Optimal Reproduction Matrices for Multispeaker Stereo

3180 (U-3)

Michael A. Gerzon
Oxford OX1 2JQ, UK

6E/U-3

Presented at
the 91st Convention
1991 October 4–8
New York

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

Optimal Reproduction Matrices for Multispeaker Stereo

Michael A. Gerzon

52 Walton Crescent, Oxford OX1 2JQ, U.K.

Abstract

This paper describes the psychoacoustically optimal matrices for reproducing stereophonic signals intended for reproduction via n_1 loudspeakers via a larger number n_2 of loudspeakers, using velocity and sound-intensity theories of sound localisation, for the specific cases $n_1 = 2,3,4$ and $n_2 = 3,4,5$. Optimal decoder designs for reproducing 2-channel stereo via 3 loudspeakers are described in detail, using frequency-dependent matrix coefficients, along with uses in music and TV applications where a stable stereo image over a large listening area is required.

1. INTRODUCTION

Since the earliest days of stereo, it has been realised that reproduction of a stereo signal intended for a given number of loudspeakers over a greater number of loudspeakers might give improved results. In particular, the reproduction of two channel stereo via three loudspeakers has been proposed and used by Bell Telephone Laboratories [1] in 1933 and Paul Klipsch in the 1950s [2-4] among others, where the centre loudspeaker was fed with the average of the left and right channels with an additional gain factor. This "bridged centre speaker" system has a number of claimed advantages, including better phantom imagery with spaced-microphone recordings and a more stable central image across a large listening area, but the improvement over two-speaker stereo has not been so overwhelming that this system has come into widespread use.

With the advent of stereo television, the instability of central images over two-speaker stereo has become a severe problem, since the position of the central sound is not matched to the accompanying visual image unless one sits in a single "stereo seat" location or the stereo loudspeakers are placed close together. While this problem can be solved by adaptive gain-riding or variable matrix circuitry, such signal-dependent reproduction has audible side effects which are incompatible with high quality results.

This paper presents the results of investigations to find the optimum method of deriving loudspeaker feeds from two stereo channels for

feeding three (or more) loudspeakers. Using a combination of a theoretical method of analysing auditory localisation and extensive listening tests to an adjustable decoding matrix, an optimised three-speaker decoder for two-channel stereo has been found that not only improves central image stability over a large listening area, but which retains a wide stereo spread over that listening area, and actually improves the quality of stereo images even for a listener seated in the ideal stereo seat.

This optimum 3×2 matrix decoder turns out to be frequency-dependent in order to take into account the different properties of human hearing at different frequencies. Because it involves no gain-riding or signal-dependent adaptive behaviour, the results give very low listening fatigue - certainly lower than ordinary two-speaker stereo - as well as convincing stereo images for listeners across a large listening area, with markedly better results than the older "bridged centre speaker" method.

The discovery and optimisation of the 3×2 decoder case led to a general theory and the discovery of optimal decoders for converting n_1-speaker stereo signals for reproduction via a larger number n_2 of loudspeakers for other values of n_1 and n_2. In this paper we present an outline of the general theory and explicit solutions for the $n_2 \times n_1$ matrix decoder for $n_1 = 2,3$ and 4 and for $n_2 = 3,4$ and 5, so that the optimal method of, say, reproducing 4-speaker stereo over 5 loudspeakers can be determined.

In connection with HDTV applications, there has been considerable controversy over how many loudspeakers are required for front-stage stereo to match the high-definition image, with different workers claiming that three and four is the optimum number, each with an associated transmission channel (e.g. see [5-7]). While the trade-off between localisation accuracy and a convenient number of loudspeakers is ultimately a subjective choice, the use of a 4×3 matrix decoder permits 4-speaker reproduction from a 3-speaker stereo transmission, and a 5×3 or 5×4 matrix decoder permits 5-speaker stereo reproduction from a 3- or 4-speaker stereo transmission, so that a given transmission standard does not limit the number of loudspeakers that can be used by the listener/viewer.

The optimum $n_2 \times n_1$ matrix decoder is not "obvious" - indeed its computation from the localisation theory given here requires both some use of the theory of orthogonal matrices and the numerical solution of systems of simultaneous nonlinear equations. The number of variables, and hence the complexity, of these equations grows rapidly with the number n_2 of reproduction loudspeakers, and computers are required to solve the more complex cases. In this paper, we have only solved the case with up to 5 loudspeakers, although the methods used are applicable also to six or more loudspeakers.

The psychoacoustic requirements on matrix decoders for multispeaker stereo fall into two areas: First we require that such decoders should preserve the total energy of all input stereo signals. We have found that such energy preservation is not merely an aesthetic requirement to preserve the level-balance among different sounds, but has psychoacoustic significance as will be explained. Secondly, we require that the localisation properties of the reproduced sound over n_2 loudspeakers should be as similar as possible to that originally intended via n_1 loudspeakers - and possibly better. To evaluate localisation performance, we use two main classes of sound localisation theory: one based on acoustic velocity, and the other based on energy and sound intensity. These theories have previously been applied to Ambisonic surround sound [8-10], but the front-stage stereo application involves some detailed differences in approach.

The energy-preservation requirement leads us to make use of the theory of orthogonal matrices, and this leads to a parameterisation of possible decoder matrices having a geometrical character. The localisation theory is used to find the optimum values of these parameters. Especially in the cases n_1 equals 3 or 4, the values of these parameters is quite tightly constrained if optimum subjective results are required, and we have also found in practice that the range of parameter values for a conventional 2-channel stereo source has little leeway if optimum results are required.

2. STEREO SPEAKER LAYOUTS AND MS MATRICES

Figures 1a to 1e shows possible loudspeaker layouts using from one to five loudspeakers. From left to right, the loudspeakers, and their associated feed signals, are denoted by C_1 for a 1-speaker (mono) speaker layout as in figure 1a, L_2 and R_2 for a 2-speaker stereo layout as in fig. 1b, L_3, C_3 and R_3 for a 3-speaker layout as in figures 1c, 1f or 1g, L_4,L_5, R_5 and R_4 for a 4-speaker layout as in fig. 1d, and L_6, L_7, C_5, R_7 and R_6 for a 5-speaker stereo layout as in fig. 1e.

In these layouts, a central loudspeaker is denoted by C_p for a numerical subscript p, and left loudspeakers and their mirror-image right counterparts by L_p and R_p respectively. The speakers L_p are assumed to lie at an angle θ_p anticlockwise from due front, and the mirror--image speakers R_p are assumed to lie at an angle θ_p clockwise from due front at the listener. While a wide variety of n-speaker stereo layouts are possible, in the main we shall deal with layouts such as shown in figs. 1b to 1f where all loudspeakers are at the same distance from an ideally-positioned listener, and where the angle subtended at that listener between all adjacent pairs of loudspeakers are identical, so that $\theta_5 = \frac{1}{3}\theta_4$ and $\theta_7 = \frac{1}{2}\theta_6$.

In figures 1a to 1e, we show the case where all loudspeakers face the ideally situated listener. However, in practice, there are often practical advantages in 'toeing in' the outer loudspeakers so that their axes cross in front of the listener, as shown in figures 1f and 1g, since this is found to enlarge the listening area. In figure 1g, in addition, the loudspeakers lie along a straight line rather than a circle centred at the listener - a layout that generally gives less good results than the circular layout of fig. 1f, but which is often easier to accomodate in a listening environment.

For reasons of theoretical convenience, most of the analyses of psychoacoustic performance of matrix decoders is done for layouts in which all loudspeakers are equally distant from the ideal listener position as shown in figs. 1a to 1e, although these decoders can be applied to other layouts.

It is generally more convenient to describe matrix decoders not directly in terms of the loudspeaker feed signals L_p and R_p, but in terms of signals in "sum-and-difference" or MS form, where we define "sum" signals to be those either of the form C_p or of the form

$$M_p = 2^{-\frac{1}{2}}(L_p+R_p) \tag{1}$$

and "difference" signals to be those of the form

$$S_p = 2^{-\frac{1}{2}}(L_p-R_p) \ . \tag{2}$$

Signals in MS form can be reconverted back into direct or left/right form by the inverse MS matrix equations:

$$L_p = 2^{-\frac{1}{2}}(M_p+S_p)$$
$$R_p = 2^{-\frac{1}{2}}(M_p-S_p) \ . \tag{3}$$

A matrix circuit satisfying this equation will be termed an "MS matrix". It will be seen that an MS matrix is its own inverse, i.e. the effect of cascading two MS matrices is to restore the original signal.

The advantage of examining signals in MS form rather than left/right form is that in matrix decoders, sum signals are converted into sum signals, and difference signals into difference signals, so that one needs to consider fewer matrix parameters.

Also, the total energy of stereo signals in MS form is the same as in left/right form, since

$$M_p^2 + S_p^2 = L_p^2 + R_p^2 \ , \tag{4}$$

as can easily be checked using simple algebra.

3. MATRIX DECODERS AND ENERGY PRESERVATION

Figure 2 shows the general schematic of a matrix reproduction decoder accepting n_1 signals from a stereo source intended for n_1-speaker stereo reproduction, and converted by the matrix reproduction decoder into n_2 signals intended for reproduction via an n_2-speaker stereo layout. The action of the matrix reproduction decoder can be described by an $n_2 \times n_1$ matrix $R_{n_2 n_1}$. If the matrix reproduction decoder is left/right symmetric in its behaviour, it can be implemented in the form shown in figure 3, where the n_1 input speaker-feed signals are converted by MS matrices into sum-and difference form, the sum signals are passed into a "sum-signal" matrix A, the difference signals are passed into a "difference-signal" matrix B, and the results passed into a further set of output MS matrix circuits to produce the

n_2 output loudspeaker-feed signals.

We have found that it is highly desirable that all matrix reproduction decoders should substantially preserve the total energy reproduced into the room. The obvious reason for this is that if the total stereo energy is preserved, the the level-balance between different sounds within a stereo mix will be preserved, which is of some artistic importance.

However, this is not the most important reason for decoders having an energy-preserving property, since the ears are actually quite tolerant of alterations in mere level-balance that can cause other more serious psychoacoustic effects. There are two important psychoacoustic reasons why energy-preservation is desirable.

The first is that the level-balance between direct sounds and early reflections in a recording environment conveys important cues about sound-source distance (e.g. see Mershon & Bowers [11].)
If a decoder is not energy-preserving to a significant degree, it will significantly degrade the sense of sound-source distance in those recordings (e.g. UHJ recordings using a SoundField microphone) having it.

The second reason is a little more complicated to explain, but is related to the fact that there is no concensus as to the ideal type of stereo recording technique. The reason for this lack of concensus probably lies in the fact that no recording technique can capture all aspects of the stereo illusion perfectly, and different people, quite legitimately, have different balances of priorities about which aspects are most important. Many stereo microphone techniques, and electronic stereo panning techniques, make use of both amplitude and time delay to create a stereo illusion. If such recordings are passed though a network that, to a significant degree, fails to preserve the total stereo energy, then such time-delay recordings will suffer from comb-filter addition and cancellation effects between the different delays, resulting in audible colourations. We have found empirically over the years by investigating forms of stereo signal processing that do and do not preserve energy that the ears are much less sensitive to mere redistributions of the stereo position with frequency due to comb filter effects than to actual comb filter variations in the total reproduced energy. (e.g. see [12])

Thus in general, it is found that energy-preserving decoders suffer from far less audible colouration over a wide variety of recording techniques than do decoders with marked variations of energy gain for different input stereo signal components.

There is a third quite subtle reason for why the energy-preserving property is particularly important in systems handling frontal-stage stereo, to do with total stage width. It is generally found that most attempts at matrixing n_1-speaker stereo signals for reproduction through a larger number of loudspeakers cause a significant loss of the stereo width as a proportion of the total angular width subtended

at the listener by the stereo loudspeaker layout. This is certainly the case with the bridged centre speaker system - and Klipsch [2-4] has noted the need to use a very wide loudspeaker layout to mitigate this effect.

If one places stringent requirements of the quality of directional localisation of such reproduction, as we shall do in this paper, it is found that the widest stereo images are obtained if the decoder is energy-preserving. This is not obvious nor is it easily proved mathematically, but appears to be true to a high degree of approximation in cases we have investigated in detail.

$n \times n$ matrices that preserve total signal energy for all signals passing through them are termed "unitary" if they have complex-valued coeficients (e.g. if they are implemented using frequency-dependent filter networks) and "orthogonal" if they have real coefficients, as is the case when a matrix circuit is not frequency-dependent.

For $n_2 \times n_1$ matrix circuits with n_1 inputs and a greater number n_2 of outputs, the network is energy-preserving if its matrix is the first (or indeed any) n_1 columns of an $n_2 \times n_2$ unitary or orthogonal matrix in the case of respective complex-valued and real matrix coefficients. For n_2 less than n_1, it is not possible for an $n_2 \times n_1$ matrix to be energy-preserving since some signal inputs result in a zero signal output.

