{pmatrix} \Sigma \\ \Delta \\ T \end{pmatrix},$$

where $$\begin{pmatrix} \alpha & j\beta & j\gamma \\ \delta & j\varepsilon & j\zeta \\ j\chi & \psi & \omega \end{pmatrix} = \begin{pmatrix} a & c & je \\ jb & jd & f \\ jg & jh & i \end{pmatrix}^{-1}.$$

In this case the coefficients $\alpha$, $\gamma$, $\varepsilon$, $\beta$, $\delta$, $\zeta$, $\chi$, $\psi$, $\omega$ are real, and the gains of the signals W, X, Y are

$$W_{\text{gain}} = 1, \quad X_{\text{gain}} = \cos\theta, \quad Y_{\text{gain}} = \sin\theta.$$

The signal feed to the speaker at azimuth $\phi$ will be

$$W + 2X\cos\phi + 2Y\sin\phi$$

for regular polygon loudspeaker layouts, having gain $1 + 2\cos(\theta - \phi)$ .

When the gain of the third channel is diminished (say by a factor $0 \leq t < 1$), the decode equations become

$$\begin{pmatrix} W_t \\ X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} \alpha & j\beta & j\gamma t \\ \delta & j\varepsilon & j\zeta t \\ j\chi & \psi & \omega t \end{pmatrix} \begin{pmatrix} \Sigma \\ \Delta \\ T \end{pmatrix}$$

and we require that the encoding coefficients a to i be chosen so that this decoder still satisfies the Makita conditions,

i.e. $\mathrm{Re}\ (X_t W_t^*) = r \cos \theta_t$

$\mathrm{Re}\ (Y_t W_t^*) = r \sin \theta_t$

with $\theta_t \cong \theta$ .

In fact, given a,b,c,d, e, f (i.e. the 2-channel encoding) it is possible to choose g, h, i so that $\theta_0 = \theta$ for $\theta = 0°, \pm 60°, \pm 120°, 180°$, and it transpires then that for all $\theta$, $| \theta_t - \theta |$ is normally less than 2° for all T-gains t. In other words, for all reasonable 2-channel encoding systems, the 3rd channel encode may be chosen so that the inverse 3-channel decoder still satisfies the Makita localisation condition to within an accuracy of about 2° even when the T-channel gain is diminished.

The general kernel decode equation for diminished third channel satisfying the Makita localisations conditions (including forward preference) is of the form

$$\begin{pmatrix} W' \\ X' \\ Y' \end{pmatrix} = \begin{pmatrix} k_1\alpha & k_1 j\beta & (k_1 j\gamma)t \\ k_2\delta & k_2 j\varepsilon & (k_2 j\zeta)t \\ k_2 j\chi - \frac{1}{2}k_3 j\alpha & k_2\psi + \frac{1}{2}k_3\beta & (k_2\omega + \frac{1}{2}k_3\gamma)t \end{pmatrix} \begin{pmatrix} \Sigma \\ \Delta \\ T \end{pmatrix}$$

with real quantities $k_1$, $k_2$, $k_3$ and t. It is possible in the case of circle-locus systems (not necessarily rotationally invariant ones) to juggle the values of $k_1$, $k_2$, $k_3$ so that as the third channel gain t changes.

(i) Makita localisation and energy vector localisation stay true,

(ii) energy gain stays approximately uniform with direction and frequency,

(iii)depending on the frequency, whichever localisation criterion of (2)

and (5) is most apt holds, and

(iv) if the encode system is suitably chosen, the front phasiness automatically is lower than the back phasiness.

The above decode equation is the most general left/right symmetric one satisfying requirement (i) above; the computation and implementation of the coefficients $k_1$, $k_2$, $k_3$ is tedious but routine provided the balance of psychoacoustic requires (1) - (5) is decided upon. System 45 J was designed using the above requirements

a) The 2-channel encode system was chosen both for stereo and mono compatibility and for good 2-channel decode performance.

b) The 3rd channel was then computed to ensure that diminished T-gain did not upset localisation in decoding, and

c) The coefficients $k_1$, $k_2$, $k_3$ have been determined for a variety of decoding options.

By these means the advantages of TMX [6] have been retained, but with improved front-sector performance in system 45 J .

## VI. Interconvertibility of 2-channel encodings

In investigating the possible methods of 2-channel and $2\frac{1}{2}$ channel encoding and decoding, there is no point in distinguishing 2-channel systems that are interconvertible (i.e. convertible from one to the other via a 2 x 2 matrix) as far as surround sound decoding goes. Also it is clear that a third channel suitable for use with one 2-channel system is equally suitable for use with any interconvertible 2-channel system. We state here (without mathematical proof) a simple test for interconvertibility. Let the 2-channel encoding equations be

$$\left.\begin{aligned} L + R &= a + c \cos\theta + ej \sin\theta \\ L - R &= bj + dj \cos\theta + f \sin\theta \end{aligned}\right\} .$$

Then two systems are interconvertible if and only if they have parameters u and v having the same value, where

$$u = \frac{cf + ed}{bc - ad}$$

$$v = -\left(\frac{be + af}{bc - ad}\right) .$$

For systems having full rotational symmetry (i.e. interconvertible with BMX or conjugate BMX), we have respectively

$$u = 0, \quad v = 1$$

or $u = 0, \quad v = -1.$

Thus to investigate encoding/decoding results, we effectively have at the encoding end only a 2-parameter family of systems. The "system J" family of systems has

$$u = -0.3536 = -1/\sqrt{8}$$

$$v = 1.0607 = 3/\sqrt{8} ,$$

and a system has a circle pan locus if and only if

$$v = \pm\sqrt{1+u^2} .$$

## VII. Some general comments

The detailed computation and implementation of non-rotationally invariant systems is somewhat tedious mathematically, but design procedures are now available to compute the 3rd-channel encoding coefficients and all decoder parameters. Only in the case when the 2-channel system has a circle locus will the $2\frac{1}{2}$-channel results have a substantially flat frequency response with direction (although we have at NRDC built a $2\frac{1}{2}$-channel encode/decode system based on the bent-locus BBC matrix H system; this has frequency response faults which theory shows cannot be eliminated.)

