NRDC/FCC 2

SIMON BROWNE

Encoding Standards for NRDC Universal HJ Surround-Sound Encoding System : "System UHJ"

N.R.D.C. Ambisonic Technology. 1977 November 22

I. PREAMBLE

The enclosed proposal for a surround sound encoding standard (termed "System UHJ") is designed to be a general purpose standard with the following properties and applications:

(i) Good mono compatibility for all encoded directions.

(ii) Good stereo compatibility for a wide variety of recording philosophies.

(iii) Capability of surround sound decoding from 2 channels, 3 channels, 4 channels in a horizontal plane, 4 channel with height, 2 channels plus bandlimited third channel, and 2 channels plus bandlimited 3rd and 4th channel.

(iv) Suitability for broadcasting of music and non-musical program.

(v) Suitability for disc recordings either with or without subcarrier channels.

(vi) Good compatibility when reproduced over existing UD-4 system surround sound decoders, and acceptable compatibility when reproduced via existing Regular Matrix and speaker matrix reproducers.

(vii) Capability of optimum psychoacoustic results in surround reproduction via suitably designed horizontal-only or full-sphere decoders, on both pan-potted and ambient recordings.

(viii) Unambiguous specifications for encoding any intended directional effect, so as to enable decoder designs to be rationally optimised.

(ix) Recommendations for approximating the specification from pairwise mixed program material.

The present proposal for an encoding standard also includes suggested methods of decoding (which do not themselves form a part of the standard, but are given for reference purposes), and data on the various properties of the proposed standard.

REFERENCES

P.B. Fellgett and M.A. Gerzon, "Encoding standards for 2-channel surround-sound", Electronics Letters, 11, pp.518-9 (1975)

M.A. Gerzon, "Compatible 2-channel encoding of surround sound", Electronics Letters, 11, pp.615-7 (1975)

M.A. Gerzon, "The Optimum Choice of Surround-Sound Encoding Specification", Preprint No. 1199 (A-5), 56th Audio Engineering Society Convention, Paris, 1 March 1977.

II. SYSTEM HJ ENCODING TOLERANCES

A 2-channel surround-sound encoding system is said to meet System HJ encoding tolerances if the eight directions at 45° around the horizontal circle commonly denoted by CF, RF, CR, RB, CB, LB, CL, LF (where L = left, R = right, F = front, B = back, C = centre) are encoded within the tolerance regions shown in figure 1 and its caption. This figure displays the encoding in terms of the Schieber coordinates α and β, where

$$\alpha = 2\tan^{-1}|R/L|$$

$$\beta = \arg L - \arg R$$

where L and R denote the complex gains of the sound in the left and right encoded channels respectively, and where for real k and θ, $\arg(ke^{j\theta})$ denotes θ. System HJ defines encoding tolerances agreed by the BBC and NRDC for 2-channel encoding systems and for the basic 2 channels in three and four channel surround-sound encoding systems.

System UHJ, or "universal HJ encoding system", described in this report, is a precise encoding standard for all directions, and for 2, 3 or 4 channels, whose basic 2-channel signals satisfy the HJ encoding tolerances. The additional precision of the UHJ specification permits a precise specification of the encoding for additional channels enhancing directional resolution in reproduction.

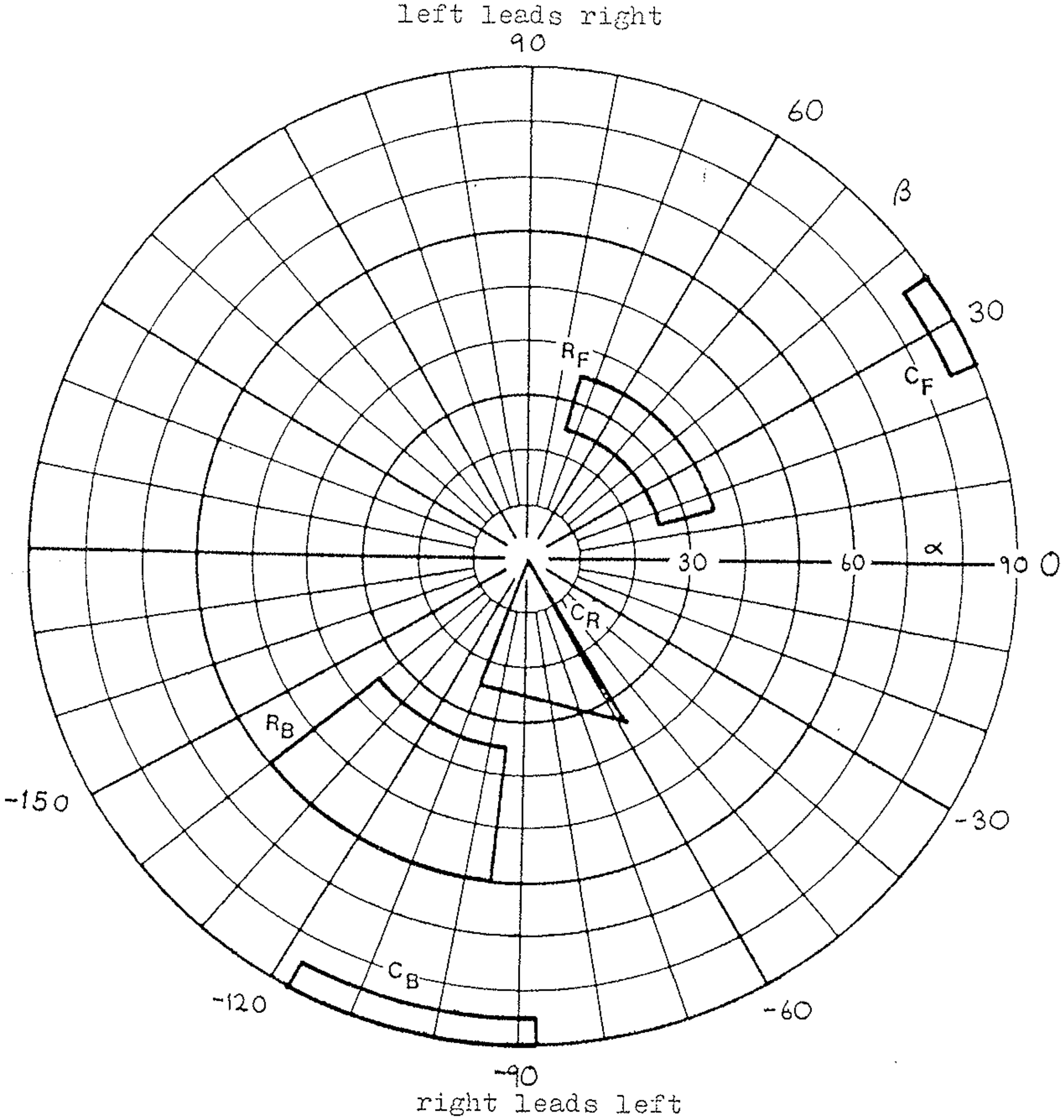


Figure 1. SYSTEM HJ TOLERANCE ZONES

Stage location	α	β
C_F	90 ±5	30 ±6
R_F	30 ±5	45 ±30
C_R	Bounded by points (α = 0), (α = 25, β = −110), (α = 35, β = −58)	
R_B	47.5 ±12.5	−117.5 ±22.5
C_B	90 ±5	−102.5 ±15
L_B	132.5 ±12.5	−117.5 ±22.5
C_L	Bounded by points (α = 180), (α = 155, β = −110), (α = 145, β = −58)	
L_F	150 ±5	45 ±30

III. System UHJ encoding standards

3.1 INTRODUCTION

System UHJ is a mutually compatible family of standards for encoding surround sound. The encodings exist in 2-, 3- and 4-channel versions, using 4 channel signals denoted L,R,T,Q. The channels L and R are the conventional stereo channels, and the encoding specification is such as to permit a horizontal surround sound reproduction to be derived from these two channels alone. In this two-channel form, we have "System UHJ , basic", denoted as "System BHJ ". The third channel (T), when available, is used to improve the accuracy of directional reproduction in the horizontal plane, and the system standards are so designed that good reproduction may still be obtained even when the T channel is bandlimited or reduced in gain. The fourth channel (Q), when available, may be used for either of two purposes: (i) in standard System UHJ it may be used to emphasise the loudspeaker positions in a square speaker layout. This "speaker emphasis" effect, often termed "separation" or "discreteness", degrades the localisation effect for non-speaker directions, but seems to be sought after by some users. (ii) In the System HHJ version of the system, the Q channel is used to convey full-sphere with-height directional effects, including elevation and depression.

These different uses of the Q channel are compatible, since the system standards are designed so that a "speaker emphasis" effect decoder does not reproduce an HHJ encoding with incorrect azimuthal directional effect, and a full-sphere with-height decoder does not reproduce a "speaker emphasis" recording with any marked unintended elevation effects. Both "speaker emphasis" and full-sphere decoders reproduce recordings with Q = 0 with correct directional effect, and conversely, all 3-channel decoders reproduce all types of 3- and 4-channel UHJ recordings correctly. Thus the consumer has no need to be informed which of the 3- and 4-channel options are in use. For the most accurate horizontal 4-speaker decoding of directional effect, a decoder using only the L,R,T channels, and not Q, is recommended, since it is now known that the fourth channel can only degrade the accuracy of reproduction of phantom inter-speaker images.