In this paper, we shall confine our attention to matrix decoders that have substantially real coefficients and that are either totally frequency independent, or which are broadly frequency independent across two or three broad frequency bands and whose frequency dependence is largely confined to the transition regions between those frequency bands. Thus we shall only analyse in detail the case with real matrix coefficients.

If the overall decoder of fig. 2 is energy preserving, then so is matrix A and matrix B in the implementation of figure 3, since MS matrices preserve energy as we have seen earlier. Thus for real coefficients, matrices A and B are each formed from columns of respective orthogonal matrices.

Now the general form of $n \times n$ orthogonal matrices is well understood by mathematicians. Such matrices either describe rotations in n-dimensional space, or else describe the effect of a rotation preceeded by a reflection about an axis. For the purposes of this paper, it is not necessary to describe the general case in detail, although it is worth mentioning that $\frac{1}{2}n(n-1)$ free parameters are required to specify an arbitrary rotation matrix in n dimensions.

For this paper, we need only examine the general form of 2×2 and 3×3 orthogonal matrices. All 2×2 rotation matrices have the form

$$\begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}, \tag{5}$$

where ϕ is an angle parameter, although by replacing ϕ by $90^o-\phi$, one can interchange the sines and cosines in this expression. The other 2×2 orthogonal matrices, associated with a reflection followed by a rotation have the form

$$\begin{bmatrix} \cos\phi & \sin\phi \\ \sin\phi & -\cos\phi \end{bmatrix}. \tag{6}$$

The general form of 3×3 rotation matrices is a little more complicated, and moreover can be parameterised in many different ways. One parameterisation, representing a rotation by an angle ϕ about an axis vector (a,b,c) of unit length, i.e. such that

$$a^2 + b^2 + c^2 = 1 , \tag{7}$$

has a matrix of the form

$$\begin{bmatrix} a^2+(1-a^2)\cos\phi & ab(1-\cos\phi)+c\sin\phi & ac(1-\cos\phi)-b\sin\phi \\ ab(1-\cos\phi)-c\sin\phi & b^2+(1-b^2)\cos\phi & bc(1-\cos\phi)+a\sin\phi \\ ac(1-\cos\phi)+b\sin\phi & bc(1-\cos\phi)-a\sin\phi & c^2+(1-c^2)\cos\phi \end{bmatrix} \tag{8}$$

and the 3×3 orthogonal matrices involving a reflection have the same form as equation (8) except that the signs of the last column are reversed.

In practice, the parameterisation of equation (8) is rarely the most convenient, and in general, an $n\times n$ matrix with real coefficients is orthogonal if and only if (i) all its columns are vectors of unit length, i.e. the sum of squares of its entries is one, and (ii) any two columns have a zero inner product, i.e. the sum of the products of the corresponding entries of the two columns is zero. This characterisation is often a useful way of constructing orthogonal matrices for particular applications.

We shall now describe the general form of the matrix A and matrix B equations for energy-preserving decoders having the form of figure 3, taking into account that we have general desired properties such as that the reproduced stereo should be the right way round.

For an energy-preserving 3×2 matrix decoder, the matrix A equation is of the form

$$\begin{bmatrix} M_3 \\ C_3 \end{bmatrix} = \begin{bmatrix} \sin\phi \\ \cos\phi \end{bmatrix} \begin{bmatrix} M_2 \end{bmatrix} \tag{9a}$$

and the matrix B equation is simply

$$S_3 = S_2 , \tag{9b}$$

where the angle parameter ϕ lies between 0^o and 90^o.

For an energy preserving 4×3 matrix decoder, the matrix A equation is of the form

$$\begin{bmatrix} M_4 \\ M_5 \end{bmatrix} = \begin{bmatrix} \cos\phi_3 & -\sin\phi_3 \\ \sin\phi_3 & \cos\phi_3 \end{bmatrix} \begin{bmatrix} M_3 \\ C_3 \end{bmatrix} \tag{10a}$$

and the matrix B equation is of the form

$$\begin{bmatrix} S_4 \\ S_5 \end{bmatrix} = \begin{bmatrix} \cos\phi_D \\ \sin\phi_D \end{bmatrix} \begin{bmatrix} S_3 \end{bmatrix} \tag{10b}$$

where ϕ_3 and ϕ_D are angle parameters, where ϕ_D certainly lies between 0^o and 90^o and ϕ_3 certainly lies somewhere in the range say of -30^o to $+60^o$.

The form of an energy-preserving 5×4 decoding matrix is more complicated, and we give a parameterisation here with no detailed explanation of how we arrived at it, but the reader can check that the columns of the matrix A equations are indeed orthogonal and of unit length. The matrix A equations of the 5×4 matrix decoder have the form

$$\begin{bmatrix} M_6 \\ M_7 \\ C_5 \end{bmatrix} = 2^{-\frac{1}{2}} \begin{bmatrix} a+\mu_1\cos\phi_4+\nu_1\sin\phi_4 & a-\mu_1\cos\phi_4-\nu_1\sin\phi_4 \\ b-\mu_2\cos\phi_4+\nu_2\sin\phi_4 & b+\mu_2\cos\phi_4-\nu_2\sin\phi_4 \\ c-\lambda\sin\phi_4 & c+\lambda\sin\phi_4 \end{bmatrix} \begin{bmatrix} M_4 \\ M_5 \end{bmatrix} \tag{11a}$$

where ϕ_4 is an angle parameter, (a,b,c) is a unit length vector with positive entries a, b and c, and where

$$\lambda = (a^2+b^2)^{\frac{1}{2}} , \mu_1 = b/\lambda , \mu_2 = a/\lambda , \nu_1 = ac/\lambda , \nu_2 = bc/\lambda , \tag{12}$$

and the matrix B equations have the form

$$\begin{bmatrix} S_6 \\ S_7 \end{bmatrix} = \begin{bmatrix} \cos\phi_5 & -\sin\phi_5 \\ \sin\phi_5 & \cos\phi_5 \end{bmatrix} \begin{bmatrix} S_4 \\ S_5 \end{bmatrix} \tag{11b}$$

where ϕ_5 is an angle parameter that is likely to be well within the range from -45^o to $+45^o$.

The form of an energy preserving 4×2 decoding matrix has a matrix A equation of the form

$$\begin{bmatrix} M_4 \\ M_5 \end{bmatrix} = \begin{bmatrix} \sin\phi_{42} \\ \cos\phi_{42} \end{bmatrix} \begin{bmatrix} M_2 \end{bmatrix} \tag{13a}$$

and the matrix B decoding equation has the form

$$\begin{bmatrix} S_4 \\ S_5 \end{bmatrix} = \begin{bmatrix} \cos\phi_D \\ \sin\phi_D \end{bmatrix} \begin{bmatrix} S_2 \end{bmatrix} \qquad (13b)$$

where ϕ_{42} and ϕ_D are angle parameters between 0^o and 90^o.

Figures 4, 5 and 6 show the forms of energy-preserving 3×2, 4×2 and 4×3 matrix decoders satisfying the above equations, where the decoders with 2-channel inputs are also provided with a width control gain adjustment of the difference signal S_2 by a gain factor w so that adjustments of stereo width are possible. Other $n_2\times n_1$ energy prerserving matrix decoders are similarly implemented as shown in figure 3.

For a 5×2 decoder, the matrix A equations are a 3×1 matrix whose entries are positive and whose squares add up to one, and the matrix B equations are a 2×1 matrix whose entries are positive and whose squares add up to one. For a 5×3 energy-preserving matrix decoder, the matrix A equations are a 3×2 energy preserving matrix and the matrix B equations are a 2×1 energy preserving matrix. The detailed form of these matrix A and matrix B equations is indicated in the next section of this paper.

For a 6×5 energy-preserving matrix decoder, the matrix A equations use a 3×3 rotation matrix with large positive diagonal elements, and the matrix B equations use a 3×2 energy-preserving matrix that is the first 2 columns of a 3×3 rotation matrix. We shall not give the details of this case here. Cases involving more than $n_2 = 6$ reproduction loudspeakers in the stereo playback involve the need to consider $n\times n$ orthogonal matrices with n equal to four or more, and 4×4 rotation matrices have a quite complicated general form involving 6 free parameters. Fortunately, the case with n_2 greater than six rarely needs to be considered in practice for front-stage stereo.

4. COMPOSITE DECODERS

If we have $n_1 < n_3 < n_2$, then we can form an $n_2\times n_1$ matrix decoder accepting n_1-speaker stereo input signals and giving outputs for reproduction via an n_2-speaker stereo layout by series connection of an $n_3\times n_1$ matrix decoder with an $n_2\times n_3$ matrix decoder as illustrated in fig.7. Such a series connection of two matrix decoders, or a decoder having the same matrix equations as such a series connection, is termed a <u>composite</u> decoder. If the component decoders are energy-preserving, then so evidently is their series connection. If the $n_3\times n_1$ matrix decoder is described by a matrix $R_{n_3n_1}$ and the $n_2\times n_3$ matrix decoder is described by the matrix $R_{n_2n_2}$, then the $n_2\times n_1$ composite decoder of figure 7 is described by the product matrix

$$R_{n_2n_1} = R_{n_2n_3} R_{n_3n_1} \qquad (14)$$

The 4×2 energy-preserving matrix decoder referred to in equations (13) above is a composite decoder formed from the series connection of a 3×2 decoder as in equations (9) and a 4×3 decoder as in equations (10), where the ϕ_D parameter is the same in the two representations of the decoder, and

$$\phi_{42} = \phi - \phi_3 . \qquad (15)$$

In general, the composite decoder formed by the series connection of two energy preserving decoders of the form shown in fig. 3 can also be implemented as in fig. 3 using a matrix A that is the product or series connection of the component matrix A 's, and a matrix B that is the product or series connection of the component matrix B 's.

Thus a 5×3 energy preserving matrix decoder can be expressed as the series connection of the 4×3 matrix decoder of equations (10) with the 5×4 matrix decoder of equations (11), by multiplying the associated matrices. Similarly, a 5×2 energy-preserving matrix decoder can be expressed as the series connection of a 4×2 matrix decoder as in equations (13) with a 5×4 matrix decoder as in equations (11) - although a given decoder matrix can generally be expressed as a product of two component matrices in more that one way.

Composite decoders are important since they tend to preserve the more desirable properties of their component decoders. This applies not only to the energy-preservation property, but to other aspects of sound that are preserved. For example, if the component decoders are so designed so as to substantially preserve a given aspect of the stereo localisation of sounds passed through them, so will the composite decoder formed by their series connection.

This leads to the observation that, for arbitrary $n_2 > n_1$, an $n_2\times n_1$ reproduction matrix decoder can be expressed as a product of $(n+1)\times n$ reproduction matrix decoders for all intermediate values of n. Figure 8 shows how signals intended for n_1-speaker stereo can be passed through a number of $(n+1)\times n$ matrix reproduction decoders to achieve reproduction via a larger number n_2 of stereo loudspeakers.

This figure only shows the case up to $n_2 = 5$, although its extension to larger numbers of loudspeakers is obvious. If one can find psychoacoustically good $(n+1)\times n$ decoders for each n, then one can form the composite decoders for every $n_2 > n_1$ simply by forming the matrix product or series connection of the component decoders as shown in figure 8, and also mix a stereo signal intended for one number of loudspeakers with those for any other numbers of loudspeakers as indicated by the addition nodes in figure 8.

The possibility of this kind of hierarchy of matrix reproduction decoders for n-speaker stereo means that to design good-sounding $n_2\times n_1$ matrix decoders, one needs only to do detailed design work in the case that $n_2 = n_1+1$, and implement the other cases as composite decoders.

5. DIRECTIONAL PSYCHOACOUSTICS

A summary of a theoretical framework for the perception of directional stereo effect is now given. It is not claimed that the following theory is a perfect predictor of all aspects of stereo perception, since no simple mechanistic model can capture all the complexity of auditory perception. However, the theory given here, which was first presented in a simple form in [8], is quite a good predictor at least up to frequencies around 3.5 kHz. Its real function is to provide detailed guidelines on the design of stereo reproduction systems without claiming to be exact or exhaustive.

The most detailed account of the theory has been provided in [9], where the theory was set in a general "metatheoretic" context which, in principle can be extended to provide an exact description of directional psychoacoustics. A useful summary of the present theory was provided in the preprint [10]. These previous descriptions were applied to Ambisonic surround-sound, and gave little explicit information about the applications of the theory to non-central listeners, whereas in the present stereo application, it is the intention to stabilise the relative position of stereo images with respect to the frontal stereo speaker layout for listeners well away from the ideal stereo seat. (In Ambisonics [13], the intention is more to stabilise the absolute direction of images).