Designing systems with rotational symmetry (i.e. with decode results dependent only on the difference $\theta - \phi$ of the encode and decode azimuths) is suspect on 2 grounds.

(i) It is not based on psychoacoustic theory of any depth.

(ii) Those psychoacoustic parameters that in any case cannot be satisfied for all directions (e.g. phasiness) can with advantage be rendered less harmful in some directions from others by departing from full symmetry.

However, it is important that those psychoacoustic parameters that can be satisfied for all directions should be so satisfied. To this extent, a system must have a partial degree of 'rotational symmetry'; for example, it should not be assumed that a listener will necessarily always face in one direction or that loudspeakers will be in a fixed position.

Note that for non-regular polygon loudspeaker layouts (e.g. non-square rectangles), even for BMX or TMX one does not get correct results by feeding the speakers with signals dependent only on $\theta - \phi$. What is fed to a speaker depends not only on its position,

but also on the position of all the other loudspeakers. Here also, a theory based on simple symmetry breaks down in giving the wrong psychoacoustic results. In the 'rectangle shape' control for loudspeaker layout on ambisonic decoders [7], the adjustment is in the opposite direction to that suggested by rotational symmetry.

VIII. Choosing the third channel

The problem of choosing (for a given 2-channel encoding) a third channel encoding so that the inverse decoder still satisfies the Makita localisation requirement at azimuths 0°, ± 60°, ± 120°, + 180° after the third channel is removed - this is solvable.

The following formula gives the 3rd channel encoding for the 2-channel encoding

$$L + R = a + c \cos\theta + ej \sin\theta$$

$$L - R = bj + dj \cos\theta + f \sin\theta$$

in terms of the quantities u and v defined above. The 3rd channel encoding is given by:

$$T_{gain} = gj + hj \cos\theta - \sin\theta \quad ,$$

where:

$$h = v^{-1}\left\{\frac{4 + 3u^2}{4 - (u/v)^2}\right\}^{\frac{1}{2}} \quad ,$$

$$g = \frac{u(1+3v^2)}{4+3u^2} \times \frac{h^2}{1+vh} \quad .$$

This choice of third channel ensures that when the third channel is removed from the input of the resulting inverse decoder, the Makita localisation is exact at azimuths 0°, 180° and at 4 azimuths equal to ± 60° and ± 120°; the Makita localisation is exact for all azimuths for the special cases

(i) $v = \pm\sqrt{1 + u^2}$ ,

and (ii) $u = 0$, $v \neq 0$.

If the inverse decoder equation is, for speaker azimuth $\phi$ in a regular loudspeaker layout

$$P_\phi = (\alpha\Sigma + j\beta\Delta + j\gamma T)$$
$$+2(\delta\Sigma + j\epsilon\Delta + j\zeta T)\cos\phi$$
$$+2(j\chi\Sigma + \psi\Delta + \omega T)\sin\phi \ ,$$

then the following decoding equation for real constants $k_1$, $k_2$, $k_3$, $t$ (with $k_1$, $k_2 > 0$, $0 \leq t \leq 1$) also satisfies Makita's localisation theory

$$P_\phi = k_1(\alpha\Sigma + j\beta\Delta + j\gamma T)$$
$$+2k_2(\delta\Sigma + j\epsilon\Delta + j\zeta tT)\cos\phi$$
$$+2k_2(j\chi\Sigma + \psi\Delta + \omega tT)\sin\phi$$
$$+ k_3(-j\alpha\Sigma + \beta\Delta + \dot{\gamma}tT)\sin\phi \ .$$

It may be shown that the above kernel decoding equation is the most general one having left/right symmetry and substantially satisfying Makita's localisation theory, as well as having a velocity magnitude $r_V$ not varying significantly with the encoded azimuth.

Actually, these decoders show small departures from Makita's theory (giving localisation errors usually not exceeding 2°), but the differences are very small and in principle capable of correction with sufficiently elaborate circuitry.

Thus the study of all $2\frac{1}{2}$-channel encode/decode possibilities reduces, for each case $0 \leq t \leq 1$, to the study of the 4-parameter system determined by the variables $u$, $v$, $k_2/k_1$, $k_3/k_1$. This system is manageable and has been surveyed both analytically and numerically. Studies show that substantially flat frequency response for all directions cannot be obtained unless the 2-channel system uses a circle locus (i.e. $v = \pm\sqrt{1 + u^2}$). Decoder design involves just choosing $k_2/k_1$ and

$k_3/k_1$, once u,v and t have been fixed. One chooses $k_2/k_1$ to satisfy the velocity magnitude or energy vector magnitude conditions (respectively at low and at high frequencies), and $k_3/k_1$ is chosen either so as to minimise front-stage phasiness and/or to ensure a maximally uniform energy gain with encoded direction. Thus the quantities $k_2/k_1$ and $k_3/k_1$ are determined by psychoacoustic criteria, t is determined by the availability or otherwise of the third channel, and we have required $v = (1+u^2)^{\frac{1}{2}}$. Thus only u remains to be chosen. If it is important that front sounds be reproduced with low phasiness in surround reproduction, but that the reproduced energy vary not more than about 0.6 dB with encoded direction (so as to give good ambience reproduction and flat $2\frac{1}{2}$-channel decoder frequency response), then u should lie between -0.28 and -0.385, and for analytic convenience and its good results, we have chosen for System 45 J, $u = -1/\sqrt{8}$, $v = (9/8)^{\frac{1}{2}}$.

These values for an optimally-decodable system may be compared with the values

$u = -0.1700$, $v = 1.4731$

found for the kernel system H encoding system described in Appendix B. System H is based on the 2-channel system obtained when the BBC's "matrix H" encoding system [2] is fed with hypercardioid inputs with nulls 135° off-axis, but the specification as given in Appendix B is entirely the author's responsibility.