The System UHJ encoding specification is a kernel specification, not a matrix specification. This means that the specification assigns to each possible direction of sound arrival (in the horizontal plane for UHJ , and on the full sphere for HHJ) a set of complex gains, one for each channel, with which a sound from that direction must be encoded. Kernel specifications do not assume any prior 4-channel mastertape with an undefined method of directional encoding. This is realistic in terms of current recording practice, where recordings are stored in the form of multitrack (e.g. 16 track) mono, or else taken direct from elaborate microphone arrays.

3.2 Notation

All azimuths are measured anticlockwise from due front. The angle θ is used to indicate the intended azimuth of arrival of an encoded sound, and (where relevant) the angle η is used to indicate the intended elevation of a sound above the horizontal plane. Thus $\theta = 45^\circ$ indicates the direction commonly denoted LF, $\theta = -90^\circ$ indicates CR (due right), and $\theta = -45^\circ$, $\eta = -30^\circ$ indicates a direction 30° below RF.

The usual left and right stereo transmission channel signals are denoted respectively by L and R. The corresponding sum and difference channel signals are denoted respectively by $\Sigma = L+R$ and $\Delta = L-R$. It is convenient to give encoding and decoding equations in terms of Σ and Δ rather than L and R, since this simplifies the equations. To recover L and R from these equations, note that $L = \frac{1}{2}(\Sigma+\Delta)$ and $R = \frac{1}{2}(\Sigma-\Delta)$.

The third and fourth channels are denoted respectively by T and Q. We shall write Σ_{gain}, Δ_{gain}, T_{gain}, Q_{gain} for the complex gains of the respective channels to sounds arriving from a stated azimuth θ (and, where relevant, elevation η).

In the case of pairwise mixed 4-channel material, the usual designations LB, LF, RF, RB are used for the signals in the left back, left front, right front and right back channels respectively.

The symbol j is used here in its usual meaning, to indicate a relative 90° phase lead; for the purposes of computations, j may be treated as $\sqrt{-1}$.

3.3 Kernel Encoding Equations for System UHJ

System BHJ 2-channel kernel equations

$$\Sigma_{gain} = 0.9397 + 0.2624\cos\theta - 0.0241j\sin\theta$$

$$\Delta_{gain} = -0.3420j + 0.7211j\cos\theta + 0.9121\sin\theta$$

System THJ 3-channel kernel equations

$$\Sigma_{gain} = 0.9397 + 0.2624\cos\theta - 0.0241j\sin\theta$$

$$\Delta_{gain} = -0.3420j + 0.7211j\cos\theta + 0.9121\sin\theta$$

$$T_{gain} = -0.1365j + 0.9218j\cos\theta - \sin\theta$$

System QHJ 4-channel "speaker emphasis" kernel equations

$$\Sigma_{gain} = 0.9397 + 0.2624\cos\theta - 0.0241j\sin\theta$$

$$\Delta_{gain} = -0.3420j + 0.7211j\cos\theta + 0.9121\sin\theta$$

$$T_{gain} = -0.1365j + 0.9218j\cos\theta - \sin\theta$$

$$Q_{gain} = aj\sin 2\theta$$

where $0 \leqslant a \leqslant 1$. The attenuation a may be chosen by the producer to obtain any desired degree of speaker emphasis.

System HHJ 4-channel with-height kernel equations

$$\Sigma_{gain} = 0.9397 + 0.2624\cos\theta\cos\eta - 0.0241j\sin\theta\cos\eta$$

$$\Delta_{gain} = -0.3420j + 0.7211j\cos\theta\cos\eta + 0.9121\sin\theta\cos\eta$$

$$T_{gain} = -0.1365j + 0.9218j\cos\theta\cos\eta - \sin\theta\cos\eta$$

$$Q_{gain} = 1.3804\sin\eta$$

3.4 B-format encoding equations for UHJ

B-format refers to a signal format particularly convenient for studio signal handling purposes, and for reasons which will become apparent when discussing the handling of pairwise mixed and related material, it is convenient to give the encoding equations in terms of B format signals.

B-format consists of four signals W, X, Y, Z (for horizontal or with-height material) or W, X, Y, F (for speaker-emphasis horizontal material) having the directional gains given by the following equations:

With-height case

$$W_{gain} = 1 \ ,$$

$$X_{gain} = 2^{\frac{1}{2}}\cos\theta\cos\eta \ ,$$

$$Y_{gain} = 2^{\frac{1}{2}}\sin\theta\cos\eta \ ,$$

$$Z_{gain} = 2^{\frac{1}{2}}\sin\eta \ .$$

Horizontal-only case

$$W_{gain} = 1 \ ,$$

$$X_{gain} = 2^{\frac{1}{2}}\cos\theta \ ,$$

$$Y_{gain} = 2^{\frac{1}{2}}\sin\theta \ ,$$

$$F_{gain} = a\sin 2\theta \quad (0 \leq a \leq 1).$$

Note that all these signals have real gains for all directions (and so involve no phase shifters in their production). The W signal is omnidirectional, whereas the X, Y and Z signals have "figure of eight" directional characteristics pointing respectively forward, to the left and upwards with maximum gain $\sqrt{2}$. This gain is most convenient for signal handling purposes and ensures a roughly equal energy in all B-format signals. (A format of this type, but with slightly different channel gains was first proposed in M.A. Gerzon, "Periphony: With-Height Sound Reproduction", J. Audio Eng. Soc., Jan/Feb 1973 and Audio Engineering Society Convention preprint, Munich, March 1972.)

The B-format encoding equations for System UHJ are as follows:

$$\Sigma = 0.9397W + 0.1856X - 0.0171jY$$

$$\Delta = -0.3420jW + 0.5099jX + 0.6449Y$$

$$T = -0.1365jW + 0.6518jX - 0.7071Y$$

$$Q = 0.9761Z \text{ or } Q = jF .$$

3.5 Energy Sphere Locus for System BHJ

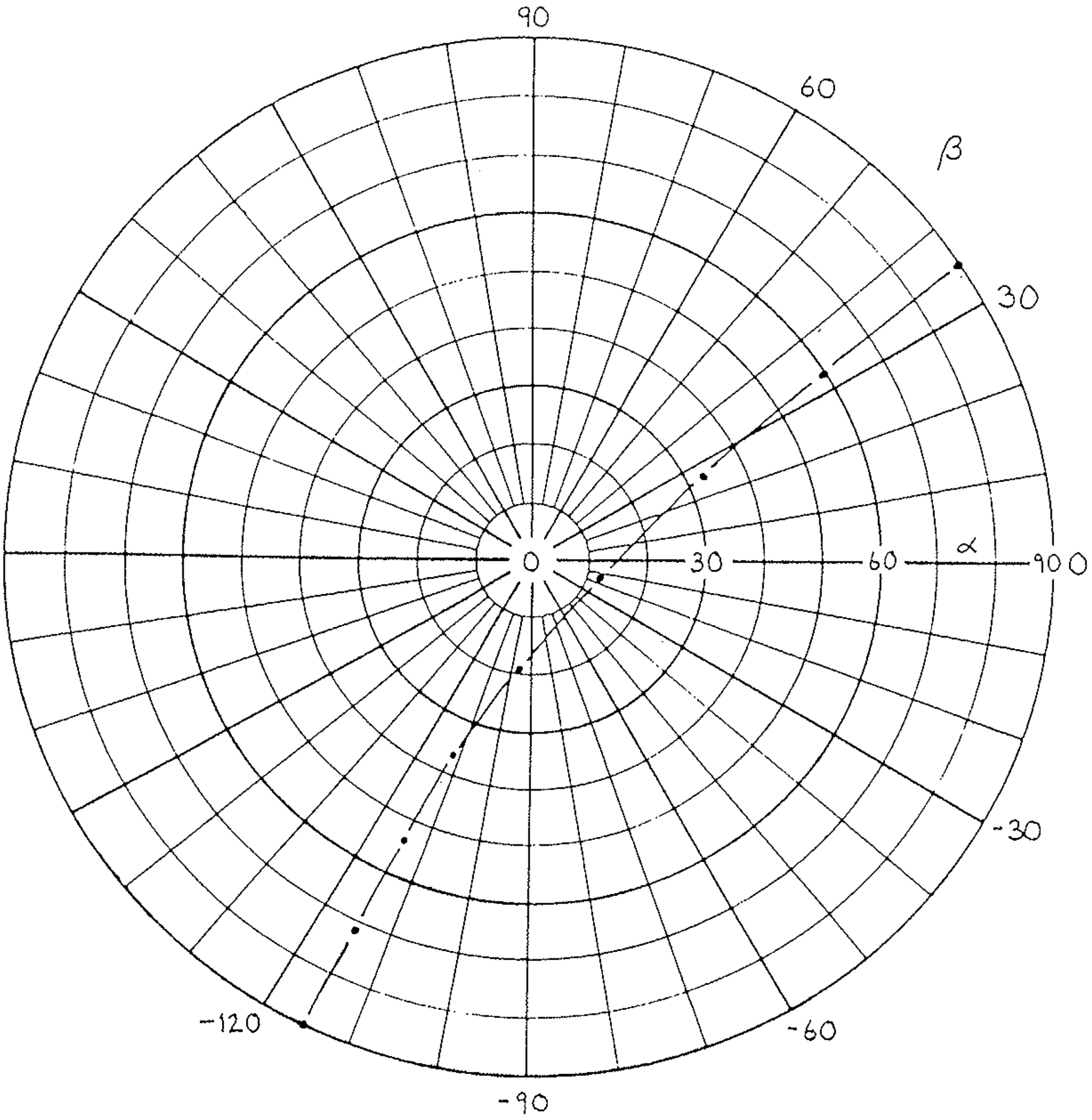


Figure 2. System BHJ locus plotted showing azimuths $\theta = 0^\circ, -22\frac{1}{2}^\circ, -45^\circ, -67\frac{1}{2}^\circ, -90^\circ, -112\frac{1}{2}^\circ, -135^\circ, -157\frac{1}{2}^\circ$ and 180° indicated by points in locus. The plot is made on Schieber coordinates α and β, defined by

$$\alpha = 2\tan^{-1}|R/L|, \quad \beta = \arg L - \arg R \quad ,$$

where L and R are the respective complex gains in the left and right transmission channels. The locus is left/right symmetric on the Energy Sphere, with azimuths between 0° and 180° on the left half of the sphere.