Figure 9 shows a stereo speaker layout disposed on a sector of a circle centred at an ideal stereo seat position, with all speakers equally distant from the ideally situated listener. We use rectangular (x,y) axes with the x-axis pointing forward and the y-axis pointing due left as shown. For n loudspeakers, indicated by subscripts i = 1 to n, the i'th loudspeaker is disposed in a direction measured by the angle $\hat{\theta}_i$ measured anticlockwise from the x-axis, i.e. from the x-axis towards the y-axis, so that $\hat{\theta}_i$ is positive for speakers to the left of centre and negative for speakers to the right of centre.

Suppose that a sound is fed to all of the n loudspeakers of fig. 9 with a gain G_i, which may be real or complex, to the i'th loudspeaker. Then one can compute objective physical properties of the resulting sound field at the ideal stereo seat, including the (complex) gains of the pressure and the velocity vector, and of the total energy and of the sound intensity vector, and relate these to the resulting stereo illusion. This gives rise to two sub-theories of localisation. One, termed the velocity theory, relates the localisation to the vector obtained by dividing the velocity vector gain by the pressure gain at the listener. The velocity theory is based on the ears using interaural phase cues to localise sounds, and is most apt at frequencies below 700 Hz - a frequency at which the ears are effectively spaced apart by half a wavelength.

The other sub-theory, which we term the energy vector theory, relates localisation to the ratio of the sound intensity vector gain (which describes the flow of sound energy) to the total energy. This theory is apt mainly at frequencies between about 700 Hz and 5 kHz for

central listeners, but it can be shown that if the sound arrivals from the n loudspeakers are phase-incoherent with one another (such as might be the case with very non-central listening positions), then the velocity vector theory gives the same predictions as the energy vector theory, so that for such listeners, the energy vector theory is useful even at frequencies well below 700 Hz. In addition, in practice, there is no sharp transition frequency such as 700 Hz for the two sub-theories, but rather a fuzzy band of frequencies at which both theories have some applicability. Certainly, interaural phase cues can still be used up to 2 kHz for sounds quite close to a frontal direction.

While this is not the place to go into details, it is in addition worth mentioning that it is possible to demonstrate that the ears use different localisation strategies for transient and steady-state continuous sounds, and that the theories just described apply in the main to continuous components of sounds. For transient sounds up to around 3.5 kHz, the Haas or precedence effect [14, 15] is used whereby earlier sound arrivals tend to influence localisation more than later arrivals, at least for delays less than about 40 ms. However, the energy vector theory can be used as a predictor also for Haas--effect transient localisation.

Above about 3.5 kHz, the Haas effect is largely inoperative, and the ability to form convincing phantom images from several sources is decreased, with a tendency for individual sound sources to become more audible. Nevertheless, the energy vector theory provides information about localisation at these frequencies, but is no longer a reasonably accurate predictor of the actual localisation direction.

With all these qualifications in mind, we state the two sub-theories in a usable mathematical and computational form.

For a central listener in figure 9, the total pressure gain at the listener is

$$P = \sum_{i=1}^{n} G_i \; , \qquad (16)$$

and the total energy gain at the listener is

$$E = \sum_{i=1}^{n} |G_i|^2 \; . \qquad (17)$$

The sound intensity vector gain at the listener is the vector quantity

$$e = (e_x, e_y) \qquad (18)$$

where the components e_x and e_y along the respective x- and y-axes are given by

$$e_x = \sum_{i=1}^{n} |G_i|^2 \cos\hat{\theta}_i \qquad (19a)$$

$$e_y = \sum_{i=1}^{n} |G_i|^2 \sin\hat{\theta}_i . \quad (19b)$$

Localisation according to the energy vector localisation theory is determined by the vector ratio e/E which has length $r_E \geqslant 0$ and angular direction θ_E given by

$$r_E \cos\theta_E = e_x/E$$
$$r_E \sin\theta_E = e_y/E \quad . \quad (20)$$

This <u>energy vector direction</u> θ_E is the apparent direction of sounds according to the energy vector localisation theory if the listener faces the apparent sound source. The <u>energy vector length</u> r_E is less than or equal to one, and is only equal to one if the sound emerges only from a single loudspeaker. Ideally $r_E = 1$ for a natural sound localisation, and $1 - r_E$ is roughly proportional to the angular movement of the apparent image direction relative to that of the reproducing loudspeakers when listeners move either laterally or rotate their heads. As a rule of thumb, this applies both to localisation of continuous sounds and Haas-effect localisation of transient sounds, although the actual degree of image movement is larger for the latter. Thus $r_E = 0.95$ gives about one third of the image movement of $r_E = 0.85$. For central images over the standard 60^o 2-speaker stereo layout, $r_E = 0.866$.

Thus, between them, θ_E and r_E provide good predictors both of image localisation for centrally placed listeners and of image stability for other listeners. Above 3.5 kHz, θ_E is not an accurate predictor of localisation, with a tendency for sounds to be pulled more strongly to the direction of the loudest loudspeakers than is predicted by the energy vector theory.

The velocity vector gain for a central listener is given by the vector

$$v = (v_x, v_y) \quad , \quad (21)$$

where

$$v_x = \sum_{i=1}^{n} G_i \cos\hat{\theta}_i \quad (22a)$$

$$v_y = \sum_{i=1}^{n} G_i \sin\hat{\theta}_i \quad . \quad (22b)$$

As before, the localisation according to the velocity vector localisation theory is given by the ratio

$$v/P = (v_x/P, v_y/P) \quad , \quad (23)$$

with the complication that in general this vector has complex-valued entries. However, low-frequency interaural phase localisation theories predict [8,9] that only the real part

$$(\mathrm{Re}(v_x/P), \mathrm{Re}(v_y/P)) \quad (24)$$

of this vector contribute to the apparent localisation, although it

is found that the imaginary component

$$(\mathrm{Im}(v_x/P), \mathrm{Im}(v_y/P)) \quad (25)$$

causes image broadening and an unpleasant subjective sensation termed "phasiness", and also contributes to localisation at middle frequencies around 600 Hz. In practice, if the length of the "phasiness vector" of equation (25) is less than about 0.2, then such phasiness effects can largely be ignored. The phasiness vector is zero for central listeners if all the gains G_i are real-valued, as is the case for all non-frequency-dependent decoders considered in this paper, although a small degree of phasiness is introduced by phase shifts in the filters in frequency-dependent decoders unless special phase-compensation precautions are taken.

The real component (24) of the ratio of velocity gain to pressure gain can be used to predict localisation. This vector has length $r_V \geqslant 0$ and angular direction θ_V given by the equations

$$r_V \cos\theta_V = \mathrm{Re}(v_x/P)$$
$$r_V \sin\theta_V = \mathrm{Re}(v_y/P) \quad . \quad (26)$$

The <u>velocity vector localisation</u> θ_V, also termed the <u>Makita localisation</u> after the work of Makita [16,17], is the apparent direction of sounds according to interaural phase theories if the listener faces the apparent direction of the sound. For natural single sound sources, $r_V = 1$, and departures of r_V from one cause instability of images as the listener rotates his/her head. r_V greater than one increases apparent image width, and r_V less than one reduces it, but in general, values of r_V close to one, say in the range 0.85 to 1.1, are found not to degrade image quality greatly, but much smaller (e.g. less than 0.7) or much larger (e.g. greater than 1.3) values of r_V are generally unacceptable.

For listeners at different distances from the loudspeakers, the above theories still apply, except that the gains G_i must be relaced by

$$(G_i/d_i)\, e^{-j\omega\tau_i} , \quad (27)$$

where d_i is the distance of the i'th loudspeaker, ω is the angular frequency of a sound and τ_i is the time delay of sound arrivals from the i'th loudspeaker.

6. 3 × 2 DECODERS

The above psychoacoustic theories are now applied to the reproduction of standard 2-channel stereo signals L_2 and R_2 via a 3×2 energy preserving matrix decoder of the form shown in figure 4 via 3 stereo loudspeakers L_3, C_3 and R_3, as shown in figure 10. We assume the use of a 3-speaker stereo layout for which $\theta_3 = 45^o$, i.e. which subtends a total angle of 90^o at the listener, as shown in figs. 1c or 1f.

It is convenient to describe the stereo position of sounds in a 2-channel stereo signal in terms of a <u>panpot angle</u> parameter θ.

where 45° θ -45°, such that the L_2 signal has gain $\cos(45^{\circ}-\theta)$ and the R_2 signal has gain $\cos(45^{\circ}+\theta)$. This is the gain law of a constant--power or sine/cosine panpot, of the kind discussed for example by Orban [18], and is such that the sound is panned to the left for $\theta = 45^{\circ}$, at the centre for $\theta = 0^{\circ}$, and to the right for $\theta = -45^{\circ}$.

Figure 11 shows the psychoacoustic localisation parameters r_V, θ_V, r_E and θ_E computed for a sound with panpot angle θ when reproduced via a conventional 2-speaker stereo layout, such as that of fig. 1b, with $\theta_2 = 35^{\circ}$, i.e. subtending a total of 70° at the stereo seat. This value has been chosen rather than the more conventional 60° sector in order to provide a better comparison with the 3-speaker reproduction data. It will be seen that, as expected, the velocity and energy vector localisation directions are the same for left, centre and right positions, but that at intermediate positions, the energy vector localisation angle θ_E is wider than the velocity localisation θ_V, being about twice as wide for near-centre locations. This agrees with the well-established observation [19-21] that high-frequency 2-speaker stereo localisation gives wider images than low-frequency localisation. Also note the 'detent' effect for energy-vector localisation whereby the localisations of sounds towards the edges of the stereo stage are pulled into the nearest loudspeaker - a phenomenon noted by Harwood [19] experimentally.

It will be seen that the values of r_V and r_E at the two edges of the stereo stage equal one, denoting the expected good image stability of sounds emerging from only one loudspeaker, but that centre-stage images have markedly reduced values of r_V and r_E, indicative of poor image stability.

Figures 12 to 15 show the psychoacoustic localisation parameters r_V, θ_V, r_E and θ_E computed for a 2-channel stereo sound with panpot angle θ when reproduced via the 3×2 energy-preserving matrix decoder of figure 4 with $w = 1$ and of equations (9) over the layout of figure 1c or 1f with $\theta_3 = 45^{\circ}$, for respective values of the decoder angle parameter $\phi = 90^{\circ}$, $\arctan\sqrt{2} = 54.74^{\circ}$, $\arctan 2^{-\frac{1}{2}} = 35.26^{\circ}$ and and $\arcsin\frac{1}{3} = 19.47^{\circ}$. The $\phi = 90^{\circ}$ decoder gives exactly the same results as 2-channel stereo reproduction via 2 speakers with $\theta_2 = 45^{\circ}$, and it will be seen that the results shown in fig. 12 are broadly similar to the narrower 2-speaker results shown in fig. 11 except that: (i) all reproduced direction angles are wider, and (ii) the values of $1 - r_V$ and of $1 - r_E$ are increased by a factor of about 1.62, resulting in stereo images that are considerably more unstable, as found in practice over such a wide loudspeaker layout.

Figure 13, with $\phi = 54.74^{\circ}$, shows a somewhat narrower stage width of reproduction with the relative width of the energy vector localisation relative to the velocity localisation even greater than for the 2-speaker stereo of figure 11. However, the centre-stage values of r_V and r_E indicate that the image stability of central images is now only a little poorer than for figure 11. The broad trend of r_V is somewhat similar to that of fig. 11, but r_E is slightly reduced at the stage edges, giving a somewhat degraded image stability at these stereo positions, but still very good as compared to centre-stage images.

Figure 14, with $\phi = 35.26^{\circ}$, shows a virtually unchanged velocity vector localisation direction, but the energy vector localisation direction is now narrowed to the point where it is virtually the same as the velocity vector direction for all but the most extreme left and right panpot angles θ. The velocity vector magnitude r_V is now generally closer to the ideal value 1 than even for figure 10, and the energy vector magnitude r_E is very nearly constant in value across the whole of the stereo stage, indicating that the degree of instability of images is about the same for all stereo positions, and certainly $1-r_E$ has about one third of its value in figure 12 for centre-stage images, and is about 0.54 of its value in figure 11, giving much improved central image stability.

The main flaw with the localisation behaviour shown in fig. 14 is the excessively narrow energy vector localisation for extreme left and right images, which causes this decoder ($\phi = 35.26^{\circ}$) to lose a sense of adequate width.

As ϕ is reduced further to 19.47°, as in figure 15, the velocity vector localisation direction θ_V has hardly changed, but the energy vector localisation direction θ_E is now considerably narrower, causing a very narrow effect at high frequencies. r_V and r_E are now much closer to one, giving excellent image stability, for central images, but poorer image stability for edge-of-stage images. As ϕ tends to 0° (not illustrated), central image stabilty becomes perfect, θ_V is not much changed, but $\theta_E = 0^{\circ}$ at all positions, thereby losing high frequency image width.

Informal listening tests to the decoder of fig. 4 with different values of the parameter ϕ confirm the above theoretical predictions, and in particular that, for a central listener, $\phi = 35.26^{\circ}$ gives the sharpest and most convincing phantom images across about 75% of the stereo stage.