## IX. Conclusions

We have summarised the reasoning that leads to our proposal of System 45 J as an encoding standard both for 2-channel and for 3-channel use. Of course, those having different balances of priorities might arrive at marginally different proposals, but the range of options available is narrow if it is required that no reasonable use and recording philosophy should "drop out" or be impracticable. Most existing encoding systems obtain their advertised excellence in some respects by abandoning some desirable features completely. System 45 J appears to be the first "balanced" encoding system that handles all philosophies of sound production and all requirements of sound reproduction (mono, stereo, and surround) with reasonable results. That the resulting balanced compromise is a good one is suggested by the fact that different members of the sound industry have urged modifications that pull in precisely opposite directions!

While three channels is the optimum number for psychoacoustically good horizontal surround reproduction via domestically feasible speaker layouts (a regular heptagon is the simplest layout taking full advantage of 4 channels), a fourth channel Q can be added either to obtain the "speaker emphasis" effect whereby all sounds are drawn towards 4 speakers, or to achieve full-sphere with-height reproduction.

Much existing surround material is in the form of pairwise mixed 4-channel tapes. Pairwise mixing is psychoacoustically very suboptimal, and where possible, new material should be pan-potted with kernel pan-pots (see [7]), which actually have a simpler circuit than pairwise panpots. It is nevertheless possible to approximate the system 45 J specification from pairwise material when archive recordings make this necessary (see Appendix C).

System 45 J meets the often conflicting requirements of both broadcasting and disc recording. An additional

aspect of System 45 J that recommends it as a universal standard for encoding surround sound is that, by the standards of "quadraphonic" systems, decoders for various existing systems give reasonable results with 45 J recordings. Existing reproduction equipment for the UMX and RM systems has been found to give substantially correct localisation from System 45 J signals, although results are not as good as psychoacoustically optimised 45 J decoders. It is likely that existing owners of UMX or RM equipment will not hear any marked degradation when using System 45 J recordings. Thus a transition to a System 45 J standard would be relatively painless. In addition, by "quadraphonic" standards, the difference between the BBC "Matrix H" proposal and the 2-channel version of System 45 J is small (see Appendices A and B). Both systems have good mono and stereo compatibility, each being better than the other for some program material.

While System 45 J gives the best decoded results according to the methods of this paper, there is nothing to stop consumer equipment manufacturers from following other decoder design philosophies (e.g. variable matrix designs) if they see a demand for the particular qualities of reproduction associated with such designs. The adoption of a good encoding standard does not inhibit decoder innovations, although a suboptimal encoding standard would do so.

It can be shown that the tolerance of System 45 J to various normal recording or transmission faults (e.g. interchannel gain or phase errors) is good. Thus it seems that System 45 J is well suited to be a universal encoding standard suitable for broadcasting, disc and cassette in its 2-channel version, and suitable for both broadcasting and the high--quality subcarrier disc technology used for UMX [6] in its 3-channel version. In addition, it offers the prospect that three existing systems (UMX, RM, Matrix H) can be reduced to just one universal system. Also, because of the powerful and comprehensive theoretical methods summarised in this paper, there is the assurance that no essentially better horizontal system of encoding can be found, so that the standard will be a permanent one, not subject to further significant change.

Appendix A. Specifications for System 45 J.

System 45J encodes a sound with azimuth $\theta$ (measured anticlockwise from due front) into the channels L, R, T with gains as follows. Let $\Sigma = L+R$ and $\Delta = L-R$ ; then

$$\begin{pmatrix} \Sigma \\ \Delta \\ T \end{pmatrix}_{gain} = \begin{pmatrix} 0.9530 & 0.2554 & 0.0661\,j \\ -0.3029\,j & 0.8034\,j & 0.9593 \\ -0.1716\,j & 1.0000\,j & -1.0000 \end{pmatrix} \begin{pmatrix} 1 \\ \cos\theta \\ \sin\theta \end{pmatrix} .$$

The basic kernel decoding equation feeds a loudspeaker at azimuth $\phi$ within a regular polygon loudspeaker layout with the signal

$$\begin{aligned} P_\phi = {} & (\ 0.9857\ \Sigma + 0.1058\,j\Delta + \ 0.1667\,jT) \\ & +(\ 0.5228\ \Sigma - 1.0785\,j\Delta - \ 1.0000\,jT)\ \cos\phi \\ & +(\ 0.1846\,j\Sigma + 1.1148\ \Delta \ - 0.9428\,T\ )\ \sin\phi\ . \end{aligned}$$

This equation causes a sound encoded at azimuth $\theta$ to be decoded with gain $1 + 2\cos(\theta-\phi)$ through the loudspeaker at azimuth $\phi$. The general kernel decoding equation is of the form

$$\begin{aligned} P_\phi = {} & k_1(\ 0.9857\ \Sigma + 0.1058\,j\Delta + 0.1667\,jtT) \\ & + k_2(\ 0.5228\ \Sigma - 1.0785\,j\Delta - 1.0000\,jtT)\cos\phi \\ & + k_2(\ 0.1846\,j\Sigma + 1.1148\Delta \ - 0.9428\,tT\ )\sin\phi \\ & + k_3(-0.9857\,j\Sigma + 0.1058\Delta + 0.1667\,tT\ )\sin\phi\ , \end{aligned}$$

where $0 \leq t \leq 1$ is the attenuation of the third channel T, and where $k_1, k_2, k_3$ are positive numbers chosen to optimise the psychoacoustics of reproduction. The apparent sound azimuth reproduced by such a decoder according to Makita's theory of localisation agrees with the encoded azimuth to within about $2^\circ$. The coefficients $k_1, k_2, k_3, t$ may vary with frequency if desired.

For rectangular loudspeaker layouts with speaker azimuths $\phi$, $180^\circ-\phi$, $-180^\circ+\phi$, amd $-\phi$, the respective speaker feed signals should be $P_{90^\circ-\phi}$ , $P_{90^\circ+\phi}$ , $P_{-90^\circ-\phi}$ , and $P_{-90^\circ+\phi}$ , provided that the effect of loudspeaker distance is neglected.