IV. Realisation and Approximation of UHJ Encoding Specification.

4.1 Introduction

It is possible to devise pan-potting devices that accurately meet B-format kernel specifications (e.g. see M. Gerzon, "Ambisonics. Part two: Studio Techniques", Studio Sound, Aug. 1975 pp. 24, 26, 28-30), and to design coincident directional microphone arrays whose output signals accurately meet B-format specifications. Such B-format signals can then easily be encoded using a phase-amplitude matrix into a System UHJ encoding as described in section 3.4. Many existing surround-sound recordings, and the outputs of many mixing desks in current use, do not conform to B-format and cannot be encoded to meet the UHJ kernel specification accurately. The aim of the present section is to describe various means of approximating the UHJ encoding specification for a variety of studio situations likely to be encountered. Clearly, such approximation should be avoided if the means are available to meet the kernel specification exactly, but practical guidance on acceptable methods of approximation is clearly necessary for the day-to-day operation of the system with non-ideal source material.

Rather than giving the System UHJ encoding equations directly for each type of imperfect source material, it is much more convenient to describe how approximate B-format signals may be realised, denoted by W_{app}, X_{app}, Y_{app}, F_{app} and Z_{app}. These may then be encoded to System UHJ by the B-format encoding equations of section 3.4. The reason for giving the approximate encoding in this two-stage form is not only that the equations become simpler, but that the apparatus generally used to derive approximate B-format is generally very simple, since it involves no phase shifters. For this reason, it is also recommended that the System UHJ encoder have a block diagram as illustrated in figure 3, with a converter (designed for the specific type of source material being handled) to approximate B-format followed by a B-format-in, System UHJ-out phase-amplitude matrix encoder. The converter to approximate

B-format may be included as a part of the System UHJ encoder or may be implemented at another part of the system (e.g. as an integral part of a particular surround-sound directional microphone array). For this reason, it is strongly recommended that the System UHJ encoder should provide inputs both for the imperfect signal format in use and for true B-format signals.

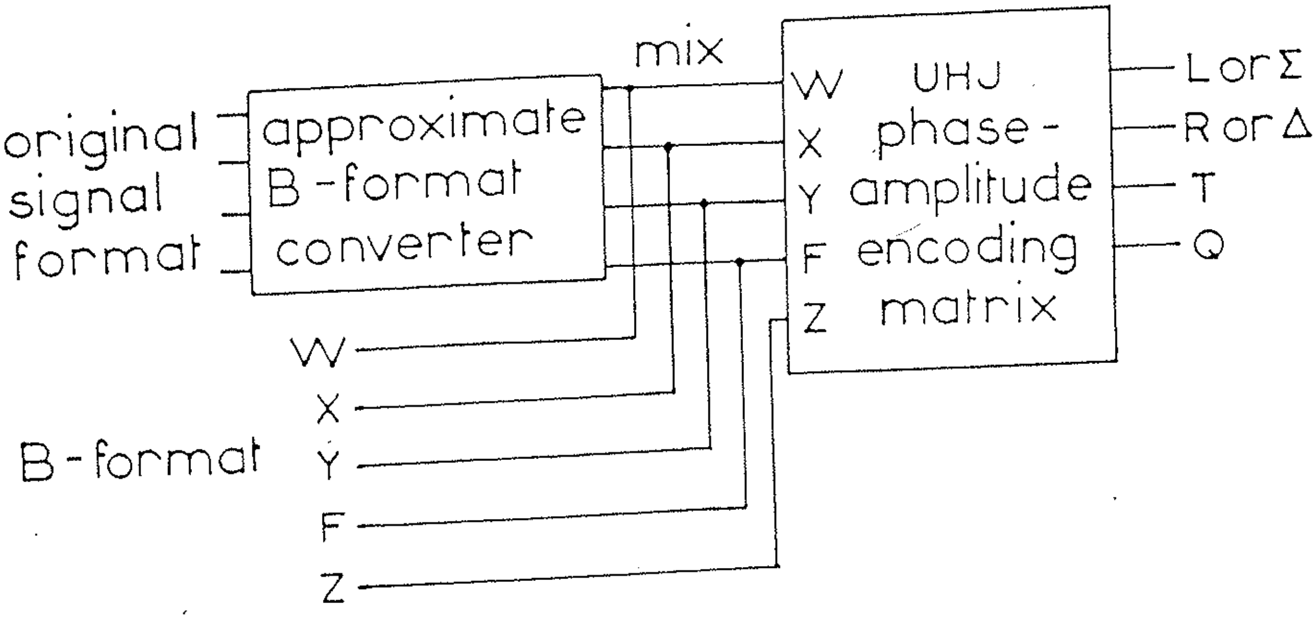


Figure 3. UHJ encoding using B-format approximate conversion.

We remark that B-format is a particularly useful method of handling signals within the studio, because while it involves no phase-amplitude matrixing, it turns out that nevertheless it is subjectively more "robust" under mild abuses such as interchannel gain or phase errors than are other proposed studio formats. Also, if a Q channel is not used, the UHJ encoding may be made just from the three signals W, X, Y. We recommend the following track assignments for recording B-format on a three or four tape tracks in order to minimise the effects of tape azimuth errors:

track 1	X
track 2	W
track 3	Y
track 4	F or Z

Figure 4. Tape track assignments for B-format signals.

4.2 Four-Channel Pairwise Mixed Material

The most common method of production for surround sound in current studio practice is the 4-channel "pairwise mixed" format, intended for reproduction via a conventional square speaker layout LB, LF, RF, RB wherein one of the channels is assigned to each of the loudspeaker positions, sounds assigned to a loudspeaker position are recorded only in the associated channel, and sounds assigned to azimuths between two of the loudspeaker azimuths are assigned just to the two channels corresponding to the adjacent loudspeakers,with identical phases and with an amplitude ratio indicative of the encoded azimuth. It can be shown that it is not possible to matrix such pairwise mixed material (hereafter denoted as PWM) accurately into either B-format or into any of the UHJ systems. There is a narrow range of approximate conversion matrices which accurately encode between four and eight azimuths accurately into B-format, while giving a degree of error for other azimuths. Such errors are, in part, responsible for the poor quality of side-images of the PWM format, and the possibilty of choosing the azimuths at which errors occur means that a careful choice of approximate B-format conversion will allow producers to improve the average image localisation quality of some program material. Alternatively, a particular choice of approximate B-format conversion can be used to obtain optimum quality of monophonic and/or stereophonic reproduction of the resultant UHJ signals. It is expected that many organisations will not wish to use the full range of approximate encoding options for PWM material, but will standardise on one particular choice suited to their typical requirements.

The approximate B-format conversion equations involve three parameters k_F, k_B and g_B , and take the form:

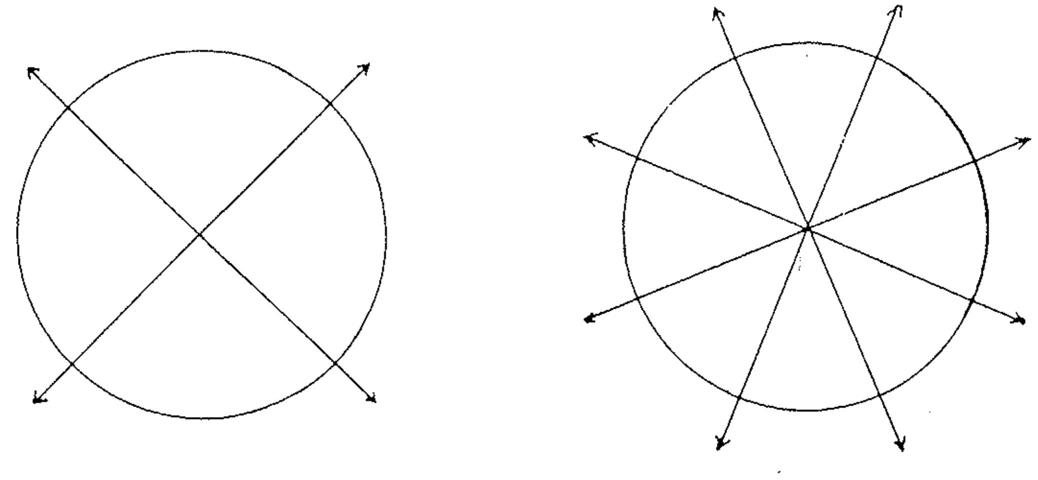
$$W_{app} = 2^{-\frac{1}{2}}k_F^{-1}(LF+RF) + 2^{-\frac{1}{2}}k_B^{-1}g_B(LB+RB)$$
$$X_{app} = (LF + RF) - g_B(LB + RB)$$
$$Y_{app} = (LF - RF) + g_B(LB - RB)$$
$$Z_{app} = 0$$
$$F_{app} = (LF - RF) - g_B(LB - RB)$$

where $0.7071 \leqslant k_F \leqslant 1$, $0.7071 \leqslant k_B \leqslant 1$, and $0.7071 \leqslant g_B \leqslant 1.4142$.