However, its loss of width at the two extremes of the stereo stage is found to be a marked defect. While not much can be done about this, one can greatly improve the subjective performance of a 3×2 matrix decoder by making ϕ frequency-dependent, so that different trade-offs apt at different frequencies can be made.

Figure 16 shows the block diagram of an energy-preserving frequency-dependent version of the 3×2 matrix decoder of figure 4. In this decoder, ϕ is made frequency dependent by preceeding the sine/cosine gain circuit in the M_2 signal path by a bandsplitting network, shown as a high-pass and complementary low-pass filter, with high frequencies going to a pair of gains $\cos\phi_H$ and $\sin\phi_H$ corresponding to a high--frequency value ϕ_H of ϕ , and with low frequencies going to a pair of gains $\cos\phi_L$ and $\sin\phi_L$ corresponding to a low-frequency value ϕ_L of ϕ.

If this bandsplitting filter has outputs that sum to its inputs, then the difference S_2 channel processing in figure 16 is unaltered, but if instead they sum to an all-pass response, then a parallel all-pass response should be placed in the S_2 signal path, indicated by the extra block before the width control gain w, to match the phase responses of the sum and difference channels. The extension of fig. 16 to the case with 3- or more way bandsplitting to provide three or more frequency bands in which ϕ takes on differing values is obvious.

For a decoder with two bands in which ϕ takes on different values, there is the problem of finding the optimum transition frequency between the bands and the optimum values of ϕ in each band. At very low frequencies, say below 150 Hz, the value of ϕ is uncritical, since we have seen that the low-frequency velocity vector localisation is largely independent of the value of ϕ, and in any case, the ears' sensitivity to stereo quality is somewhat diminished at very low frequencies. Over most of the stereo stage, we have seen from fig. 14 that $\phi = 35^\circ$ tends to be the best choice up to perhaps 3.5 kHz. The optimum choice of ϕ above, say, 5 kHz needs to be determined empirically, both for central and noncentral listeners, due to the failure of energy-vector and Haas effect localisation models at these frequencies.

On most programme material, it is found that the contribution of the frequencies above 5 kHz to the sense of a wide stereo stage is very important, as can be verified in conventional 2-speaker stereo by filtering out these frequencies. Therefore, a good strategy for ameliorating the loss of stage width with $\phi = 35^\circ$ at lower frequencies is to provide a wider reproduction just above 5 kHz. The best value of ϕ_H above 5 kHz has been determined by listening to a wide range of programme material through a steep-cut high-pass filter at 5 kHz. It is found that values of ϕ_H significantly smaller than 55° do not retain a full sense of stage width for both central and noncentral listeners, but that $\phi_H = 55^\circ$ or larger do. However, it is also found that ϕ_H much larger than 55° have a high-frequency "hole in the middle" with inadequate centre-stage sounds. The value $\phi_H = 55^\circ$ or thereabouts gives both a wide stage and a good spread of sound among all three speakers. Informal listening tests to full-range material confirm that a decoder of the form shown in fig. 16 with $\phi_L = 35^\circ$ and $\phi_H = 55^\circ$ or thereabouts seem to be the most satisfory both for listeners at the ideal stereo seat and for listeners well away from it.

An experimental decoder as shown in fig. 16 was built, with the values of ϕ_L, ϕ_H and the characteristics of the bandsplitting filter fully adjustable, and many hundreds of hours of listening to a wide range of different kinds of both commercial and private recordings covering the full range of current recording approaches has verified that the above values of ϕ_L and ϕ_H are consistently preferable to other choices. However, within limits, the characteristics of the bandsplitting filter are not so critical.

We have found that the transition frequency of the bandsplitting filter should not be below about 5 kHz, but that otherwise, the transition frequency is not very critical. Also, the cross-over between the two frequency ranges should not be too sharp since the ears object to sudden changes of behaviour with frequency, but otherwise the cross--over characteristics again seem not to be critical.

The uncritical nature of ϕ at very low frequencies means that one can alter ϕ below say 150 Hz to match loudspeaker bass characteristics. For example, if the centre speaker has less bass power handling than the outer speakers, then it is convenient to choose ϕ near 90° at these frequencies, whereas if only the centre loudspeaker can handle deep bass, a value of ϕ approaching 0° at very low frequencies is more apt. If all three speakers have limited bass power handling, a value of ϕ approaching 54.74° at very low frequencies is best, since this shares the extreme bass power of central images equally among the three speakers, maximising bass power handling and giving maximum reinforcement of bass from the three speakers.

If a frequency-independent decoder has to be used, we have found that $\phi = 45^\circ$ provides a reasonable compromise between image width and the image quality and stability requirements, and that even this decoder sounds markedly better than the Bell/Klipsch bridged centre speaker proposal [1-4], while being audibly inferior to an optimised frequency-dependent decoder, both for central and noncentral listeners.

In applications where central image stability is critical, such as stereo TV sound, image stability can be improved by reducing ϕ_L to say 25° or 20° while retaining ϕ_H at 55°. However, this central image improvement is at the expense of image quality at the edges of the stage, which affects not only, say, the width of musical and "stage-off" dramatic sounds, but also the sense of naturalism of ambient sound effects such as crowd noises, acoustics and rainfall sounds, such as are important for atmosphere in dramatic programmes, news material and sports broadcasts.

7. PRESERVATION DECODERS

While the 3×2 matrix decoders are optimally designed to improve on the inherent limitations of 2-speaker reproduction, the degree of improvement possible by reproducing 3- or 4-speaker stereo via more loudspeakers is somewhat reduced. In this part of the paper, we present our main results, which are the determination of what we term <u>preservation decoders</u>, i.e. $n_2\times n_1$ energy-preserving matrix decoders that substantially preserve the values of the stereo localisation parameters (apart from the effects of any overall change of angular stage width caused by the overall stage width of the speaker layout).

Because of the possibility indicated in fig. 8 of forming composite decoders, we only consider the design of $(n+1)\times n$ matrix preservation decoders. The exact form of a preservation decoder depends on the values of the angles θ_p of the speakers in the layouts assumed shown in figures 1b to 1e. However, using the design procedures described below, we have found that the form of the $(n+1)\times n$ preservation decoder matrix does not vary greatly for quite large variations in the values of the angles θ_p.

We have chosen, for the purposes of computations, to use the following standard reference values of the layout angles θ_p :-

$$\theta_2 = 35^o,\ \theta_3 = 45^o,\ \theta_4 = 50^o,\ \theta_5 = \tfrac{1}{3}\theta_4,\ \theta_6 = 54^o,\ \theta_7 = \tfrac{1}{2}\theta_6. \quad (28)$$

These reference values have been chosen because: (i) they are such that the angle between any adjacent pair of speakers in a layout is the same as between other adjacent pairs in the same layout, and (ii) the $(n+1)\times n$ preservation decoders computed below give almost unchanged angular width of images before and after the decoder; this lack of change of overall angular width makes before-and-after comparisons rather easier than they would otherwise be.

Unlike in the 2-channel case, there is no standard panpot law for 3- or more channel stereo - see the discussion in [22], so that one cannot optimise a decoder, as we did in the 3×2 case, simply by looking at its effect on sounds panned across the stereo stage. Rather, one has a wide variety of possible methods of "encoding" directional effect into the n-channel stereo signal. The n speaker gains G_i of fig. 9 can be chosen over qite a wide range of values to create images. The n-entry vector $(G_1,\ldots,G_n)$ thus lies in a region of n-dimensional space, and in designing preservation decoders, one needs to find convenient representative points broadly covering this region representative of different stereo positions, compute their localisation parameters r_V, θ_V, r_E and θ_E via n loudspeakers and also, via an $(n+1)\times n$ energy-preserving decoder, via (n+1) loudspeakers. One then needs to adjust the free parameters of the energy-preserving decoder until the reproduced localisation parameters via the matrix decoder are broadly similar or identical to the original localisation parameters.

This procedure is quite complicated in that the mathematical equations for the localisation parameters are highly nonlinear in the free parameters describing the decoders, and solutions are only obtainable by numerical solution using a computer.

Another problem is that we cannot expect every localisation parameter to be exactly preserved. For example, if r_E is close to one for a sound via n loudspeakers, a sound in the same direction via n+1 loudspeakers will have a smaller value of r_E since the sound will no longer be in a direction near one of the loudspeakers. One therefore has to prioritise localisation parameters in deciding exactly what constitutes a good approximation to "preserving" them. In practice we do this by concentrating on the localisation angles θ_V and θ_E and attempting to preserve these.

More precisely, we choose the following n-speaker "stereo test signal" speaker gains G_i: the cases where only one speaker is excited with $G_k = 1$ and $G_i = 0$ for $i \neq k$, and the cases where a pair of speakers is excited with $G_k = G_l = 1$ and $G_i = 0$ when $k \neq i \neq l$, i.e. with equal in-phase gains. This yield $\tfrac{1}{2}n(n-1)$ different test signals, for which $\theta_V = \theta_E$ and $r_V = r_E$, since $G_i = |G_i|^2$ for all loudspeakers. For a feed only to the k'th speaker $\theta_V = \theta_E = \hat{\theta}_k$ and $r_V = r_E = 1$, and for an equal feed to the k'th and l'th speaker

$$\theta_V = \theta_E = \tfrac{1}{2}(\hat{\theta}_k + \hat{\theta}_l) \quad \text{and} \quad r_V = r_E = \cos\tfrac{1}{2}(\hat{\theta}_k - \hat{\theta}_l). \quad (29)$$

Because of left/right symmetry in the speaker layouts and decoder equations, we need only consider one of these signals and not its mirror image. If the localisation parameters reproduced via an $(n+1)\times n$ energy-preserving decoder are indicated by r_V', θ_V', r_E' and θ_E', then the preservation decoder requirement is the equations :-

$$\theta_V' = \theta_E' \quad (30)$$

for all $\tfrac{1}{2}n(n+1)$ stereo test signal gains considered above. For left/right symmetric test signals, this is automatically true since symmetry implies $\theta_V = \theta_E = \theta_V' = \theta_E' = 0^o$, and if the equation (30) is true for a given stereo test signal, then it is also true for its left/right mirror image.

Thus left/right symmetry reduces the number of equations (30), and it can be shown that the number of equations (30) to be satisfied exactly equals the number of free parameters describing a left/right symmetric energy-preserving $(n+1)\times n$ matrix decoder, for any n.

In particular, for the 3×2 preservation decoder, we must find that value of the decoder parameter ϕ for which (30) is satisfied for (L_2,R_2) gains (1,0); for the 4×3 preservation decoder, we must find those values of the 2 decoder parameters ϕ_3 and ϕ_D for which (30) is satisfied for the (L_3,C_3,R_3) gains (1,0,0) and (1,1,0); for the 5×4 preservation decoder, we must find those values of the 4 free decoder parameters (a,b,c) with $a^2+b^2+c^2=1$, ϕ_4 and ϕ_5 for which (30) is satisfied for (L_4,L_5,R_5,R_4) gains (1,0,0,0), (0,1,0,0), (1,1,0,0) and (1,0,1,0); for the 6×5 preservation decoder, we must find those values of the six free decoder parameters for which equ.(30) is satisfied for the (L_6,L_7,C_5,R_7,R_6) gains (1,0,0,0,0), (0,1,0,0,0), (1,1,0,0,0), (0,1,1,0,0), (1,0,1,0,0) and (1,0,0,1,0) ; and so forth for larger n.

The system (30) of simultaneous equations is highly nonlinear, and so must be solved by numerical methods, e.g. hill-climbing methods, which depend for convergence on finding or guessing a reasonable initial approximation to the final decoder parameter values, so that a preliminary seach of plausible values must be undertaken.

By such numerical methods, which are quite time-consuming in all but the the very simplest cases, we have found that for loudspeaker layouts figs. 1b to 1e with the standard reference values of equations (28), the preservation decoder parameters are:

$$\phi = 50.36^o \quad (9p)$$

for the 3×2 decoder of equations (9),

$$\phi_3 = 10.57^o \text{ and } \phi_D = 28.64^o \quad (10p)$$

for the 4×3 decoder of equations (10), and

$$(a,b,c) = (0.6164, 0.6558, 0.4359),\ \phi_4 = 51.64^o \text{ and } \phi_5 = 9.64^o \quad (11p)$$

approximately for the 5×4 decoder of equations (11).

Tables 1 to 4 show the computed values of the original values r_V, θ_V, r_E and θ_E and the matrix-decoded values r_V', θ_V', r_E' and θ_E' of the localisation parameters of various preservation decoders for the stereo test signal gains, respectively for the 3×2, 4×3, and 5×4 presevation decoders and also the 5×3 preservation decoder obtained by cascading the 4×3 with the 5×4 preservation decoder, via the speaker layouts of figs. 1b to 1e with the standard reference angles of equ. (28).