The parameters $k_1$, $k_2$, $k_3$, t in the horizontal kernel decoding equations for System 45 J have different values according to the number of channels available, the desired complexity of the decoder, and whether account is taken of the frequency-dependence of human sound localisation. We list suitable values for various applications, although it must be realised that further research may suggest improved values.

* Basic 3-channel decoder

$k_1 = k_2 = t = 1$ , $k_3 = 0$.

* Psychoacoustically compensated 3-channel decoder

$k_1 = k_2 = t = 1$, $k_3 = 0$ at frequencies $F \ll 400$ Hz

$k_1 = 1.2247$, $k_2 = 0.8660$, $t = 1$, $k_3 = 0$ for $F \gg 400$Hz.

Basic 2-channel decoder $k_1 = k_2 = 1$, $t = k_3 = 0$.

* Basic 2-channel decoder with almost uniform directional gain

$k_1 = 1$, $k_2 = 1.15$, $k_3 = 0.3622$, $t = 0$.

* Psychoacoustically compensated 2-channel decoder $t = 0$ and:

$k_1 = 0.6592$, $k_2 = 1.2807$, $k_3 = 0.1545$ for $F \ll 400$ Hz,

$k_1 = k_2 = 1$, $k_3 = 0.4175$ for $F \gg 400$ Hz.

Basic "$2\frac{1}{2}$-channel" decoder

$k_1 = k_2 = t = 1$, $k_3 = 0$ at frequencies for which 3 channels are available,

$k_1 = k_2 = 1.1454$, $k_3 = 0$, $t = 0$ when only 2 channels are available.

* Intermediate "$2\frac{1}{2}$-channel" decoder with almost uniform directional gain

$k_1 = k_2 = t = 1$, $k_3 = 0$ when 3 channels are available,

$k_1 = k_2 = 1.2162$, $k_3 = 0.5077$, $t = 0$ when only 2 channels are available.

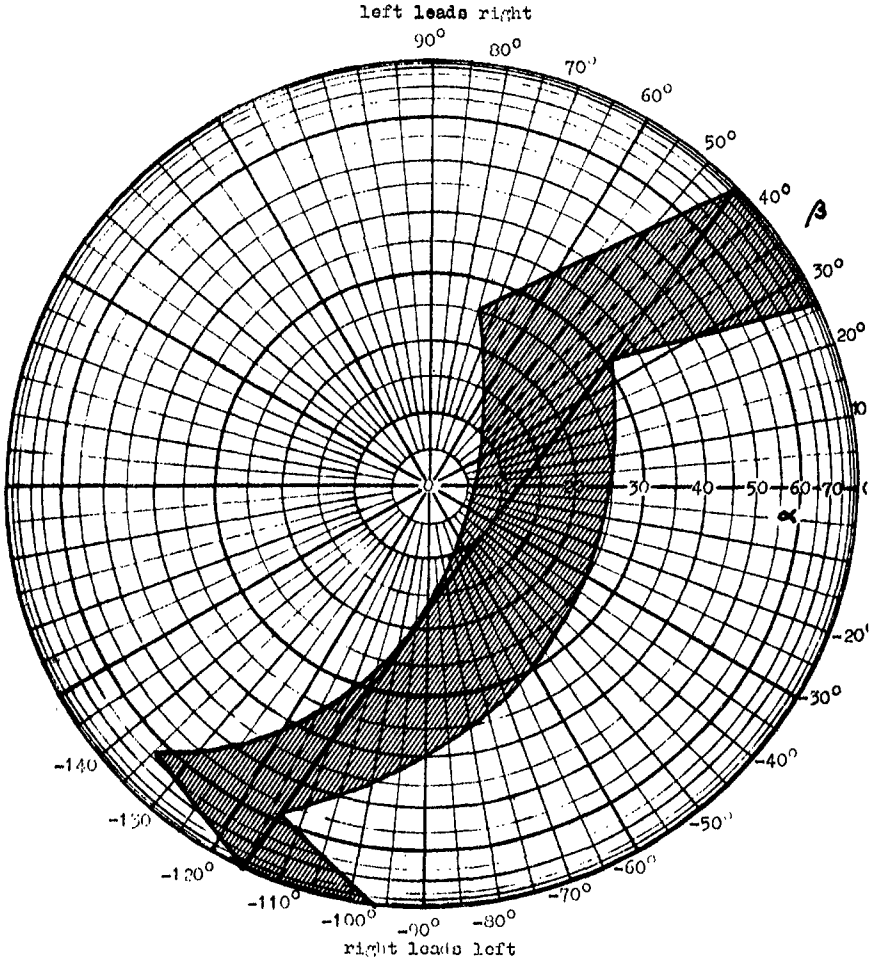
* Psychoacoustically compensated "$2\frac{1}{2}$-channel" decoder

$k_1 = k_2 = t = 1$, $k_3 = 0$ for frequencies $F \ll 400$ Hz,

$k_1 = 1.2247$, $k_2 = 0.8660$, $k_3 = 0$, $t = 1$ for $F \gg 400$ Hz when 3 channels are available,

$k_1 = k_2 = 1.2162$, $k_3 = 0.5077$, $t = 0$ at high frequencies when only 2 channels are available.

Notes: (i) The term "$2\frac{1}{2}$-channel" decoder indicates a decoder using 3 channels at lower frequencies and 2 at higher.
(ii) Decoders marked * have directional gain and frequency response uniform to within 0.51 dB variation.




System 45J encoding locus shown on Energy sphere (view from right side). Put left channel gain = $Le^{j\theta_1}$ and right channel gain = $Re^{j\theta_2}$ (with $L,R \geq 0$). Then:

$$\alpha = 2 \arctan(R/L) \quad \text{and} \quad \beta = \theta_1 - \theta_2.$$

——— System 45J kernel (optimal) encoding.

▨ Region of encoding for system 45 J pairwise mix encoding options.