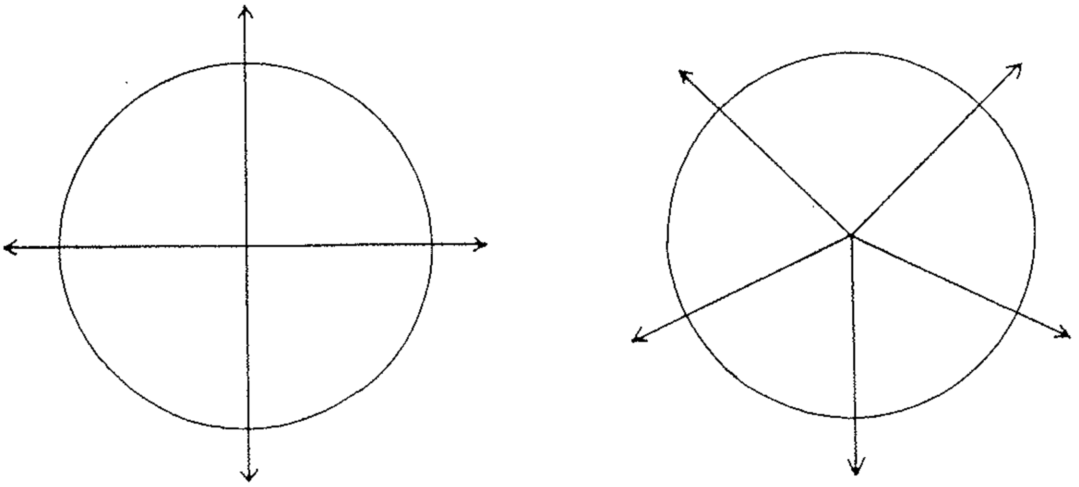
The quantity g_B determines the relative gain of reproduction of the back and the front channels, and should be chosen to ensure rough equality of gain in optimal BHJ and THJ reproduction. The quantities k_F and k_B determine which azimuths in, respectively, the front and back quadrants are accurately encoded to B-format. k_F or $k_B = 0.7071$ ensure that the corner azimuths (at respectively front and back) are correctly encoded, whereas k_F or $k_B = 1$ ensures that due front or due back are correctly encoded. If $k_F = k_B = k$, then the correctly-encoded azimuths are symmetrically disposed in the four quadrants. Figure 5 illustrates the correctly-encoded azimuths for various values of k_F and k_B.

The various B-format approximations give rise to varying degrees of corner-to-corner crosstalk when reproduced in "speaker emphasis" mode, as described in section 5.3, and values of this crosstalk in dB for various options are tabulated in table I. In no case is the corner-to-corner crosstalk worse than -22dB; the side-to-side crosstalk can reach -15.3 dB, but is in antiphase which actually helps improve phantom-image localisation especially for side positions.

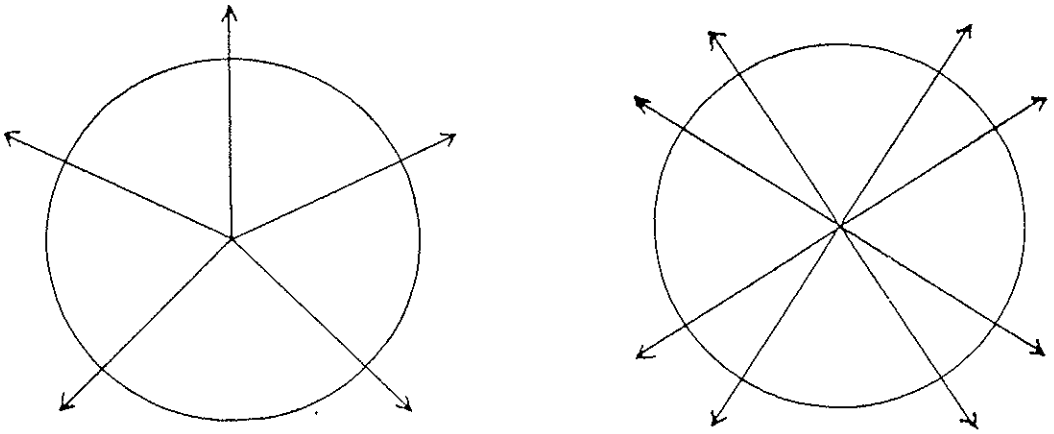
Table 2 lists the approximate optimal value for g_B to ensure maximum gain uniformity in optimal 2- and 3-channel reproduction for various values of k_F and k_B. For general-purpose use, the best overall mono and stereo compatibility may be obtained using the values $k_F = 1$, $k_B = 0.7071$ or $k_F = 0.8409$, $k_B = 0.7071$ or $k_F = k_B = 0.8409$. As seen from figure 6, which shows the



$k_F = k_B = k = 0.7071$ $k_F = k_B = k = 0.9239$



$k_F = k_B = k = 1.0000$ $k_F = 0.7071$, $k_B = 1.0000$



$k_F = 1.0000$, $k_B = 0.7071$ $k_F = k_B = k = 0.8409$

Fig. 5 Showing those azimuths of pairwise mixed material that are encoded correctly according to B-format kernel encoding standards for various values of the coefficients k_F, k_B, k in the pairwise mix encoding equations.

Encoded azimuth	$k=1.0000$	$k=0.9239$	$k=0.7071$	$k_F=1.0000$ $k_B=0.7071$	$k_F=0.7071$ $k_B=1.0000$	$k=0.8409$
CF 0°	-15.31	-17.53	$-\infty$	-15.31	$-\infty$	-21.27
LF 45°	-22.05	-24.11	$-\infty$	-22.05	$-\infty$	-27.65
CL 90°	-15.31	-17.53	$-\infty$	-22.05	-22.05	-21.27
LB 135°	-22.05	-24.11	$-\infty$	$-\infty$	-22.05	-27.65
CB 180°	-15.31	-17.53	$-\infty$	$-\infty$	-15.31	-21.27

Table I. Cross-talk (in dB) for "speaker emphasis" decoding of pairwise mixed material for various values of the pairwise encoding parameters k, k_F, k_B. Level of cross-talk is identical in all nominally unactivated speaker channels. Cross-talk is anti-phase with respect to activated speaker channels. In all cases, corner-to-corner cross-talk ("channel separation") is better than -22dB.

k_F	k_B	g_B
1.0000	1.0000	1.0000
0.9239	0.9239	1.0000
0.8409	0.8409	1.0000
0.7071	0.7071	1.0000
1.0000	0.7071	0.8660
0.7071	1.0000	1.1547
0.8409	0.7071	0.9239

Table II. Value of g_B ensuring roughly uniform front and back gain via psychoacoustically optimised 2-channel and 3-channel decoders for various values of k_F and k_B. In general, g_B is given approximately by the formula

$$g_B = \left(\frac{\frac{1}{2}k_F^{-2}+1}{\frac{1}{2}k_B^{-2}+1} \right)^{\frac{1}{2}} .$$

energy sphere loci of front and back quadrant PWM encoded material for various values of k_F and k_B for System BHJ, the encoded PWM material falls within the System HJ encoding tolerances of chapter II provided that k_F falls between 0.83 and 1.0. Figure 7 shows the BHJ approximation obtained for two choices of k_F and k_B.