It will be seen from these tables that the values of θ_V and θ_E are preserved to within about 1^o, and that the values of r_V are not altered greatly. The values of r_E are often diminished somewhat as expected in advance, but not by a great amount, and in a few cases are actually increased, e.g. for central images in the 3×2 and 5×4 decoder cases - a useful if marginal improvement.

Investigations of the before and after localisation parameters of these decoders with signal gains according to various 3- and 4-channel panpot laws have indicated that these decoders maintain the values of localisation parameters remarkably well for a variety of ways of positioning sounds. However, we do not give details of these investigations here, since a detailed discussion of panpot laws is outside the scope of this paper. Differences in angular localisation before and after decoding seem invariably to be under 2^o for reasonable stereo signals with these decoders.

It will be noted that preservation decoders give an overall image width rather less than the full angular width of the decoded speaker layout - something that is inevitable given that any cross-talk among loudspeakers narrows the energy vector reproduction. However, this loss is minimal, being about 78.6% for the 3×2 decoder, 90.2% for the 4×3 decoder, and 94.4% for the 5×4 decoder.

Appendix A lists the matrix equations of these preservation decoders in direct terms, i.e. in terms of the left/right form of the loudspeaker feed signals.

For loudspeaker layouts with other angles θ_p than the standard reference values θ_p given in equs. (28), the results are somewhat similar. Tables 5 and 6 give the computed values of the preservation decoder parameters of 3×2 and 4×3 energy-preserving decoders for final reproduction speaker layouts with various indicated values of θ_3, θ_4 and θ_5. It will be seen that the values of these preservation decoder parameters does not vary enormously with the precise angular disposition of the speaker layout. However, detailed computations of the localisation parameters indicate that the best-behaved values of r_E are obtained if θ_5 is near $\frac{1}{2}\theta_4$. For θ_5 much smaller, there is a greater loss of total stage width and a reduction of r_E for edge-of--stage sounds, and for θ_5 much larger, there is a marked loss of r_E for centre-stage sounds. Space precludes full details of these results here, but nevertheless it is expected that preservation decoders should prove useful for most reasonable multispeaker stereo layouts.

8. IMPROVEMENT DECODERS

There are two problems with the use of preservation decoders, as we have already seen in the 3×2 decoder case. First, they preserve the inherent defects as well as the desired virtues of a stereo sound via n speakers, whereas the use of a greater number of speakers should allow some improvement of at least some of the localisation parameters, notably r_E, over parts of the stereo stage for which they are poor, perhaps at the expense of slightly degrading r_E for those stereo positions at which it is very close to one. Secondly, preservation decoders were designed assuming a localisation theory that becomes inaccurate at high frequencies, particularly above 5 kHz.

Energy-preserving matrix decoders that "improve" the localisation parameters, particularly r_E, at low and middle frequencies, and that have modified behaviour above about 5 kHz to take into account the modified psychoacoustics of localisation at these frequencies, are termed improvement decoders. The design of improvement decoders is more of an art than the design of preservation decoders, since the requirements imposed on improvement decoders is a somewhat subjective trade-off among various psychoacoustic requirements.

However, it is found that, since the inherent defects of well-made 3- and 4-speaker stereo sound are markedly smaller than for 2-speaker stereo sound, the room for improvement in 3- and 4-speaker material is smaller than in the 2-speaker case. Thus the decoder parameters for improvement decoders with 3- or 4-speaker inputs are found to be close, within a few degrees of angle, to the preservation decoder values, and in many cases, there is little advantage in departing from the use of a preservation decoder.

In general, a composite improvement decoder can be formed by following an improvement decoder with either an improvement or preservation decoder. Thus, for example, a 4×2 improvement decoder can be formed by cascading the 3×2 improvement decoder with $\phi = 35^o$ below 5 kHz and $\phi = 55^o$ above 5 kHz with the 4×3 preservation decoder given above with, say $\phi_3 = 10.57^o$ and $\phi_D = 28.64^o$, giving a 4×2 decoder as in equations (13) and figure 5 for which $\phi_{42} = \phi - \phi_3 = 25^o$ approximately below 5 kHz and $\phi_{42} = 45^o$ approximately above 5 kHz and $\phi_D = 28.6^o$. The decoder of fig. 5 can be made frequency-dependent in the same way that fig. 16 is a frequency-dependent version of fig. 4.

Slight variations of these low and high frequency decoder parameters may be found to give slightly better results for the 4×2 improvement decoder, although variations of ϕ_D with frequency are generally found to be only around 2 or 3^o.

The optimum improvement decoder parameters may be dependent on the precise recording or panning method used to create the input n-speaker stereo, particularly if sounds are positioned without the benefit of carefully designed n-speaker panpots. However, with well-made material, it is expected that a single improvement decoder should always work well, and that for n greater than 2, this improvement decoder will be little different from the preservation decoder.

n × 2 improvement decoders can easily be designed as composite decoders having the performance of the earlier-described 3 × 2 frequency-dependent decoder followed by an n × 3 preservation or improvement decoder.

9. DELAY COMPENSATION

The above work has applied to stereo layouts where all speakers are equally distant from a listener at the ideal stereo seat. These decoders can be used with other stereo speaker layouts, such as that shown in fig. 1g, but the results will be less than ideal, although they will generally still be quite acceptable to noncritical listeners. Generally , such layouts will still tend to work quite well for very noncentral listeners, but give degraded results at or near the traditional stereo seat.

Figure 17 indicates a method of overcoming this problem with layouts not equidistant from a central listener. Essentially, one provides a delay line in each of the loudspeaker feed channels feeding speakers that are closest to the central listener, so that the path-length differences in the acoustic field are compensated. Thus, if a central speaker is 0.5 metres closer than the outer speakers, it is fed via a 1.47 ms delay, since sound travels 0.5 m in 1.47 ms for a speed of sound in air of 340 m/s. The use of delay lines to ensure that sounds from all speakers arrive at the same time at the listener in the ideal stereo seat is termed "delay compensation", and is preferably done for a central location in the middle of the listening area.

Some gain compensation for the closer loudspeakers may also be used, but this should be done cautiously, since it affects the energy-preservation properties of the decoder, particularly for indirect or reflected sounds arriving from the speakers via the room boundaries. In general, providing delay compensation is more important and effective than gain compensation.

Results for listeners away from the ideal stereo seat are almost always improved by 'toeing in' the outer loudspeakers of a layout, e.g. as shown in fig. 1f. This has been realised since the 1950s [23], and is based on using the speakers' polar diagrams to help compensate for the Haas effect. The optimum degree of toeing in is very dependent on the actual characteristics of the speakers used.

10. CONCLUSIONS

This paper has presented detailed results on the optimum psychoacoustic design of matrix decoders for reproducing n_1-speaker stereo via a greater number n_2 of loudspeakers. Although the mathematics, based on orthogonal matrix theory and a velocity/sound intensity theory of sound localisation, leads to horrendously complicated systems of nonlinear equations, these can be solved numerically using a computer, and the optimum matrix decoders thus derived are relatively insensitive

to the precise angular dispositions of loudspeakers within the speaker layout.

This insensitivity is fortunate, since it means that a single $n_2 \times n_1$ matrix decoder can in practice be used with a variety of speaker layout dispositions, and adjustments are necessary only for the most critical applications.

This paper has, in particular, provided a detailed optimisation of a matrix decoder for reproducing 2-channel stereo via 3 loudspeakers. This decoder is frequency-dependent, and gives far from perfect stereo imaging, but it does give excellent general results at most listening positions with a useful improvement in central image stability, and gives most listeners the illusion of a continuous sound stage between the two speakers. It has been found preferable both to traditional 2-speaker stereo and to Bell/Klipsch bridged centre speaker stereo. It is suitable for a wide variety of applications, including high quality music reproduction, TV stereo reproduction, and other application such as "ghetto blaster" portable and in-car uses.

The other more general $n_2 \times n_1$ decoders, tabulated in Appendix A for particular speaker layouts, provide a detailed solution to the problem of feeding stereo source material to a larger number of speakers for all numbers of speakers up to five, so that, for example, 3-channel stereo films can be reproduced via 4 or 5 front-stage speakers. This may be particularly useful in connection with HDTV stereo systems, in that it makes the choice of number of transmitted stereo channels less critical.

The psychoacoustic theory used allows a detailed theoretical investigation of the stereo imaging properties of these decoders, in which their virtues and defects can be examined in detail without a very high expenditure on experimental psychoacoustic testing. Such testing is only required for the final 'fine tuning' stages in the design.

The work in this paper has many other applications, notably to the optimal design of a matrix transmission system for recording and transmitting multichannel stereo, and to the optimisation of 3- and 4-channel panpots, but the details of this will be published elsewhere.

It is believed that this paper provides a systematic basis for handling multispeaker stereo signals and for getting the best out of them via the available or desired reproduction layout. By using theoretical psychoacoustics in a systematic way, it allows better results to be obtained than prior ad-hoc proposals for multispeaker stereo systems, such as have been used previously in cinema or HDTV applications, and is particularly optimised towards domestic-scale applications.

REFERENCES

[1] J.C. Steinberg & W.B. Snow, "Auditory Perspective - Physical Factors", Electrical Engineering, vol. 53 No. 1 pp. 12-15 (1934 Jan.) reprinted in ref.[24]
[2] P.W. Klipsch, "Stereophonic Sound With Two Tracks, Three Channels By Means of a Phantom Circuit (2PH3)", J. Audio Eng. Soc., vol. 6 no. 2, pp. 118-123 (1958 April). Reprinted in ref. [24]
[3] P.W. Klipsch, "Stereo Geometry Tests", IRE Trans. Audio, vol. AU-10 No. 6, pp. 174-176 (1962 Nov./Dec). Reprinted in ref. [24]
[4] P.W. Klipsch, "Signal Mutuality in Stereo Systems", IRE Trans. Audio, vol. AU-8 no. 5 pp. 168-173 (1960 Sept./Oct.). Reprinted in [24]
[5] D.J. Meares, "Sound Systems for High Definition TV", Proc. Inst. Acoustics, Vol. 11 Part 7, pp. 47-62 (1989 Nov.)
[6] S. Komiyama, "Subjective Evaluation of Angular Displacement Between Picture and Sound Directions for HDTV Sound Systems". J. Audio Eng. Soc., vol. 37 no. 4, pp. 210-214 (1989 April)
[7] G. Theile, "On the Performance of Two-Channel and Multi-Channel Stereophony", Preprint 2887, Audio Engineering Society 80th Convention, Montreux, 1990 March 13-16
[8] M.A. Gerzon, "Surround Sound Psychoacoustics", Wireless World, vol. 80 pp. 483-486 (1974 Dec.)
[9] M.A. Gerzon, "The Rational Systematic Design of Surround Sound Recording and Reproduction Systems. Part I. General Theory of Directional Psychoacoustics and Applications," Appendix C of the Comments of National Research Development Corporation to the Federal Communication Commission in the matter of FM Quadraphonic Broadcasting, Docket 21310.
[10] M.A. Gerzon, "The Optimum Choice of Surround Sound Encoding Specification", Preprint 1199, Audio Engineering Society 56th Convention, Paris 1977 March
[11] D.H. Mershon & J.N. Bowers, "Absolute and Relative Cues for the Auditory Perception of Egocentric Distance", Perception, vol. 8, pp. 311-322 (1979)
[12] M.A. Gerzon, "Fixing It Outside the Mix", Studio Sound, vol. 32 no. 9, pp. 78-93 (1990 Sept.), Especially p. 86 column 1.
[13] M.A. Gerzon, "Ambisonics in Multichannel Broadcasting and Video", J. Audio Eng. Soc., vol. 33 no. 11, pp. 859-871 (1985 Nov.)
[14] H. Wallach, E.B. Newman & M.R. Rosenzweig, "The Precedence Effect in Sound Localization," Am. J. Psychol., vol. 52 pp.315-336 (1949)
[15] H. Haas, "The Influence of a Single Echo On the Audibility of Speech," J. Audio Eng. Soc., vol. 20 no. 2 pp.145-159 (1972 Mar.)
[16] Y. Makita, "On the Directional Localization of Sound in the Stereophonic Sound Field," E.B.U. Review, Part A no. 73, pp. 102 -108 (1962)
[17] D.H. Cooper & T. Shiga, "Discrete-Matrix Multichannel Stereö," J. Audio Eng. Soc., vol. 20, pp. 346-360 (1972 June)
[18] R. Orban, "Notes on Panpots," J. Audio Eng. Soc., vol. 19 no. 11 pp. 936- (1971 Nov.)
[19] H.D. Harwood, "Stereophonic Image Sharpness", Wireless World, vol. 74, pp. 207-211 (1968)
[20] J.S. Bower, "The Subjective Effects of Interchannel Phase-Shifts on

the Sterephonic Image Localisation of Narrowband Audio Signals," BBC Research Department Report BBC RD 1975/28 (1975 Sept.)
[21] H.A.M. Clark, G.F. Dutton & P.B. Vanderlyn, "The 'Stereosonic' Recording and Reproducing System," J. Audio Eng. Soc., vol. 6 no. 2 pp. 102-117 (1958 April). Reprinted in ref. [24]
[22] M.A. Gerzon, "Three Channels. The Future of Stereo?", Studio Sound, vol. 32 no. 6, pp. 112- 125 (1990 June)
[23] W.B. Snow, "Basic Principles of Stereophonic Sound,", J. SMPTE vol. 61 pp. 567-589 (1953 Nov.). Reprinted in ref. [24]
[24] ed. J. Eargle, Stereophonic Techniques, Audio Engineering Society, New York, 1986

PATENT NOTE

Some of the work disclosed in this paper is the subject of patent applications by the author.