Compatibility Table System 45 J encoding.

| Azimuth angle degrees | | Mono Gain dB | Stereo Gain dB | Position P | Phasiness Q |
|---|---|---|---|---|---|
| $C_F$ | 0 | 0.00 | 0.00 | 0.000 | 0.414 |
| | 22½ | -0.14 | 0.08 | 0.316 | 0.363 |
| $I_F$ | 45 | -0.55 | 0.26 | 0.607 | 0.209 |
| | 67½ | -1.20 | 0.44 | 0.841 | -0.045 |
| $C_L$ | 90 | -2.04 | 0.51 | 0.980 | -0.386 |
| | 112½ | -2.98 | 0.44 | 0.980 | -0.784 |
| $L_B$ | 135 | -3.87 | 0.26 | 0.807 | -1.176 |
| | 157½ | -4.53 | 0.08 | 0.460 | -1.474 |
| $C_B$ | 180 | -4.77 | 0.00 | 0.000 | -1.586 |
| Gain variation (dB) | | 4.77 | 0.51 | | |

Note: For left and right channel gains L and R respectively,

$$P = \mathrm{Re}\{(L-R)/(L+R)\}$$
$$Q = \mathrm{Im}\{(L-R)/(L+R)\} .$$

According to Makita's sound localisation theory, P is the proportional displacement from the midpoint along the line joining the stereo speaker pair. For a further discussion of the psychoacoustic significance of P and Q in other localisation theories, see ref. [4]. The System 45 J encoding locus and its compatibility properties have previously been discussed in [1], especially the locus e of its figure 1.

System interchannel phase difference (between L & R channels):

for $C_F$ encoded sound: 45.00°

for $C_B$ encoded sound: -115.53°

Appendix B. Specifications for system H, a 3-channel kernel version of the BBC 'Matrix H' encoding system.

System H encodes a sound with azimuth $\theta$ (measured anticlockwise from due front) into the channels L, R, T with gains as follows. Let $\Sigma = L+R$ and $\Delta = L-R$ ; then

$$\begin{pmatrix} \Sigma \\ \Delta \\ T \end{pmatrix}_{gain} = \begin{pmatrix} 0.9915 & 0.2030 & -0.1305j \\ -0.1305j & 0.6580j & 0.9915 \\ -0.0733j & 0.6873j & -1.0000 \end{pmatrix} \begin{pmatrix} 1 \\ \cos\theta \\ \sin\theta \end{pmatrix} .$$

The basic kernel decoding equation feeds a loudspeaker at azimuth $\phi$ within a regular polygon loudspeaker layout with the signal

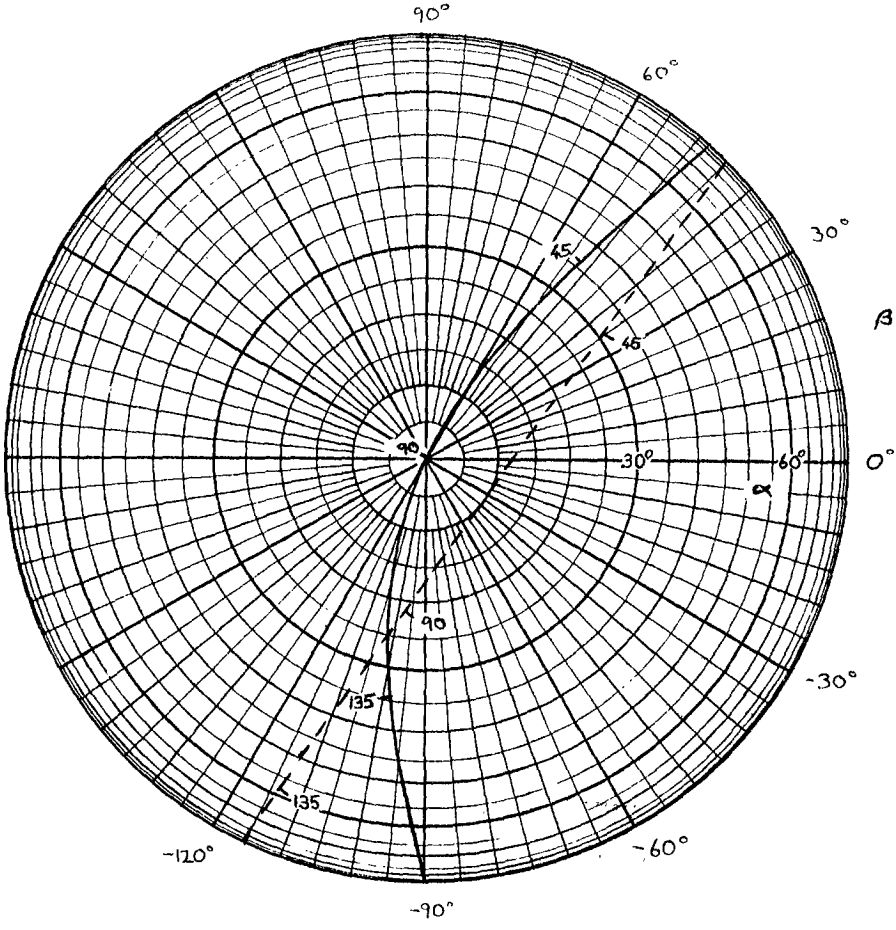
$$\begin{aligned} P_\phi = \; & ( 0.9744\,\Sigma + 0.2129j\Delta + 0.0839jT) \\ & +( 0.2956\,\Sigma - 1.4286j\Delta - 1.4549jT)\cos\phi \\ & +( 0.0603j\Sigma + 1.0131\,\Delta - 0.9877T )\sin\phi . \end{aligned}$$

This equation causes a sound encoded at azimuth $\theta$ to be decoded with gain $1 + 2\cos(\theta-\phi)$ through the loudspeaker at azimuth $\phi$. The general kernel decoding equation is of the form

$$\begin{aligned} P_\phi = \; & k_1( 0.9744\,\Sigma + 0.2129j\Delta + 0.0839jtT) \\ & + k_2( 0.2956\,\Sigma - 1.4286j\Delta - 1.4549jtT)\cos\phi \\ & + k_2( 0.0603j\Sigma + 1.0131\,\Delta - 0.9877tT )\sin\phi \\ & + k_3(-0.9744j\Sigma + 0.2129\,\Delta + 0.0839tT )\sin\phi , \end{aligned}$$

where $0 \leqslant t \leqslant 1$ is the attenuation of the third channel T, and where $k_1, k_2, k_3$ are positive numbers chosen to optimise the psychoacoustics of reproduction. The apparent sound azimuth reproduced by such a decoder according to Makita's theory of localisation agrees with the encoded azimuth to within about $\pm 1^\circ$. The coefficients $k_1, k_2, k_3, t$ may vary with frequency if desired.