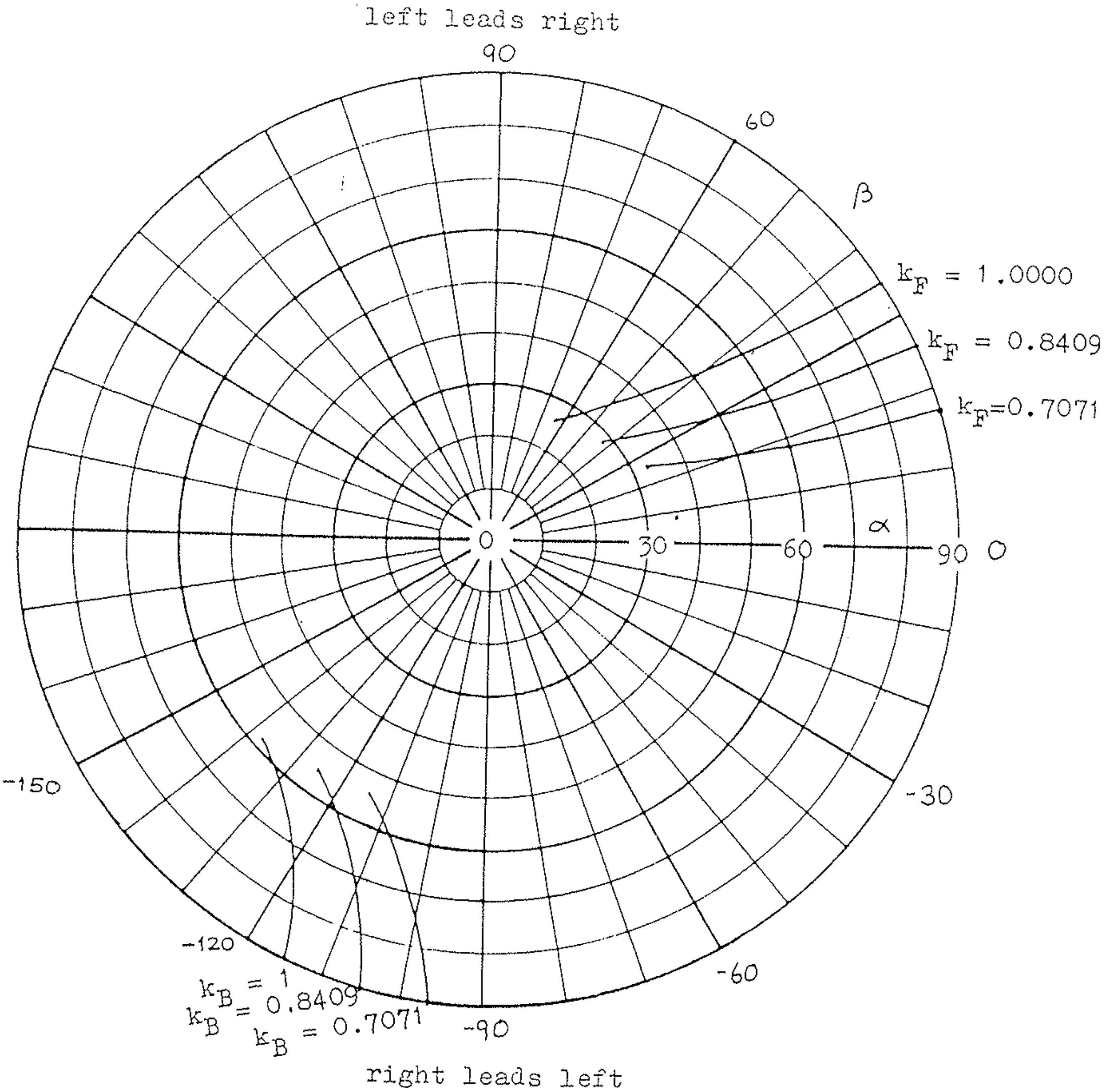


Figure 6. Energy Sphere plot, using Schieber coordinates α and β, showing the encoding of front-quadrant and back-quadrant pairwise mixed 4-channel material into System BHJ using various values of the PWM encoding constants k_F and k_B.

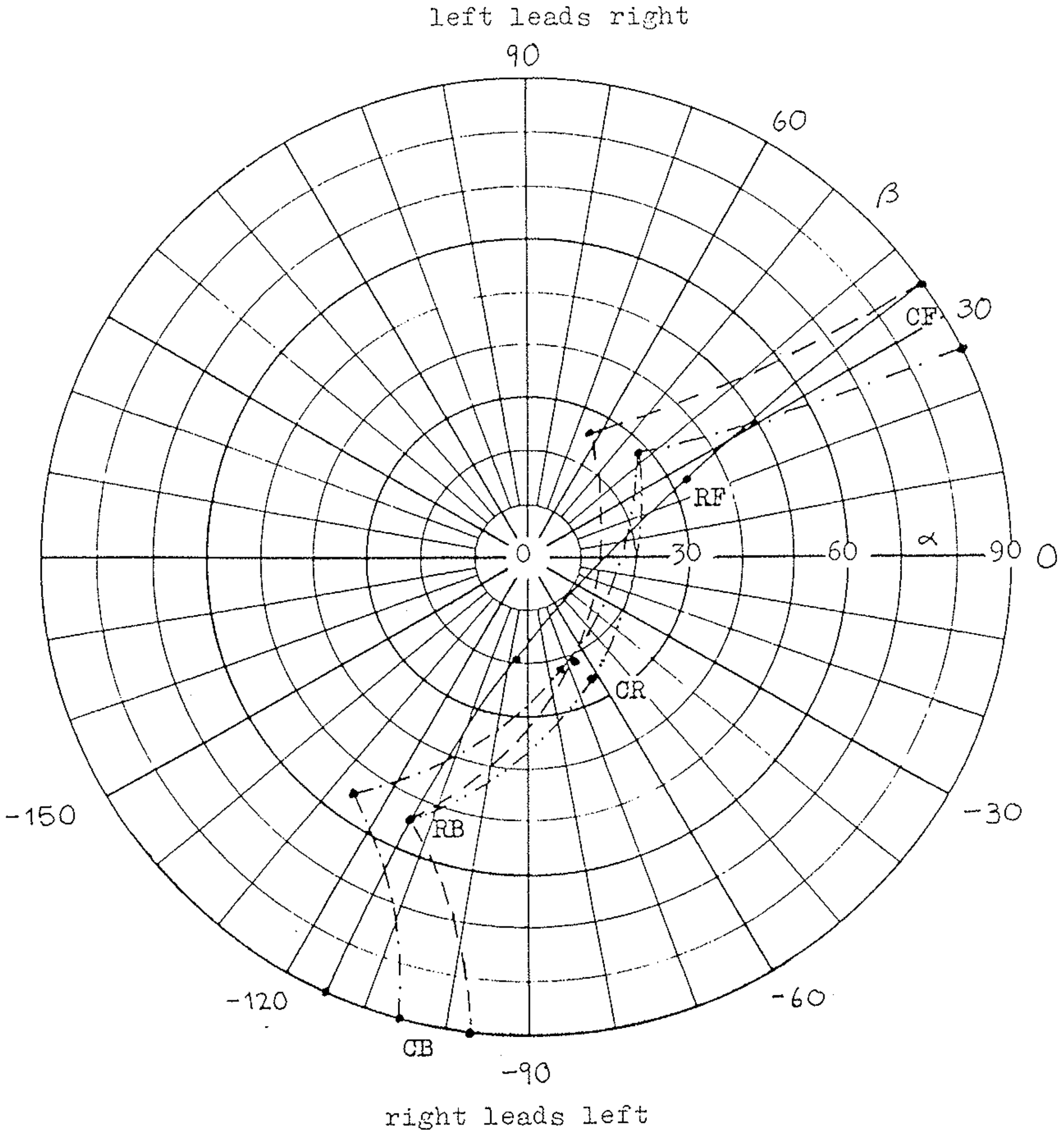


Figure 7. Energy Sphere loci of various pairwise mix encoding options for System BHJ, plotted on Schieber coordinates α & β.

———— Kernel encoding locus.

— — — — $k_F = 1$, $k_B = 0.7071$, $g_B = 0.8660$

—·—·—· $k_F = k_B = 0.8409$, $g_B = 1$

—··—··— $k_F = 0.8409$, $k_B = 0.7071$, $g_B = 0.9239$

4.3 Encoding using a mixer with 8 groups

If pan-potted material is prepared using pairwise techniques on a mixer with four output groups, the inherent errors of pressure: velocity gain cannot be reduced below 1.5 dB, and may equal up to 3 dB. However, many studios are equipped with mixers having eight or more output groups, with each fader assignable either to any number of groups or stereo-pannable between even and odd-numbered groups. This section describes a method of approximating B-format, and hence UHJ, using such mixers, with improved accuracy. Numbering the groups 1 to 8, each group is assigned to a different direction as shown in figure 8, with group 1 in front, group 2 at LF, group 3 at CL, and so on round the circle. The eight group signals G_1 to G_8 are then encoded into B-format for their respective directions, via the formulas

$$\begin{aligned}
W_{app} &= G_1+G_2+G_3+G_4+G_5+G_6+G_7+G_8 \\
X_{app} &= 2^{\frac{1}{2}}G_1+G_2-G_4-2^{\frac{1}{2}}G_5-G_6+G_8 \\
Y_{app} &= G_2+2^{\frac{1}{2}}G_3+G_4-G_6-2^{\frac{1}{2}}G_7-G_8 \\
F_{app} &= G_2 - G_4 + G_6 - G_8 .
\end{aligned}$$

Sounds may be assigned to intermediate azimuths by pairwise panning between adjacent groups, e.g. between 8 and 1 or between 3 and 4. The maximum error in the ratio of pressure to velocity is 0.69 dB using this procedure. Note that the eight phantom positions $22\frac{1}{2}^{\circ}$ away from the corners may be obtained without using any pan-pots simply by assigning a signal to the two adjacent groups. Additionally, if the mixer is equipped with 4-channel paiwise pan-pots, sounds can be circled round the stage using these by assigning the 4 outputs of the device to groups G_2, G_4, G_6, and G_8 , with some loss of accuracy in the encoding.

If an organisation has to handle both 4-channel pairwise mixed and 8-group mixed material, it is desirable to use an encoder or approximate B-format converter that will handle both requirements. This may be ensured by using the G_2, G_4, G_6, and

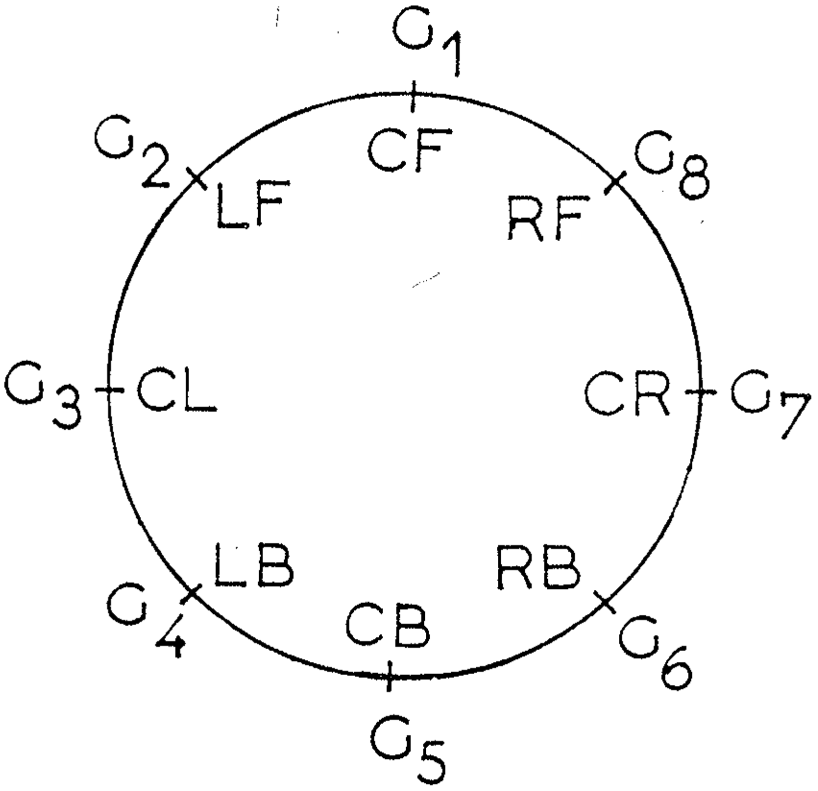


Figure 8. Typical group assignment for encoding from a mixer with eight output groups.