ACKNOWLEDGEMENTS

I would like to acknowledge the encouragement given to me in the course of this work by Peter Craven, Roger Furness and John S. Wright.

APPENDIX A. EXPLICIT EQUATIONS FOR PRESERVATION DECODERS

The preservation decoder equations for the speaker layouts of figs. 1b to 1e with angles θ_p as in equs. (28), computed from equations (9a), (9b), (9p), (10a), (10b), (10p), (11a), (11b), (11p) and (12), along with the MS equations (1-3), are given here in terms of the speaker feed signals.

3×2 preservation decoder

$$\begin{pmatrix} L_3 \\ C_3 \\ R_3 \end{pmatrix} = \begin{pmatrix} 0.8850 & -0.1150 \\ 0.4511 & 0.4511 \\ -0.1150 & 0.8850 \end{pmatrix} \begin{pmatrix} L_2 \\ R_2 \end{pmatrix}$$

4×3 preservation decoder

$$\begin{pmatrix} L_4 \\ L_5 \\ R_5 \\ R_4 \end{pmatrix} = \begin{pmatrix} 0.9303 & -0.1297 & 0.0527 \\ 0.3314 & 0.6951 & -0.1479 \\ -0.1479 & 0.6951 & 0.3314 \\ 0.0527 & -0.1297 & 0.9303 \end{pmatrix} \begin{pmatrix} L_3 \\ C_3 \\ R_3 \end{pmatrix}$$

5×4 preservation decoder

$$\begin{pmatrix} L_6 \\ L_7 \\ C_5 \\ R_7 \\ R_6 \end{pmatrix} = \begin{pmatrix} 0.9535 & -0.1084 & 0.0590 & -0.0324 \\ 0.2533 & 0.7870 & -0.1989 & 0.0859 \\ -0.1349 & 0.5708 & 0.5708 & -0.1349 \\ 0.0859 & -0.1989 & 0.7870 & 0.2533 \\ -0.0324 & 0.0590 & -0.1084 & 0.9535 \end{pmatrix} \begin{pmatrix} L_4 \\ L_5 \\ R_5 \\ R_4 \end{pmatrix}$$

5×3 composite preservation decoder

$$\begin{pmatrix} L_6 \\ L_7 \\ C_5 \\ R_7 \\ R_6 \end{pmatrix} = \begin{pmatrix} 0.8407 & -0.1538 & 0.0557 \\ 0.5304 & 0.3648 & -0.0891 \\ -0.0279 & 0.8285 & -0.0279 \\ -0.0891 & 0.3648 & 0.5304 \\ 0.0557 & -0.1538 & 0.8407 \end{pmatrix} \begin{pmatrix} L_3 \\ C_3 \\ R_3 \end{pmatrix}$$

TABLES

gains		2-speaker parameters				3-speaker parameters			
L_2	R_2	r_V	θ_V	r_E	θ_E	r_V'	θ_V'	r_E'	θ_E'
1	0	1.0000	35.00	1.0000	35.00	1.0000	35.38	0.9404	35.38
1	1	0.8192	0.00	0.8192	0.00	0.8153	0.00	0.8263	0.00

Table 1 Gains of 2-speaker stereo test signals and their localisation parameters direct via 2 speakers with $\theta_2 = 35^o$ and via 3×2 preservation decoder via 3 speakers with $\theta_3 = 45^o$.

gains			3-speaker parameters				4-speaker parameters			
L_3	C_3	R_3	r_V	θ_V	r_E	θ_E	r_V'	θ_V'	r_E'	θ_E'
1	0	0	1.0000	45.00	1.0000	45.00	0.9805	45.08	0.9690	45.08
0	1	0	1.0000	0.00	1.0000	0.00	1.0303	0.00	0.9474	0.00
1	1	0	0.9239	22.50	0.9239	22.50	0.9282	22.32	0.9254	22.32
1	0	1	0.7071	0.00	0.7071	0.00	0.6924	0.00	0.6534	0.00

Table 2 Gains of 3-speaker stereo test signals and their localisation parameters direct via 3 speakers with $\theta_3 = 45^o$ and via a 4×3 preservation decoder via 4 speakers with $\theta_4 = 50^o$ and $\theta_5 = \frac{1}{3}\theta_4$.

gains				4-speaker parameters				5-speaker parameters			
L_4	L_5	R_5	R_4	r_V	θ_V	r_E	θ_E	r_V'	θ_V'	r_E'	θ_E'
1	0	0	0	1.0000	50.00	1.0000	50.00	0.9996	50.95	0.9793	50.95
0	1	0	0	1.0000	16.67	1.0000	16.67	1.0009	16.32	0.9606	16.32
1	1	0	0	0.9580	33.33	0.9580	33.33	0.9549	33.75	0.9546	33.83
0	1	1	0	0.9580	0.00	0.9580	0.00	0.9606	0.00	0.9613	0.00
1	0	1	0	0.8355	16.67	0.8355	16.67	0.8328	17.56	0.7790	17.51
1	0	0	1	0.6428	0.00	0.6428	0.00	0.6298	0.00	0.6377	0.00

Table 3 Gains of 4-speaker stereo test signals and their localisation parameters direct via 4 speakers with $\theta_4 = 50^o$ and $\theta_5 = \frac{1}{3}\theta_4$ and via a 5×4 preservation decoder via 5 speakers with $\theta_6 = 54^o$ and $\theta_7 = \frac{1}{2}\theta_6$.

gains			3-speaker parameters				5-speaker parameters			
L_3	C_3	R_3	r_V	θ_V	r_E	θ_E	r_V'	θ_V'	r_E'	θ_E'
1	0	0	1.0000	45.00	1.0000	45.00	0.9764	45.76	0.9683	45.74
0	1	0	1.0000	0.00	1.0000	0.00	1.0378	0.00	0.9515	0.00
1	1	0	0.9239	22.50	0.9239	22.50	0.9272	22.70	0.9226	22.41
1	0	1	0.7071	0.00	0.7071	0.00	0.6812	0.00	0.6475	0.00

Table 4 Gains of 3-speaker stereo test signals and their localisation parameters direct via 3 speakers with $\theta_3 = 45^o$ and via a 5×3 composite preservation decoder via 5 speakers with $\theta_6 = 54^o$ and $\theta_7 = \frac{1}{2}\theta_6$.

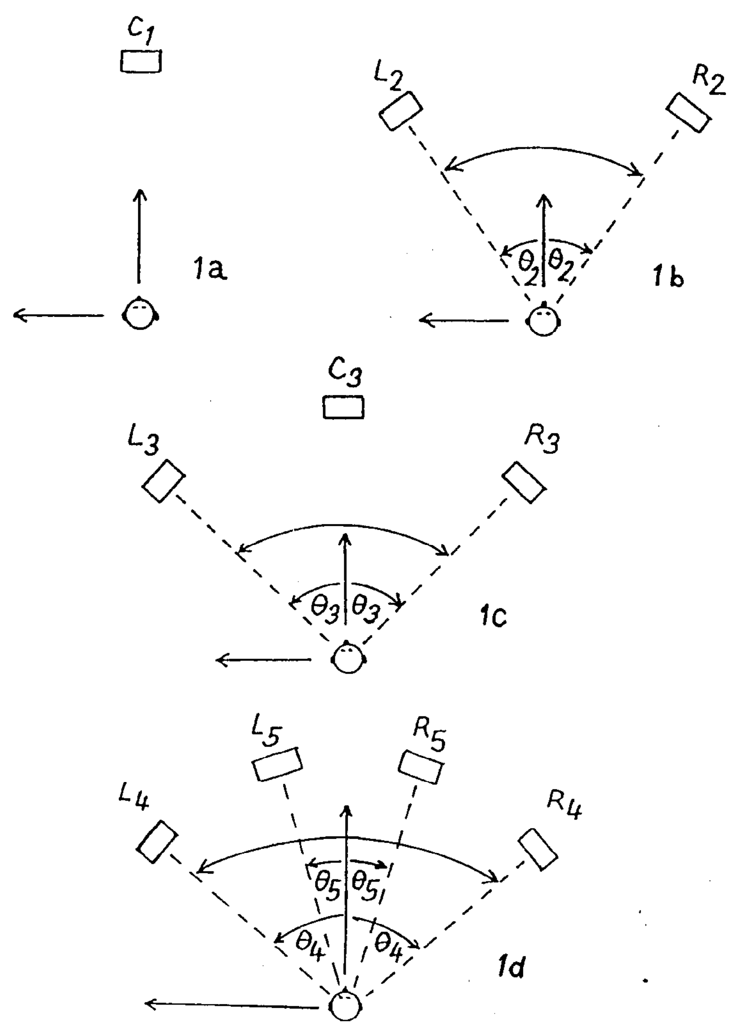
θ_3	0.00	15.00	30.00	45.00	60.00
ϕ	54.74	54.27	52.84	50.36	46.69

Table 5 Value of 3×2 preservation decoder parameter ϕ for various values of the 3-speaker layout half-angle θ_3.

θ_4	45.00	45.00	50.00	50.00	60.00	60.00	60.00	60.00	60.00	75.00	75.00
θ_5	9.00	15.00	10.00	16.67	12.00	15.00	20.00	24.00	30.00	15.00	25.00
ϕ_3	9.07	10.40	9.08	10.57	9.16	9.89	10.98	11.73	12.65	9.49	11.76
ϕ_D	33.39	28.32	32.72	28.64	31.64	30.75	29.42	28.37	26.70	30.92	31.01

Table 6 Value of 4×3 preservation decoder parameters ϕ_3 and ϕ_D for various values of the 4-speaker layout angles θ_4 and θ_5. All angles are in degrees.

FIGURES



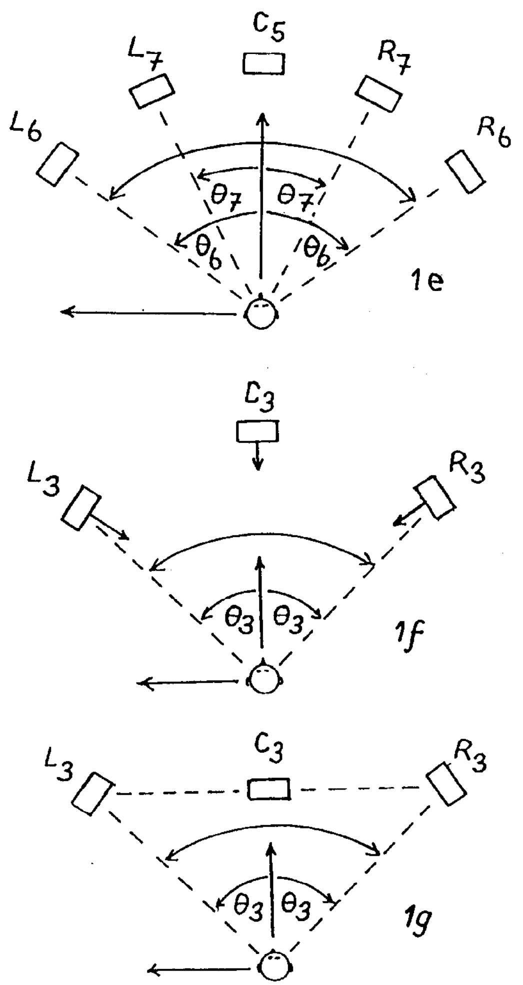


Figure 1. Various loudspeaker arrangements for frontal-stage stereo. 1a to 1f have all speakers equally distant from the central listener shown, 1a to 1e has all speakers face that listener, and 1f and 1g have toed-in outer speakers.