For rectangular loudspeaker layouts with speaker azimuths $\phi$, $180^\circ-\phi$, $-180^\circ+\phi$, and $-\phi$, the respective speaker feed signals should be $P_{90^\circ-\phi}$ , $P_{90^\circ+\phi}$ , $P_{-90^\circ-\phi}$ , and $P_{-90^\circ+\phi}$ , provided that the effect of loudspeaker distance is neglected.




Energy sphere loci for kernel versions of:
System H (solid line) and System 45 J (dashed line),
showing in each case the encoded azimuths $45^{\circ}, 90^{\circ}, 135^{\circ}$.

Put left channel gain = $Le^{j\theta_1}$ and right channel gain = $Re^{j\theta_2}$ (with $L,R \geq 0$). Then:

$$\alpha = 2 \arctan(R/L) \quad \text{and} \quad \beta = \theta_1 - \theta_2 .$$

Sphere is shown viewed from right side.

Compatibility Table  System H kernel (optimal) encoding.

| Azimuth angle degrees | | Mono Gain dB | Stereo Gain dB | Position P | Phasiness Q |
|---|---|---|---|---|---|
| $C_F$ | 0 | 0.00 | 0.00 | 0.000 | 0.442 |
| | 22½ | -0.11 | 0.15 | 0.304 | 0.418 |
| $L_F$ | 45 | -0.41 | 0.47 | 0.590 | 0.343 |
| | 67½ | -0.91 | 0.72 | 0.833 | 0.207 |
| $C_L$ | 90 | -1.54 | 0.69 | 1.000 | 0.000 |
| | 112½ | -2.25 | 0.32 | 1.040 | -0.281 |
| $L_B$ | 135 | -2.92 | -0.35 | 0.893 | -0.605 |
| | 157½ | -3.42 | -1.05 | 0.527 | -0.886 |
| $C_B$ | 180 | -3.61 | -1.37 | 0.000 | -1.000 |
| Gain variation (dB) | | 3.61 | 2.12 | | |

Note: For left and right channel gains L and R respectively,

$$P = \mathrm{Re}\{(L-R)/(L+R)\}$$
$$Q = \mathrm{Im}\{(L-R)/(L+R)\} .$$

According to Makita's sound localisation theory, P is the proportional displacement from the midpoint along the line joining the stereo speaker pair. For a further discussion of the psychoacoustic significance of P and Q in other localisation theories, see ref. [4] . The system H encoding locus and its compatibility properties have previously been discussed in [1] , especially the locus c of its figure 1.

System interchannel phase difference (between L & R channels):

for $C_F$ encoded sound: $47.66^\circ$

for $C_B$ encoded sound: $-90.00^\circ$

Appendix C :

## SYSTEM 45 J ENCODING FROM PAIRWISE MIXED MATERIAL

In view of the poor subjective results normally obtained from pairwise mixed 4-channel material, the encoding used to approximate a System 45 J specification is necessarily a compromise. Although a fixed compromise (e.g. with k = 0.7071 below) may be used for all application, users may find that the range of options given below allows the results obtained from any given pairwise mixed program to be optimized.

Put $X = \frac{1}{2}(-LB+LF+RF-RB)$

$W = \frac{1}{2}(+LB+LF+RF+RB)$

$Y = \frac{1}{2}(+LB+LF-RF-RB)$

$Z = \frac{1}{2}(-LB+LF-RF+RB)$ .

Then for a chosen positive constant k (with $0.7071 \leq k \leq 1$), the following encoding may be used to approximate ystem 45 J "speaker emphasis" encoding from pairwise mixed program:

$$\Sigma = 0.9530\ W + 0.2554\ kX + 0.0661\ jkY$$

$$\Delta = -0.3029\ jW + 0.8034\ jkX + 0.9593\ kY$$

$$T = -0.1716\ jW + 1.0000\ jkX - 1.0000\ kY$$

Putting k = 0.7071 ensures that the "corner" azimuths ($\pm 45^{\circ}$, $\pm 135^{\circ}$) are encoded correctly according to the System 45 J kernel specification, whereas k = 1 ensures that the "cardinal" azimuths ($0^{\circ}$, $\pm 90^{\circ}$, $180^{\circ}$) are encoded correctly according to the kernel encoding equations. Intermediate values of k ensure that intermediate azimuths are encoded correctly. For example, k = 0.9239 ensures correct encoding for azimuths $\pm 22\frac{1}{2}^{\circ}$, $\pm 67\frac{1}{2}^{\circ}$, $\pm 112\frac{1}{2}^{\circ}$, $\pm 157\frac{1}{2}^{\circ}$.

If required, a more flexible range of encoding options for pairwise mixed material is obtained by using different values for k (denoted respectively by $k_F$ and $k_B$) for the front and back channels. This yields the following encoding equations:

$$\Sigma = 0.9530\ W' + 0.2554\ X + 0.0661jY$$

$$\Delta = -0.3029jW' + 0.8034jX + 0.9593\ Y$$

$$T = -0.1716jW' + 1.0000jX - 1.0000\ Y$$

where the signals X, Y are as defined above, and where

$$W' = \tfrac{1}{2}(k_B^{-1}LB + k_F^{-1}LF + k_F^{-1}RF + k_B^{-1}RB)\ ,$$

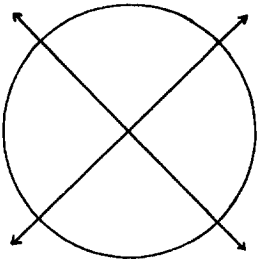
with $0.7071 \leq k_F \leq 1$ and $0.7071 \leq k_B \leq 1$ .