G_8 inputs of the converter for LF, LB, RB and RF respectively. of the 4-channel PWM material, and providing a choice of matrices matrices for the approximate PWM conversion. This may be achieved via the approximate B-format encoding equations:

$$
\begin{aligned}
W_{app} &= CF+CL+CB+CR + a_F(LF+RF) + a_B(LB+RB) \\
X_{app} &= 2^{\frac{1}{2}}(CF-CB) + b_F(LF+RF) - b_B(LB+RB) \\
Y_{app} &= 2^{\frac{1}{2}}(CL-CR) + b_F(LF-RF) + b_B(LB-RB) \\
F_{app} &= b_F(LF-RF) - b_B(LB-RB)
\end{aligned}
$$

where the positive real numbers a_F, a_B, b_F, b_B are chosen so as to ensure uniform gain via psychoacoustically optimised 2- and 3-channel decoders (see section 5.4). Table III lists suitable values of these coefficients corresponding to various values of k_F and k_B in the 4-channel encoding.

k_F	k_B	a_F	a_B	b_F	b_B
1.0000	1.0000	0.8165	0.8165	1.1547	1.1547
0.8409	0.8409	0.9102	0.9102	1.0824	1.0824
0.7071	0.7071	1.0000	1.0000	1.0000	1.0000
1.0000	0.7071	0.8165	1.0000	1.1547	1.0000
0.7071	1.0000	1.0000	0.8165	1.0000	1.1547
0.8409	0.7071	0.9102	1.0000	1.0824	1.0000

Table III. Values of coefficients a_F, a_B, b_F, b_B for various values of k_F and k_B.

V. System UHJ decoders

5.1 INTRODUCTION

No decoding equations can be regarded as part of an encoding standard, since the job of the decoder is to provide patterns of sound in the listening room capable of producing an illusion of the intended encoded directional effect at the ears of a listener. Since theoretical and empirical knowledge of psychoacoustics is constantly improving, it follows that the best available designs of decoders giving an optimal illusion will also improve, and no standard should restrict such legitimate development of the art. However, any encoding standard should be formulated to make such development possible, insofar as it is possible to forsee future developments. System UHJ has been formulated specifically to optimise a number of localisation criteria (see: M.A. Gerzon, "Surround Sound Psychoacoustics", Wireless World, 80, pp.483-6 (Dec. 1974)), and to permit others to be designed objectively. All decoders described satisfy Makita's localisation theory, and all save the "speaker emphasis" designs also satisfy other relevant criteria.
All decoders are specified in kernel form, i.e. by an equation relating the speaker feed signal to the speaker azimuth (and, where relevant, elevation).

All designs given here are given purely for reference purposes, and as an aid to understanding the encoding specification. Future research may well yield improved designs, especially since many of these designs may have the various design parameters varying with frequency.

5.2 NOTATION

The azimuth (measured anticlockwise from due front) of loudspeakers is denoted by ϕ. In all decoders, it is assumed that all loudspeakers are equidistant from a central reference position, with respect to which all loudspeaker directions are measured. The signal fed to the loudspeaker at azimuth ϕ is denoted P_ϕ for regular loudspeaker layouts. The elevation, where relevant, of loudspeakers above the horizontal plane is denoted by ξ, and for regular with-height loudspeaker layouts, the signal fed to the speaker at azimuth ϕ and elevation ξ is denoted by $P_{\phi,\xi}$. In the following decoder descriptions, no allowance is made for the finite distance of loudspeakers from the central point; optimal practical designs will include modifications to allow for finite loudspeaker distances.

Other notations are as given in the encoding specification.

5.3 SQUARE LAYOUT "SPEAKER EMPHASIS" DECODER

This decoder, which is not psychoacoustically optimal, is a reference decoder specifically intended for use with a square layout of loudspeakers, and gives the "speaker emphasis" effect (often termed "separation" or "discreteness") whereby sounds are drawn towards the four loudspeaker positions. For the loudspeaker placed at one of the azimuths $\phi = \pm 45^\circ$ and $\pm 135^\circ$, the signal feed is

$$\begin{aligned} P_\phi = {} & (0.9790\,\Sigma + 0.1785\,j\Delta + 0.1391\,jT) \\ & + (0.5849\,\Sigma - 1.1740\,j\Delta - 1.0849\,jT)\cos\phi \\ & + (0.2718\,j\Sigma + 1.1308\,\Delta - 0.9620T)\sin\phi \\ & - 1.0000jQ\sin 2\phi \ . \end{aligned}$$

OTHER "SPEAKER EMPHASIS" DECODERS

If it is desired to obtain an emphasis of the four "corner" azimuths $\phi = \pm 45^\circ$ and $\pm 135^\circ$ via a regular polygon loudspeaker layout using more than 4 loudspeakers, then the speaker at azimuth ϕ should be fed with P_ϕ as given immediately above. For a rectangular loudspeaker layout with speaker azimuths ϕ, $180^\circ - \phi$, $-180^\circ + \phi$, $-\phi$, the respective loudspeaker feed signals should be $P_{90^\circ - \phi}$, $P_{90^\circ + \phi}$, $P_{-90^\circ - \phi}$, $P_{-90^\circ + \phi}$.

5.4 3-CHANNEL HORIZONTAL DECODERS

The basic kernel decoding equation feeds a loudspeaker at azimuth ϕ within a regular polygon loudspeaker layout (using not less than 4 speakers) with the signal

$$\begin{aligned} P_\phi = {} & (0.9790\Sigma + 0.1785j\Delta + 0.1391jT) \\ & + (0.5849\Sigma - 1.1740j\Delta - 1.0849jT)\cos\phi \\ & + (0.2718j\Sigma + 1.1308\Delta - 0.9620T)\sin\phi \ . \end{aligned}$$

This equation causes a sound encoded at azimuth θ to be decoded with gain $1 + 2\cos(\theta - \phi)$ through the loudspeaker at azimuth ϕ.

The general kernel decoding equation is of the form

$$\begin{aligned} P_\phi = {} & k_1(0.9790\Sigma + 0.1785j\Delta + 0.1391jtT) \\ & + k_2(0.5849\Sigma - 1.1740j\Delta - 1.0849jtT)\cos\phi \\ & + k_2(0.2718j\Sigma + 1.1308\Delta - 0.9620tT)\sin\phi \\ & + k_3(-0.9790j\Sigma + 0.1785\Delta + 0.1391tT)\sin\phi \ , \end{aligned}$$

where $0 \leqslant t \leqslant 1$ is the attenuation of the third channel T, and where k_1, k_2 are positive numbers and k_3 a real (and usually positive) number chosen to optimise the psychoacoustics of the reproduction. The apparent sound azimuth reproduced by such a regular polygon decoder according to Makita's localisation theory agrees with the encoded azimuth exactly whenever $t = 0$ or 1, and within about 2° for intermediate values of t. The coefficients k_1, k_2, k_3, t may vary with frequency if desired.

For rectangular loudspeaker layouts with speaker azimuths ϕ, $180^\circ-\phi$, $-180^\circ+\phi$, $-\phi$, the respective loudspeaker feed signals should be $P_{90^\circ-\phi}$, $P_{90^\circ+\phi}$, $P_{-90^\circ-\phi}$, $P_{-90^\circ+\phi}$.

The parameters k_1, k_2, k_3, t in the horizontal kernel decoding equations for System UHJ 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
$k_1 = 1$, $k_2 = 1.0$, $k_3 = 0.3502$, $t = 0$.

* Psychoacoustically compensated 2-channel decoder $t = 0$ and:
$k_1 = 0.6577$, $k_2 = 1.2903$, $k_3 = 0.1493$ for $F \ll 400$ Hz,
$k_1 = k_2 = 1$, $k_3 = 0.3502$ for $F \gg 400$ Hz.

Basic "2½-channel" decoder
$k_1 = k_2 = t = 1$, $k_3 = 0$ at frequencies for which 3 channels are available,
$k_1 = k_2 = 1.16$, $k_3 = 0$, $t = 0$ when only 2 channels are available.

* Intermediate "2½-channel" decoder with almost uniform directional gain
$k_1 = k_2 = t = 1$, $k_3 = 0$ when 3 channels are available,
$k_1 = k_2 = 1.2233$, $k_3 = 0.4283$, $t = 0$ when only 2 channels are available.

* Psychoacoustically compensated "2½-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,
$k_1 = k_2 = 1.2233$, $k_3 = 0.4283$, $t = 0$ at high frequencies when only 2 channels are available.

Notes: (i) The term "2½-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 0.67 dB variation.