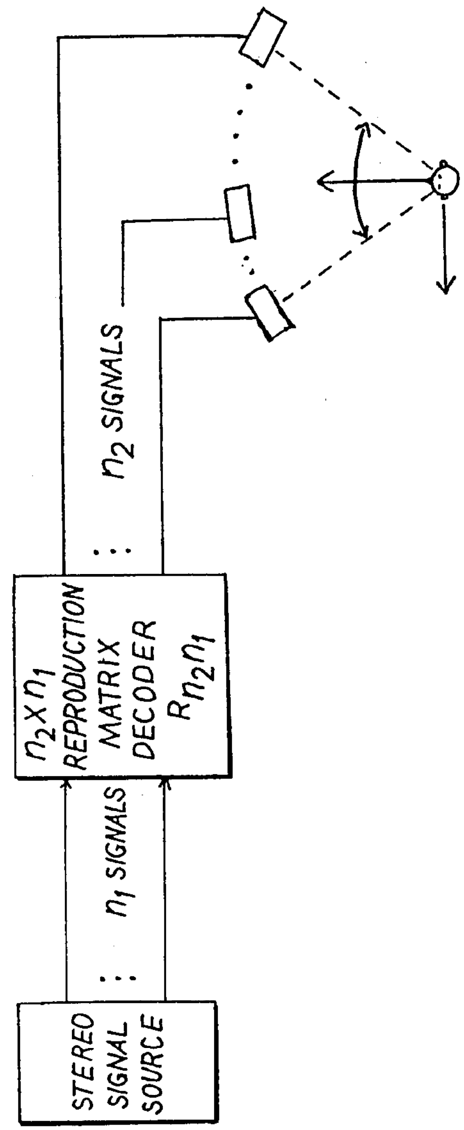


Figure 2. Schematic of n_2-speaker stereo reproduction via a decoding matrix from an n_1-speaker stereo source.

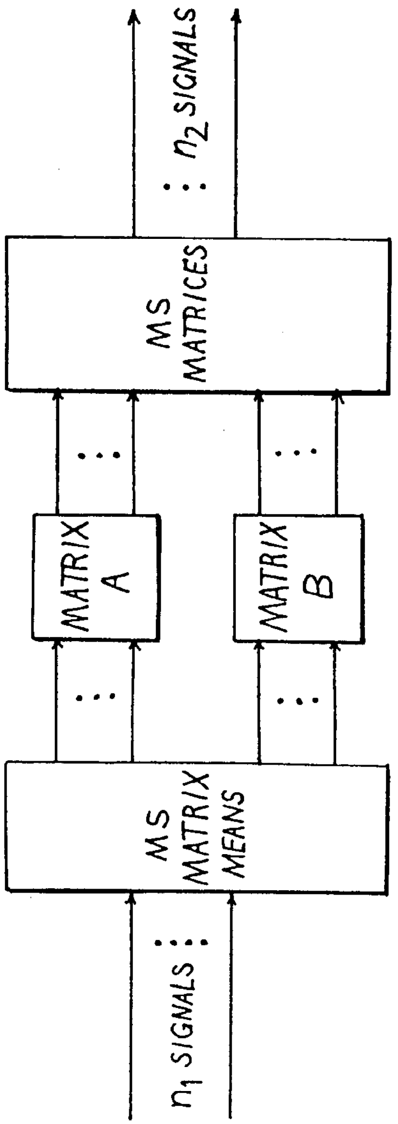


Figure 3. $n_2 \times n_1$ matrix reproduction decoder using input and output sum and difference processing, matrix A handling sum signals and matrix B handling difference signals.

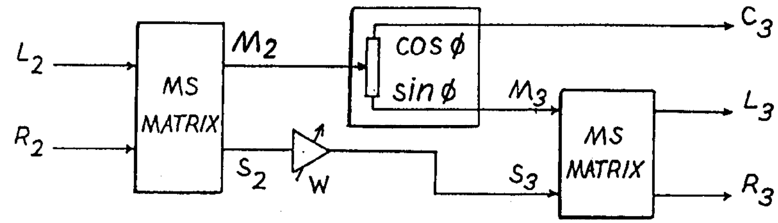


Figure 4. Energy-preserving 3×2 matrix decoder network with additional width control gain w.

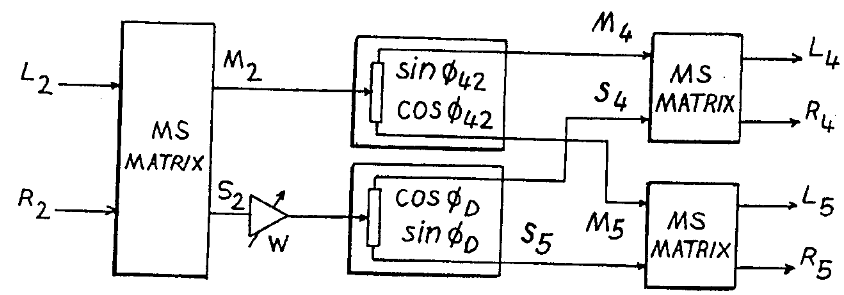


Figure 5. Energy-preserving 4×2 matrix decoder network with additional width control gain w.

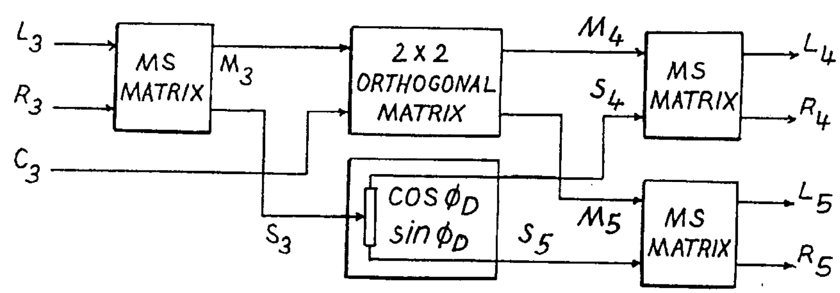


Figure 6. Energy-preserving 4×3 matrix decoder network.

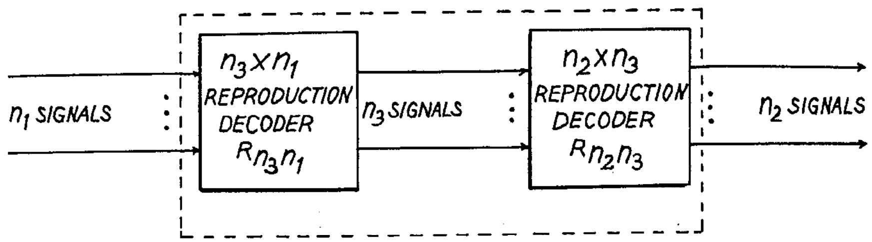


Figure 7. Composite matrix reproduction decoder formed from the series connection of two component matrix reproduction decoders.

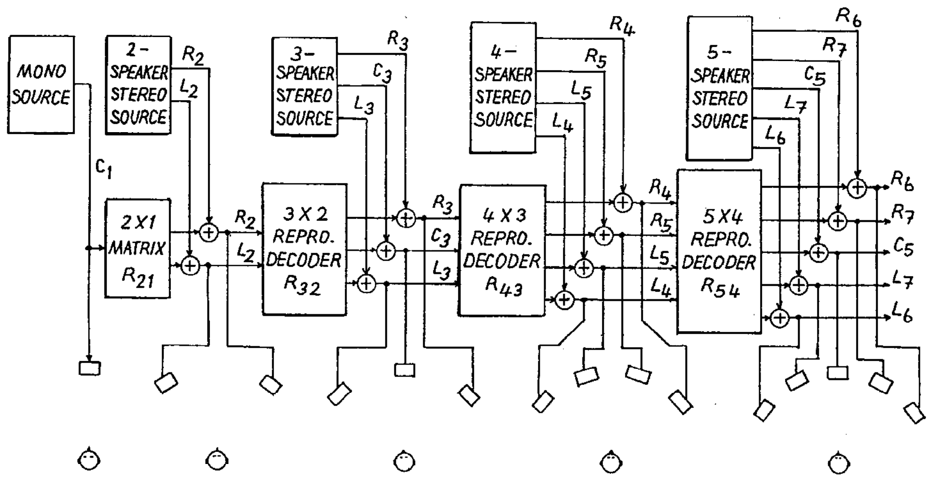


Figure 8. Schematic showing how stereo signals for any number of loudspeakers can be mixed with those intended for and reproduced via any larger number of loudspeakers using component (n+1) x n matrix reproduction decoders.

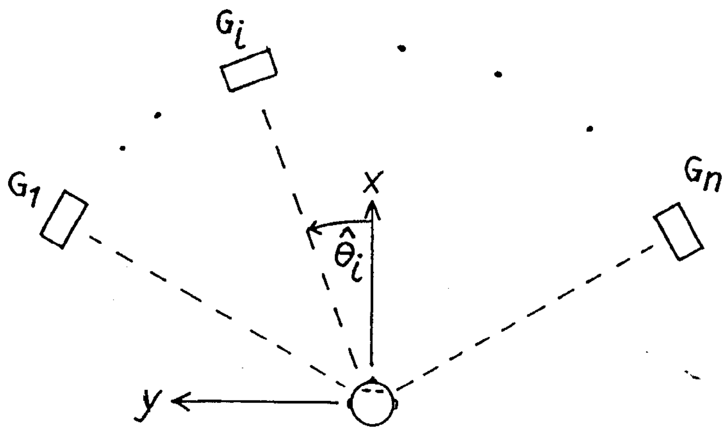


Figure 9. Showing rectangular axes x, y and polar angles $\hat{\theta}_i$ of loudspeakers fed with indicated gains G_i used in psychoacoustic localisation theories.

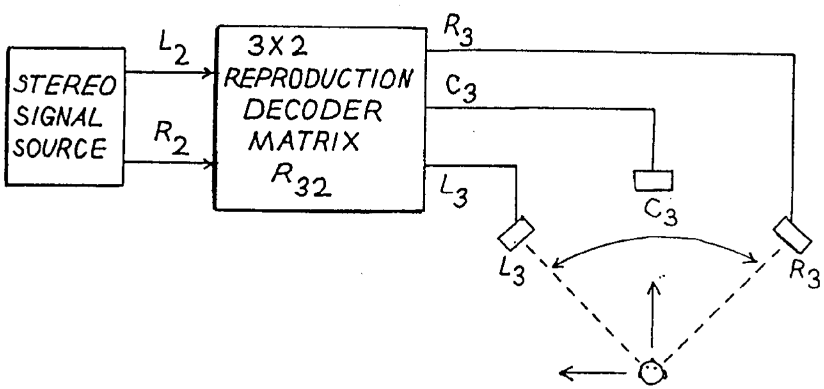


Figure 10. The reproduction of 2-channel stereo through three loudspeakers via a matrix decoder.

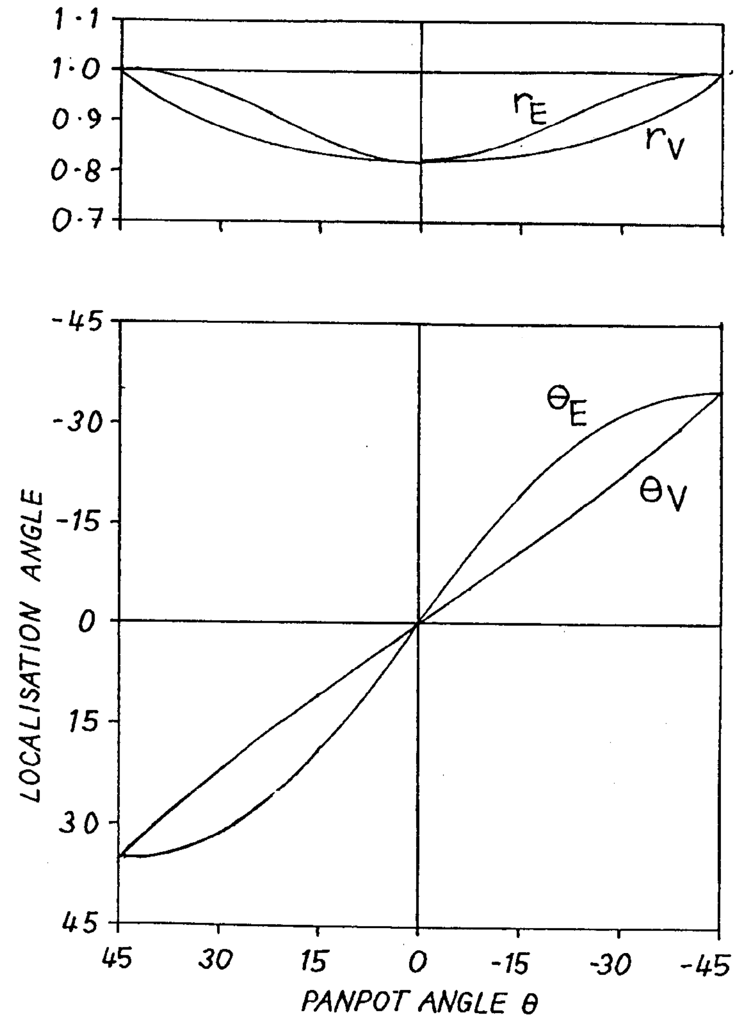


Figure 11. Psychoacoustic localisation parameters for 2-speaker stereo reproduction with $\theta_2 = 35°$.