If $k_F = 0.7071$, the azimuths $\pm 45^\circ$ are encoded correctly according to the System 45J kernel specification, whereas if $k_F = 1$, the azimuth $0^\circ$ (due front) is encoded correctly. If $k_B = 0.7071$, then the azimuths $\pm 135^\circ$ are encoded correctly, and if $k_B = 1$, the azimuth $180^\circ$ (due back) is encoded correctly.

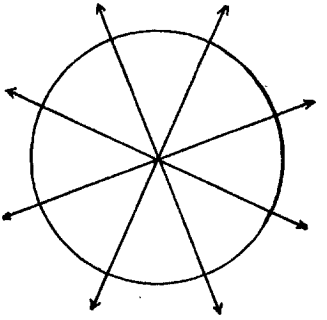
For the special case $k_F = 0.7071$, $k_B = 1$, the following azimuths are encoded correctly: $\pm 45^\circ$, $\pm 115.5^\circ$, $180^\circ$ , whereas for $k_F = 1$, $k_B = 0.7071$, the correctly encoded azimuths are $0^\circ$, $\pm 64.5^\circ$, $\pm 135^\circ$. For $k_F = k_B = k$ , the results are as described above.

An accompanying figure illustrates the azimuths that are encoded correctly according to System 45 J specifications for pairwise mixed material using various values of the parameters $k_F$ and $k_B$. We also show the energy sphere loci of various pairwise mix encoding options.

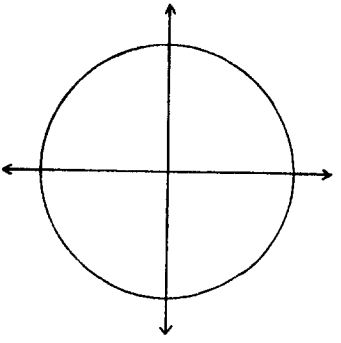
For pairwise mixed material, it is permissable to perform the above encodings with a gain for the rear channel signals LB, RB differing from that of the front signals LF, RF. The rear channels' gain may be between -3 dB and + 3 dB. An encoding option particularly suited to general use in broadcasting is $k_F = 1$, $k_B = 2^{-\frac{1}{2}}$ and rear gain = -1.25 dB, which ensures roughly uniform reproduced corner gains both in stereo and optimised 2-, $2\frac{1}{2}$- and 3-channel surround reproduction.



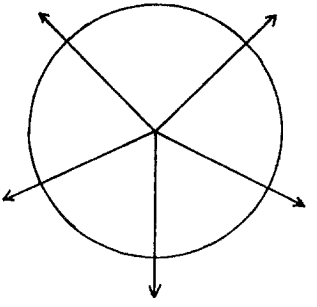
$k_F = k_B = k = 0.7071$



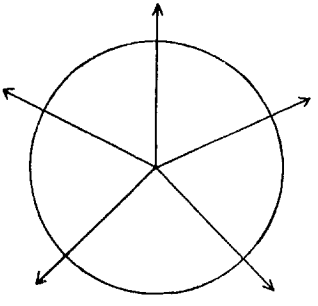
$k_F = k_B = k = 0.9239$



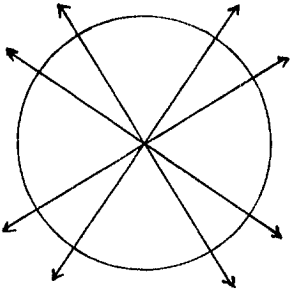
$k_F = k_B = k = 1.0000$



$k_F = 0.7071$, $k_B = 1.0000$

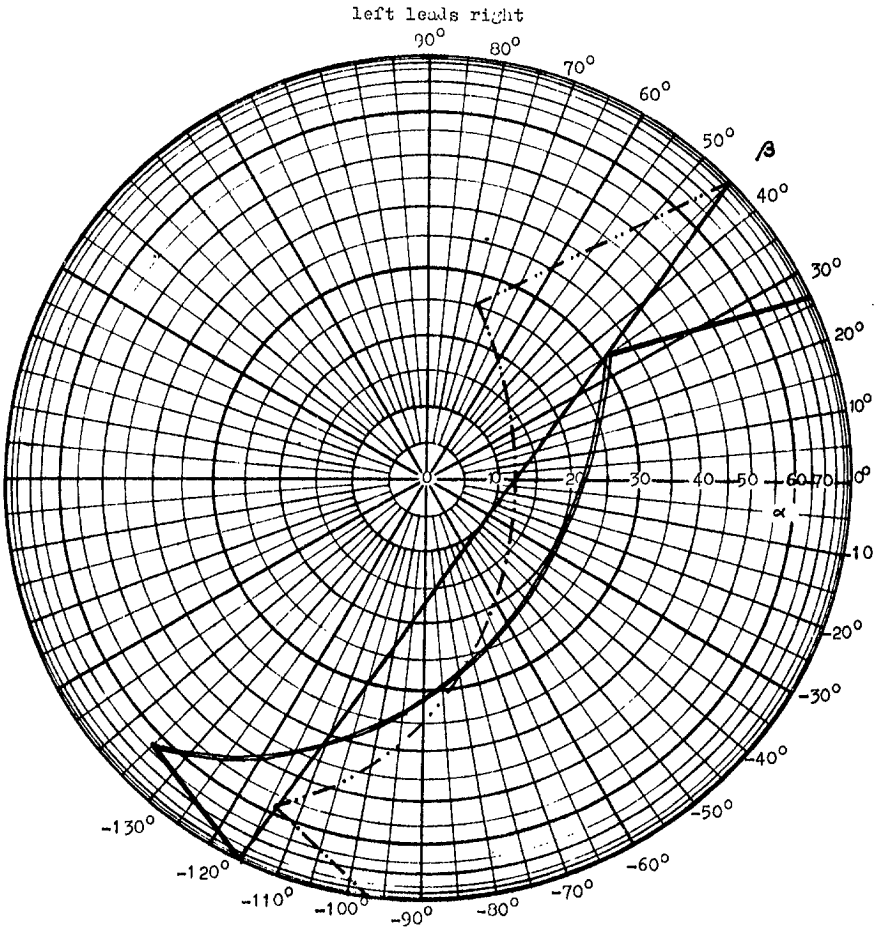


$k_F = 1.0000$, $k_B = 0.7071$



$k_F = k_B = k = 0.8409$

Showing those azimuths of pairwise mixed material that are encoded correctly according to System 4·J kernel encoding standards for various values of the coefficients $k_F$, $k_B$, $k$ in the pairwise mix encoding equations.