5.5 SIGNAL-ACTUATED 2-CHANNEL DECODERS

If it is desired to construct signal-actuated variable decoders for 2-channel BHJ encoded signals, the decoding equations for the speaker-feed signals P_ϕ for regular polygonal or rectangular speaker layouts (used to feed speakers as described in section 5.4) may be derived as follows:

$$\begin{aligned} P_\phi = {} & k_1(0.9790\Sigma + 0.1785j\Delta) \\ & + k_2(1.1476\Sigma - 2.3034j\Delta)\cos\phi \\ & + k_2(1.0578j\Sigma + 2.1231\Delta)\sin\phi \\ & + k_3(-0.9790j\Sigma + 0.1785\Delta)\sin\phi \\ & + k_4(-0.9790j\Sigma + 0.1785\Delta)\cos\phi \ , \end{aligned}$$

where at low frequencies substantially less than 400 Hz, $k_1 = k_2 = 1$, $k_3 = 1.8435\cos\theta'$ and $k_4 = -2.1698\sin\theta'$, where θ' is the deduced localisation of the dominant sound in that frequency range, and at frequencies substantially greater than 400 Hz, $k_1 = 1.2247$, $k_2 = 0.8660$, $k_3 = 1.5965\cos\theta''$ and $k_4 = -1.8791\sin\theta''$ where θ'' is the deduced localisation of the dominant sound in the higher frequency range.

The decoder just described gives correct Makita-theory reproduced azimuth for all sounds (even non-dominant ones) at all times, whatever the values of k_3 and k_4 provided that these are real. The gains k_3 and k_4 may be varied in several frequency bands according to the dominant localisation θ in each band, and a simplified variable decoder may be obtained by putting either or both of k_3 and k_4 equal to zero in one or more of the frequency bands. Other more complex decoders in which the overall gain also varies to compensate for the slight directional dependence of the gain of the above decoder can be devised.

The dominant azimuth θ can be deduced using a variety of known detection means, and the resultant decoder simulates, for the dominant direction, the "psychoacoustic" 3-channel decoder of section 5.4.

5.6 WITH-HEIGHT DECODERS

For a non-tetrahedral regular polyhedron loudspeaker layout, the loudspeaker placed at azimuth ϕ and elevation ξ has the loudspeaker feed signal:

$$\begin{aligned} P_{\phi,\xi} = {} & k_1(0.9790\Sigma + 0.1785j\Delta + 0.1391jT) \\ & + k_2(0.5849\Sigma - 1.1740j\Delta - 1.0849jT)\cos\phi\cos\xi \\ & + k_2(0.2718j\Sigma + 1.1308\Delta - 0.9620T)\sin\phi\cos\xi \\ & + k_2(1.4488Q)\sin\xi \quad , \end{aligned}$$

where k_1 and k_2 are positive numbers, with preferably $k_1 = 1.0000$ and $k_2 = 1.5000$ at frequencies substantially less than 700 Hz, and $k_1 = 1.4142$ and $k_2 = 1.2247$ at higher frequencies.

For a cuboid loudspeaker layout with eight loudspeakers placed at the azimuth/elevations $(\phi,\pm\xi)$, $(180^\circ-\phi,\pm\xi)$, $(-180^\circ+\phi,\pm\xi)$, $(-\phi,\pm\xi)$, the respective signal feeds are:

$$P_{\psi,\pm\varsigma} \; , \; P_{180^\circ-\psi,\pm\varsigma} \; , \; P_{-180^\circ+\psi,\pm\varsigma} \; , \; P_{-\psi,\pm\varsigma} \; ,$$

where $\psi = 90^\circ-\phi$ and $\varsigma = \tan^{-1}\left(\frac{\sin 2\phi}{2\tan\xi}\right)$.

Horizontal-only encoded signals may also be decoded via a regular or cuboid with-height speaker layout by using the above equations, with either $Q = 0$ or Q a speaker-emphasis signal, which has little effect on reproduced apparent sound direction.

VI. System UHJ Properties

6.1 Compatibility of UHJ kernel encoding in mono and stereo

Azimuth angle degrees	Mono Gain dB	Stereo Gain dB	Position P	Phasiness Q	Scheiber coordinates $180^\circ-\alpha$	β
C_F 0	0.00	0.00	0.000	0.315	90.0	35.0
$22\frac{1}{2}$	-0.15	0.10	0.293	0.276	59.7	33.4
L_F 45	-0.57	0.32	0.571	0.158	32.3	25.9
$67\frac{1}{2}$	-1.26	0.53	0.811	-0.046	12.2	-15.2
C_L 90	-2.14	0.62	0.979	-0.339	19.2	-96.2
$112\frac{1}{2}$	-3.12	0.53	1.023	-0.709	36.6	-111.2
L_B 135	-4.05	0.32	0.880	-1.110	54.2	-114.4
$157\frac{1}{2}$	-4.73	0.10	0.520	-1.439	71.9	-115.0
C_B 180	-4.98	0.00	0.000	-1.570	90.0	-115.0
Gain variation in dB.	4.98	0.62				

most extreme value of P : 1.027

Table IV. UHJ kernel coding compatibility data.

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 M.A. Gerzon, "A Geometric Model for 2-Channel 4-Speaker Matrix Stereo Systems", J. Audio Eng. Soc., vol.23, pp.98-106 (March 1975), appendix II.

6.2 UHJ Decoder Properties

Encoded Azimuth θ	t=0.00	t=0.25	t=0.50	t=0.75	t=1.00
0.0	0.0	0.0	0.0	0.0	0.0
22.5	22.3	22.1	22.1	22.2	22.5
45.0	44.8	44.3	44.2	44.5	45.0
60.0	60.0	59.1	59.0	59.3	60.0
67.5	67.6	66.6	66.4	66.8	67.5
90.0	90.3	88.9	88.6	89.1	90.0
112.5	112.6	111.2	111.0	111.6	112.5
120.0	120.0	118.6	118.5	119.1	120.0
135.0	134.7	133.7	133.7	134.2	135.0
157.5	157.1	156.6	156.7	157.0	157.5
180.0	180.0	180.0	180.0	180.0	180.0

Table V : Makita azimuths of decoded sounds for decoders of section 5.4 for various values of the attenuation t of the third channel T, for various encoded azimuths θ.

Encoded Azimuth θ	Makita Azimuth $\theta_V = \theta_E$	Velocity Vector Magnitude r_V	Phasiness q	Energy Vector Magnitude r_E	Energy Gain (in dB)
0.0	0.0	0.518	0.165	0.651	0.00
22.5	22.3	0.516	0.128	0.623	0.11
45.0	44.8	0.512	0.023	0.559	0.35
67.5	67.6	0.509	-0.133	0.491	0.58
90.0	90.3	0.509	-0.314	0.438	0.67
112.5	112.6	0.510	-0.493	0.404	0.58
135.0	134.7	0.508	-0.641	0.384	0.35
157.5	157.1	0.502	-0.738	0.374	0.11
180.0	180.0	0.500	-0.772	0.371	0.00

Table VI. Parameters describing subjective performance of 2-channel BHJ decoder with $k_1=k_2=1$, $k_3 = 0.3502$, t=0 of section 5.4, using square or regular polygon speaker layout shape. A description of the meaning of these parameters is to be found in : M.A. Gerzon, "Design of Ambisonic Decoders for Multispeaker Surround Sound", presented at the 58th Audio Engineering Society Convention, 4th Nov. 1977.

UHJ Encoder.

The basic UHJ encoder produces a stereo compatible UHJ encoded signal according to the BBC/NRDC jointly agreed specification from Ambisonic B-format, such as the signals from the Soundfield Microphone. The UHJ encoding is such that, via a suitable decoder, a horizontal soundfield illusion may be produced in the listeners home.

The directionality of sounds in the original soundfield is encoded by a unique relationship between the relative amplitude and relative phase of the left and right stereo compatible encoder output signals, each direction having only one possible output amplitude/phase combination. The relationship has been chosen carefully so that the resulting signals are fully stereo compatible.

To ensure the preservation of the high quality of the original B-format signals, the encoder was carefully designed for maximum audio transparency and low noise. Precision 10 pole phase-shifters are used to give an accuracy of 0.6° from 20Hz to 20kHz and the unit is constructed with high quality close tolerance components which have been further selected and matched.

Various options are available as extras to the basic encoder, including the addition of the third and fourth UHJ channels.

A new option now available is stereo transcoding. As it is expected that B-format will become the standard master tape format of the future, the basic encoder was designed to operate from such material. However, much programme material already exists in the form of either multitrack or pairwise-mixed masters. As very few mixing desks are available which can produce B-format at the moment, a transcoding input was developed for the UHJ encoder. This enables UHJ mixes to be produced from existing material via normal stereo or 'quad' output mixers. Normally if a stereo pair is first converted into B-format (eg. by panning the left channel signal to a left location and the right channel signal to a right location) and thence encoded via a UHJ encoder, a pressure-velocity ratio of up to 3dB can occur. This affects the quality and precision of the images in the mix and can an unpleasant 'in-the-head' effect for some source ections. By bypassing the B-format stage and encoding directly nto 2-channel UHJ this error can be avoided and this is the function of the transcoding input.