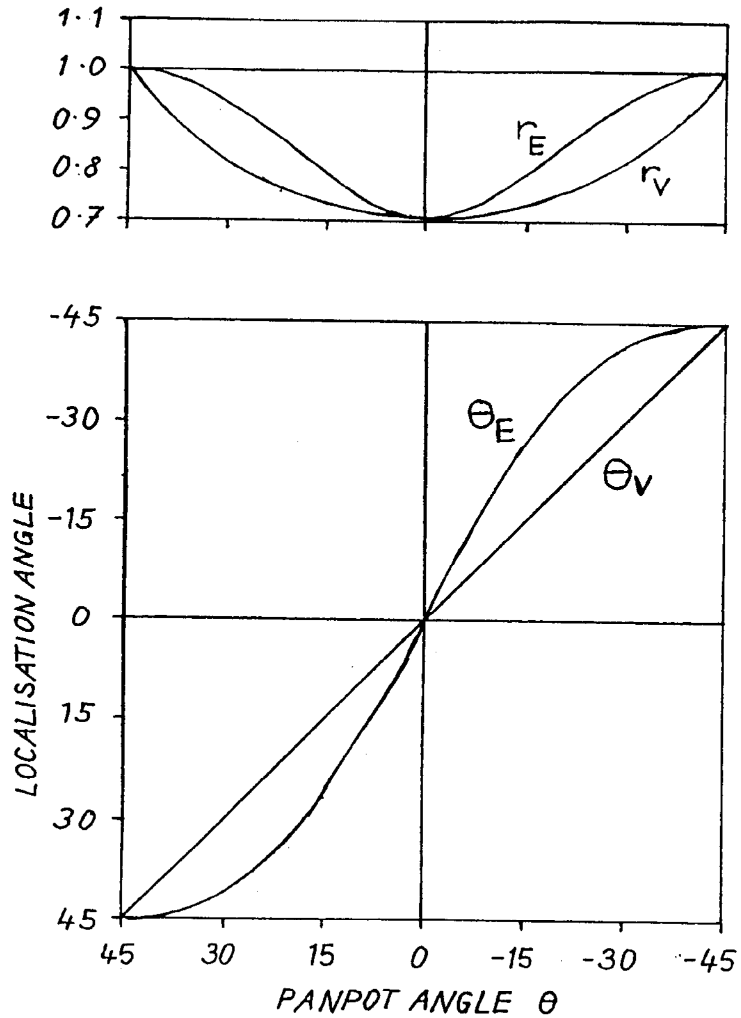


Figure 12. Psychoacoustic localisation parameters for 2-speaker stereo reproduction with $\theta_2 = 45^o$, i.e. for 3-speaker stereo reproduction with $\theta_3 = 45^o$ via the 3×2 energy-preserving matrix decoder with $\phi = 90^o$.

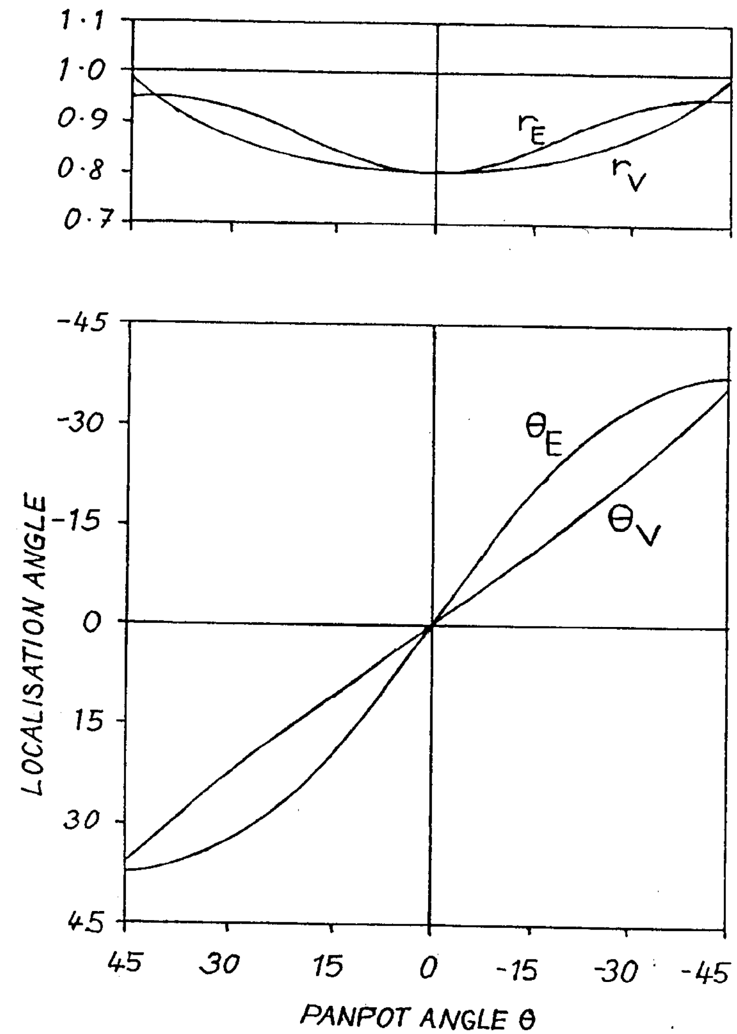


Figure 13. Psychoacoustic localisation parameters for 3-speaker stereo reproduction with $\theta_3 = 45^o$ via a 3×2 energy-preserving matrix decoder with $\phi = 54.74^o$.

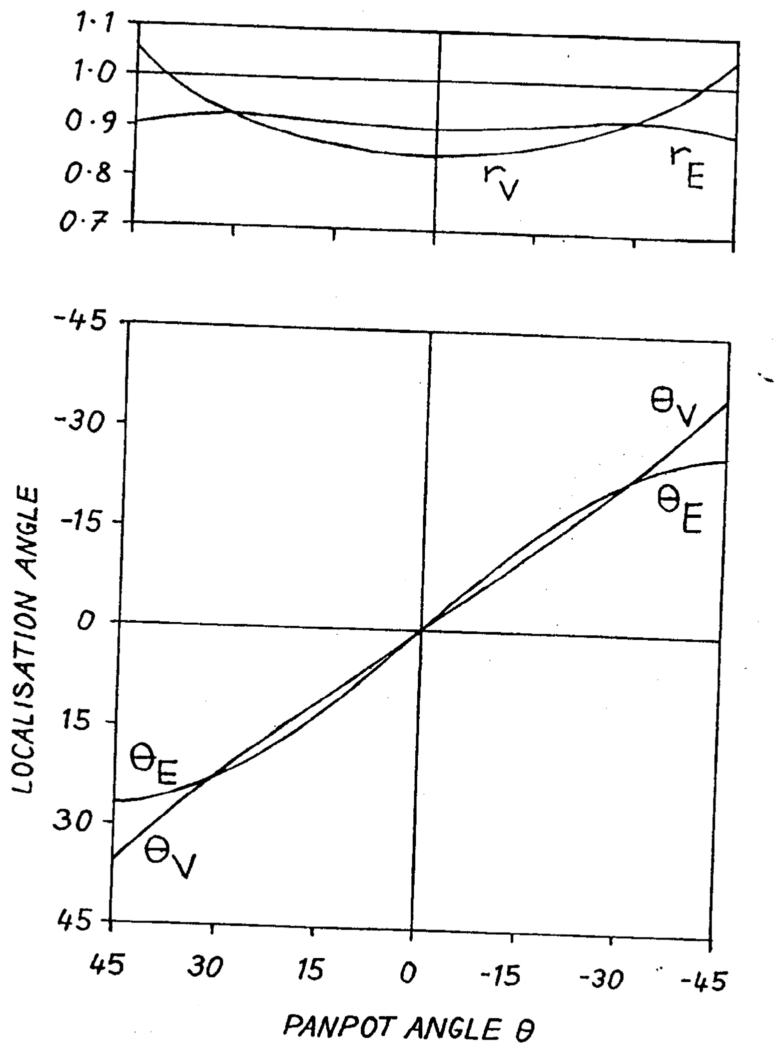


Figure 14. Psychoacoustic localisation parameters for 3-speaker stereo reproduction with $\theta_3 = 45^o$ via a 3×2 energy-preserving matrix decoder with $\phi = 35.26^o$.

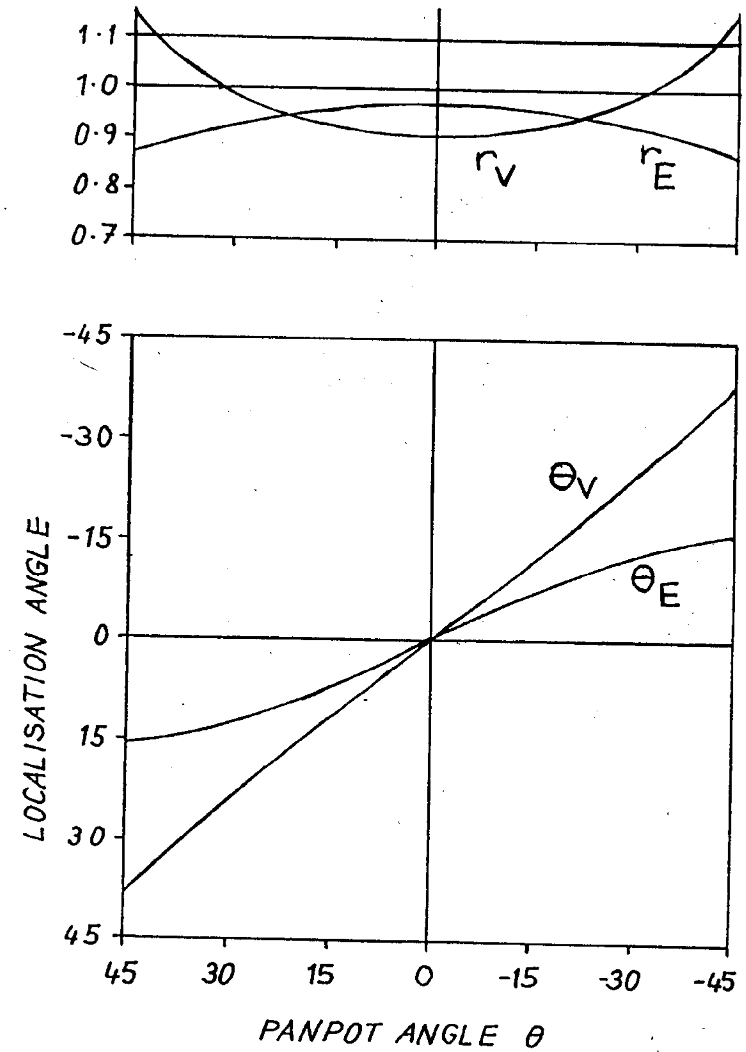


Figure 15. Psychoacoustic localisation parameters for 3-speaker stereo reproduction with $\theta_3 = 45^o$ via a 3×2 energy-preserving matrix decoder with $\phi = 19.47^o$.

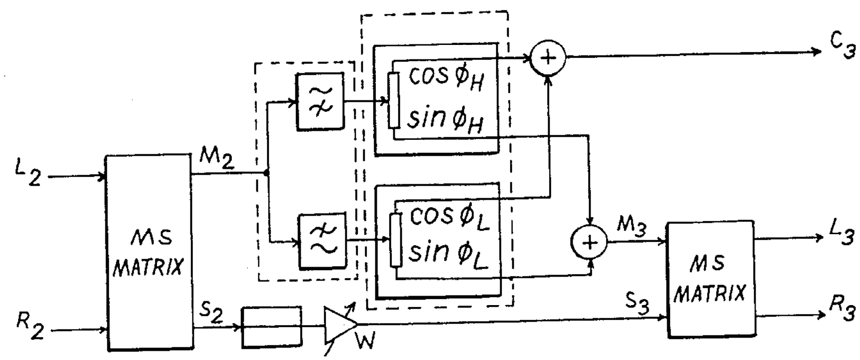


Figure 16. Frequency-dependent version of the 3 × 2 matrix decoder of fig. 4, using a bandsplitting filter network preceeding the sine/cosine gains and a possible all-pass phase compensation in the difference channel.

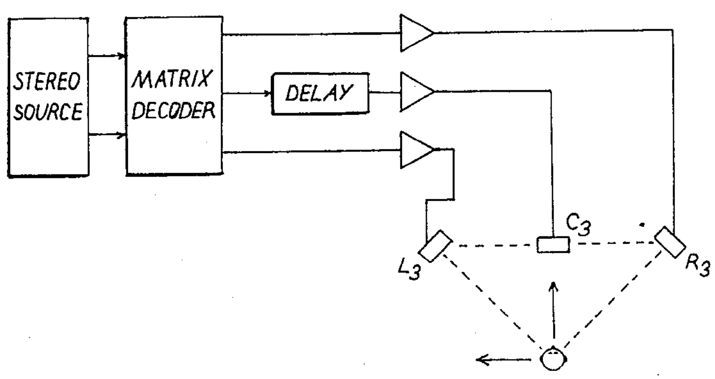


Figure 17. Use of delay compensation to compensate for differing loudspeaker distances.