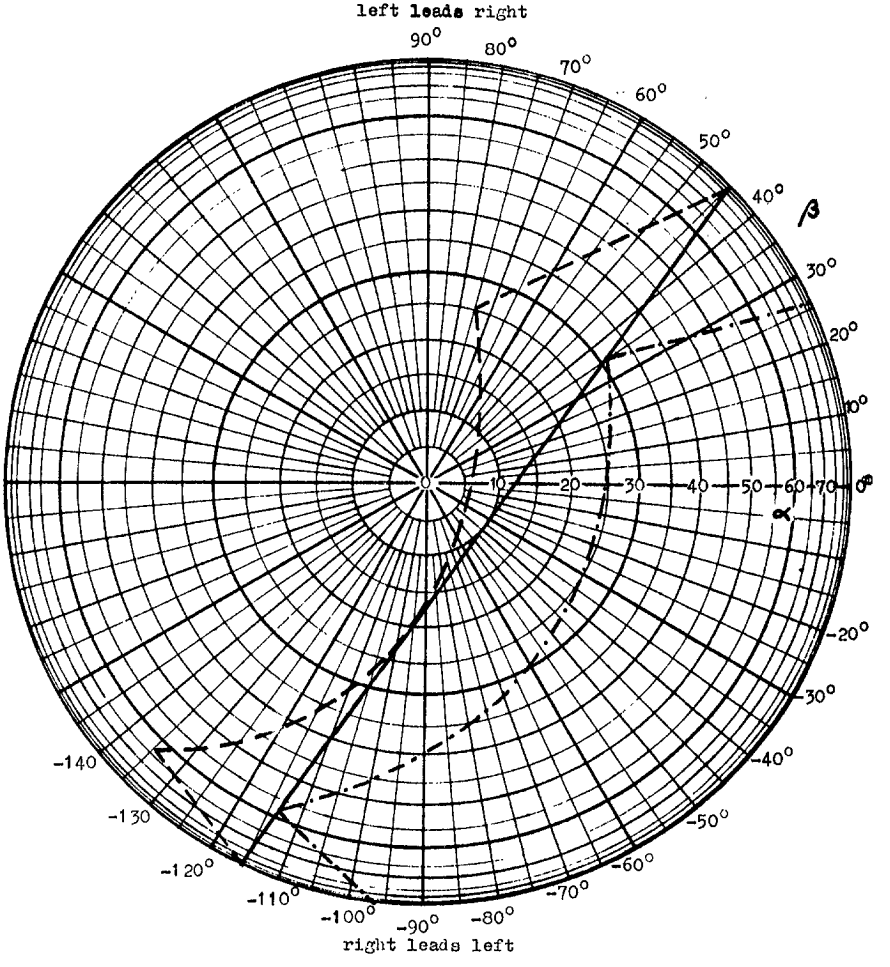

System 45J encoding locus shown on energy sphere (view from right side). Put left channel gain = $Le^{j\theta_1}$ and right channel gain = $Re^{j\theta_2}$ (with $L,R \geq 0$). Then:

$$\alpha = 2 \arctan(R/L) \quad \text{and} \quad \beta = \theta_1 - \theta_2.$$

——— System 45J kernel (optimal) encoding.

—··—··— System 45J pairwise mix encoding, $k_F=1$, $k_B= 0.7071$.

▬▬▬ System 45J pairwise mix encoding. $k_F= 0.7071$, $k_B=1$.




System 45J encoding locus shown on Energy sphere (view from right side). Put left channel gain = $Le^{j\theta_1}$ and right channel gain = $Re^{j\theta_2}$ (with $L,R \geq 0$). Then:

$$\alpha = 2 \arctan(R/L) \quad \text{and} \quad \beta = \theta_1 - \theta_2.$$

——— System 45J kernel (optimal) encoding.

—·—·— System 45J pairwise mix encoding, k = 0.7071 .

— — — — System 45J pairwise mix encoding, k = 1 .

References

[1]. M.A. Gerzon, "Compatible 2-Channel Encoding of Surround Sound", Electronics Letters, vol.11, pp.518-519 (11 Dec. 1975)

[2]. D.J. Meares and P.A. Ratliff, "The Development of a Compatible 4-2-4 Quadraphonic Matrix System, BBC Matrix H", E.B.U. Rev.-Technical, No.159, pp. (Oct. 1976)

[3]. M.A. Gerzon, "Surround Sound Psychoacoustics", Wireless World, vol.80, pp.483-486 (Dec. 1974)

[4]. M.A. Gerzon, "A Geometric Model for Two-Channel Four--Speaker Matrix Stereo Systems", J. Audio Eng. Soc., vol.23, pp.98-106 (Mar. 1975), especially Appendix II.

[5]. B. Bernfeld, "Simple Equations for Multichannel Stereophonic Sound Localization", J. Audio Eng. Soc., vol. 23, pp.553-557 (Sept. 1975)

[6]. D.H. Cooper and T. Shiga, "Discrete-Matrix Multichannel Stereo", J. Audio Eng. Soc., vol. 20, pp.346-360 (June 1972)

[7]. M.A. Gerzon, "Ambisonics, Part II. Studio Techniques", Studio Sound, vol.17 no.8, pp. 24, 26, 28-30 (Aug. 1975); correction, ibid. vol.17 no.10, p.60 (Oct. 1975)

ACKNOWLEDGEMENTS

I would like to thank my colleques in the NRDC ambisonic project, and most particularly Professor Peter Fellgett, for their considerable assistance and encouragement; also members of the BBC research dept. for useful discussions.