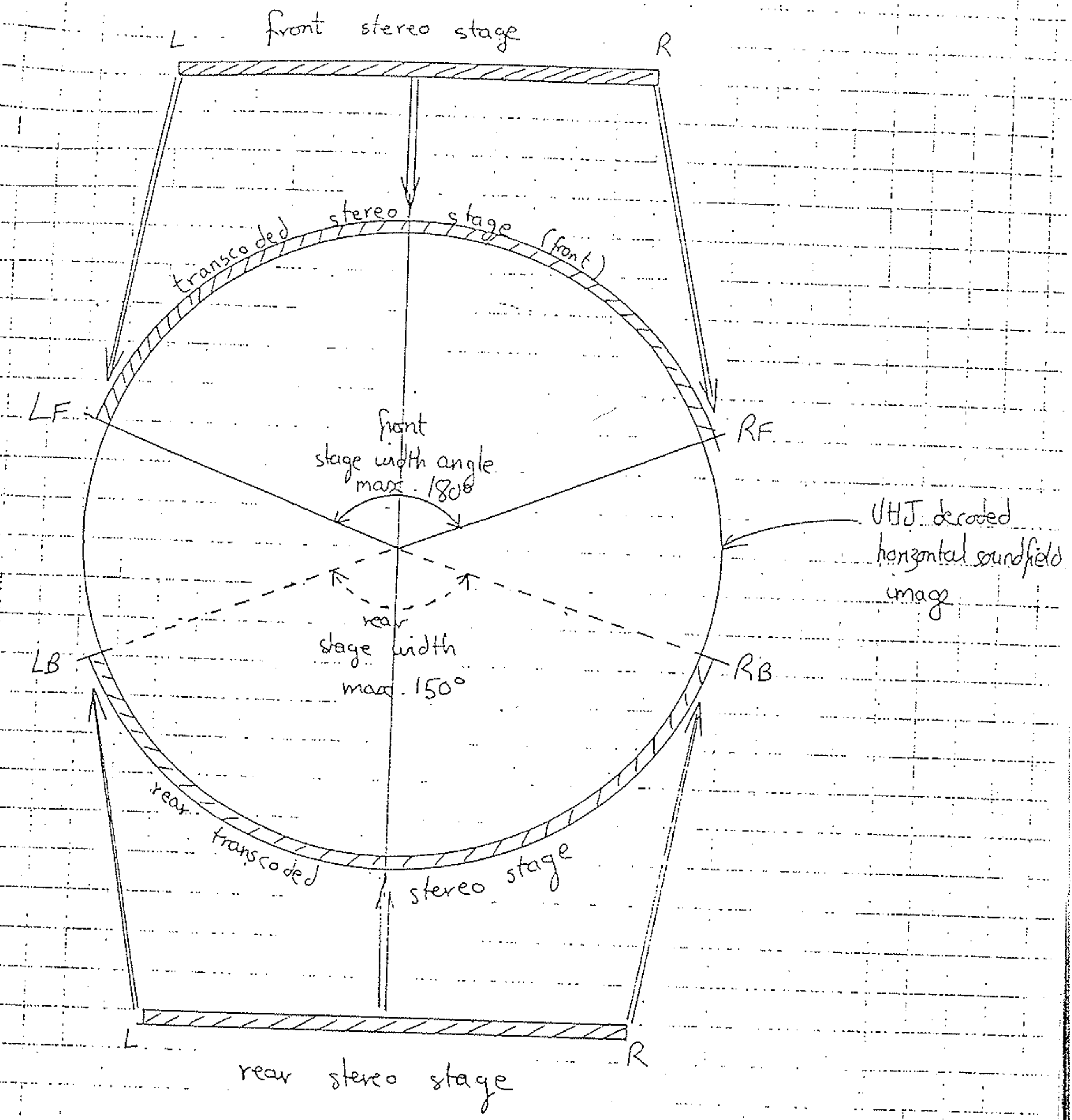
The input takes the form of two stereo pairs. One pair (LF,RF) is assigned to the front stage and is stretched through an angle , in the decoded sound stage, of up to 180°. (see figure 1). The width of the stage is continuously variable between 0° and 180° by the front stage width control on the front panel of the encoder. For example, if a stereo mix consists of three sources, full left, centre, and full right (ie. left channel only, equally left and right, and right channel only respectively) then for a width setting of, say, 150° the centre sound will appear in the centre and left and right will appear from 75° either side of centre (ie. 150° apart). If the width setting is then increased to 180° left and right will move out to 90° from centre. A 0° setting will produce a mono front centre image only. Of course signals panned between left and right will appear proportionately displaced between LF and RF on the decoded sound stage.

The second stereo pair (LB,RB) is assigned to the rear stage (ie. behind the listener). A second width control allows image widths of up to 150° at the rear.

The total transcoded sound stage coverage is therefore limited to a maximum of 180° front and 150° rear. This leaves two 15° gaps between the two stages. The maximum stage-widths were selected to keep the transcoding into the UHJ locus as accurate as possible; wider stages would have exceeded tolerable limits. Sounds panned between LF-LB or RF-RB are reproduced with substantially correct directionality but with pressure-velocity ratio errors, the error being smallest for the larger stage width settings.

The 150° front stage width setting facilitates a useful comparison. A stereo pair transcoded across the front with this setting has identical width and sound balance when compared with the original pair on two loudspeakers (ie. not decoded).

mod. for AE 0880



<u>UHJ Transcoder</u> stereo image transcoding diagram

<u>nb.</u> LF, RF, LB, RB represent the extremities of images and do not relate to loudspeaker positions. A mono source connected to the LF input of the transcoder will be reproduced at the point marked LF on the circle diagram above etc.

27 June 81

ADDITIONAL INFORMATION

After compiling the contents of the folder, some additional information arrived. It consists of the BBC Engineering Press Statement given below and the attached diagram "SYSTEM HJ TOLERANCE ZONES".

Engineering Press Statement 1530

BBC/NRDC SYSTEM HJ

Earlier this year, the BBC and NRDC announced their intention to exchange information and experience with a view to moving towards a joint specification for surround-sound (or quadraphonic) encoding in a two-channel system.

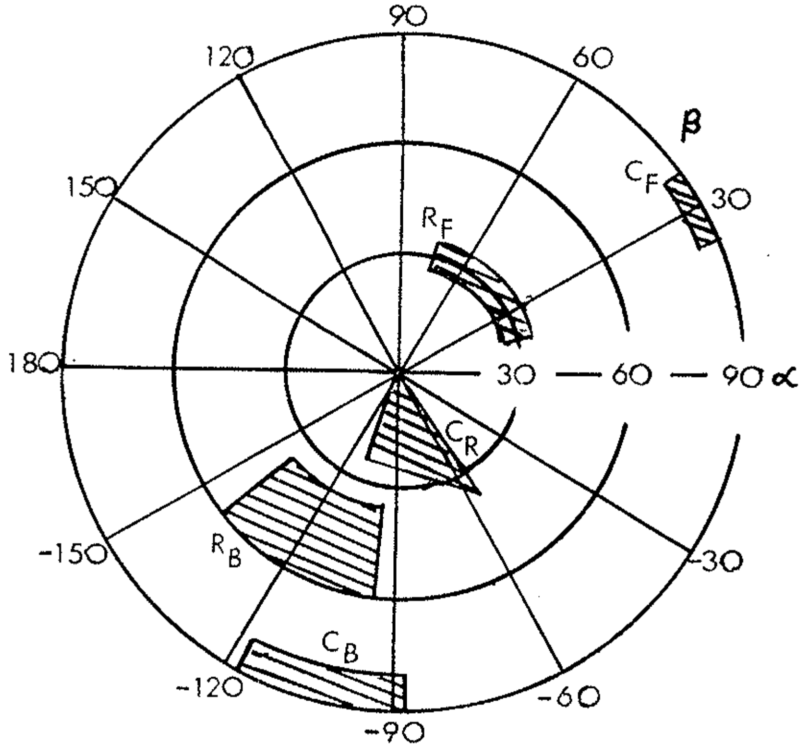
Mainly as a result of Matrix H broadcasting experience gained by the BBC during the 1977 Promenade Concert season, the BBC and NRDC jointly decided that it would be advantageous to make minor alterations to their respective systems, BBC Matrix H and NRDC System 45J. In either case the change primarily affects the phase of signals related to a centre-front source. This change improves stereo compatibility with very little effect upon either mono or surround reproduction. Thus existing decoders do not require any alteration.

As a result of these changes the two systems moved even closer together and it has become possible to agree upon a unified specification which takes the form of tolerance zones on the Scheiber Sphere in which the encoding points will lie. The agreed changes have also been noted by Nippon-Columbia, with whom NRDC has a working agreement.

The BBC and NRDC are therefore pleased to announce that their joint system will be known as BBC/NRDC System HJ to mark the fusion of two extensive programmes of research and development. Future BBC surround-sound broadcasts will use this system. Collaboration between the BBC and NRDC in the fields of production and microphone techniques and decoder technology will continue.

Engineering Information Department,
Publicity Section.

SYSTEM HJ TOLERANCE ZONES



Stage location	$\lvert\alpha\rvert$	$\lvert\beta\rvert$
C_F	90 ± 5	30 ± 6
R_F	30 ± 5	45 ± 30
C_R	Bounded by points (α=0)	(α= -25, β= -110), (α= 35 , β= -58)
R_B	47.5 ± 12.5	-117.5 ± 22.5
C_B	90 ± 5	-102.5 ± 15
L_B	132.5 ± 12.5	-117.5 ± 22.5
C_L	Bounded by points (α=180)	(α= -155, β= -110), (α= 145 , β= -58)
L_F	150 ± 5	45 ± 30

Technical note

For any particular system there always exists a multiplicity of loci for different but equally acceptable microphone techniques. The 'tolerance zones' method is a useful way of specifying the practical implementation of the system. A single-line locus is valid for only one microphone set-up and, if only the coefficients of the encoding matrix are specified, the strong influence of microphone technique on the overall system is ignored. The design of the decoder must take into account all possible loci falling within the tolerance zones given.

In favourably noting and analysing this joint specification, Nippon-Columbia have found that, in practice, intercompatibility between their existing System UD-4 and System HJ can be entirely achieved subjectively by merely readjusting the rear amplifier gains by plus or minus 2dB in reproducing UD programmes with HJ decoders or HJ programmes with UD decoders respectively.