Optimum Reproduction Matrices for Multispeaker Stereo*

MICHAEL A. GERZON**, *AES Fellow*

Oxford OX2 6DJ, UK

Psychoacoustically optimum matrices are described, which can be applied 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 localization. Optimal decoder designs for reproducing two-channel stereo via three loudspeakers are described in detail, using frequency-dependent matrix coefficients, along with uses in music and television applications where a stable stereo image over a large listening area is required.

0 INTRODUCTION

Since the earliest days of stereo it has been realized that reproduction of a stereo signal intended for a given number of loudspeakers over a greater number of loudspeakers might give improved results [1]. In particular, the reproduction of two-channel stereo via three loudspeakers has been proposed and used by Bell Telephone Laboratories [2] in 1933 and by Klipsch in the 1950s [3]–[5] among others, where the center loudspeaker was fed the average of the left and right channels with an additional gain factor. This "bridged center loudspeaker" 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-loudspeaker 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-loudspeaker 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 analyzing auditory localization and extensive listening tests to an adjustable decoding matrix, an optimized three-loudspeaker decoder for two-channel stereo has been found which not only improves central image stability over a large listening area, but which also 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 behavior, the results give very low listening fatigue—certainly lower than ordinary two-loudspeaker stereo—as well as convincing stereo images for listeners across a large listening area, with markedly better results than the older bridged center loudspeaker method.

The discovery and optimization of the 3×2 decoder case led to a general theory and the discovery of optimal decoders for converting n_1-loudspeaker stereo signals for reproduction via a larger number n_2 of loudspeakers

* Manuscript received 1991 February 23 and presented at the 91st Convention of the Audio Engineering Society, New York, 1991 October 4–8.

** Now Technical Consultant to Trifield Productions Ltd., London N7 9AH, UK.

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 $n_1 = 2, 3, 4$ and for $n_2 = 3, 4, 5$, so that the optimal method of, say, reproducing four-loudspeaker stereo over five loudspeakers can be determined.

In connection with high-definition television (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 are the optimum numbers, each with an associated transmission channel (for example, see [6]–[8]). While the tradeoff between localization accuracy and a convenient number of loudspeakers is ultimately a subjective choice, the use of a 4×3 matrix decoder permits four-loudspeaker reproduction from a three-loudspeaker stereo transmission, and a 5×3 or 5×4 matrix decoder permits five-loudspeaker stereo reproduction from a three- or four-loudspeaker 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 localization 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 five loudspeakers, although the methods used are applicable also to six or more loudspeakers.

The psychoacoustic requirements on matrix decoders for multiloudspeaker 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. Second, we require that the localization 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 localization performance, we use two main classes of sound localization 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 [9]–[11], 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 parameterization of possible decoder matrices having a geometric character. The localization theory is used to find the optimum values of these parameters. Especially in the cases $n_1 = 3$ or 4, the values of these parameters are 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 two-channel stereo source has little leeway if optimum results are required.

1 STEREO LOUDSPEAKER LAYOUTS AND MS MATRICES

Fig. 1(a)–(e) 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 one-loudspeaker (mono) layout [Fig. 1(a)], L_2 and R_2 for a two-loudspeaker stereo layout [Fig. 1(b)], L_3, C_3, and R_3 for a three-loudspeaker layout [Fig. 1(c), 1(f), or 1(g)], L_4, L_5, and R_5, and R_4 for a four-loudspeaker layout [Fig. 1(d)], and L_6, L_7, C_5, R_7, and R_6 for a five-loudspeaker stereo layout [Fig. 1(e)].

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 loudspeakers L_p are assumed to lie at an angle θ_p anticlockwise from due front, and the mirror-image loudspeakers R_p are assumed to lie at an angle θ_p clockwise from due front at the listener. While a wide variety of n-loudspeaker stereo layouts is possible, in the main we shall deal with layouts such as shown in Fig. 1(b)–(f) where all loudspeakers are at the same distance from an ideally positioned listener, and where the angles 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 Fig. 1(a)–(e) 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 Fig. 1(f) and (g), since this is found to enlarge the listening area. In Fig. 1(g), in addition, the loudspeakers lie along a straight line rather than a circle centered at the listener—a layout that generally gives less good results than the circular layout of Fig. 1(f), but which is often easier to accommodate in a listening environment.

For reasons of theoretical convenience, most of the analyses of psychoacoustic performance of matrix decoders are done for layouts in which all loudspeakers are equally distant from the ideal listener position, as shown in Fig. 1(a)–(e), 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

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

and difference signals to be of the form

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

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

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

A matrix circuit satisfying these equations will be termed an MS matrix. It will be seen that an MS matrix is its own inverse, that is, 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.

2 MATRIX DECODERS AND ENERGY PRESERVATION

Fig. 2 shows the general schematic of a matrix reproduction decoder accepting n_1 signals from a stereo source intended for n_1-loudspeaker stereo reproduction, and converted by the matrix reproduction decoder into n_2 signals intended for reproduction via an n_2-loudspeaker 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 behavior, it can be implemented in the form shown in Fig. 3, where the n_1 input loudspeaker-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 are 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 pre-

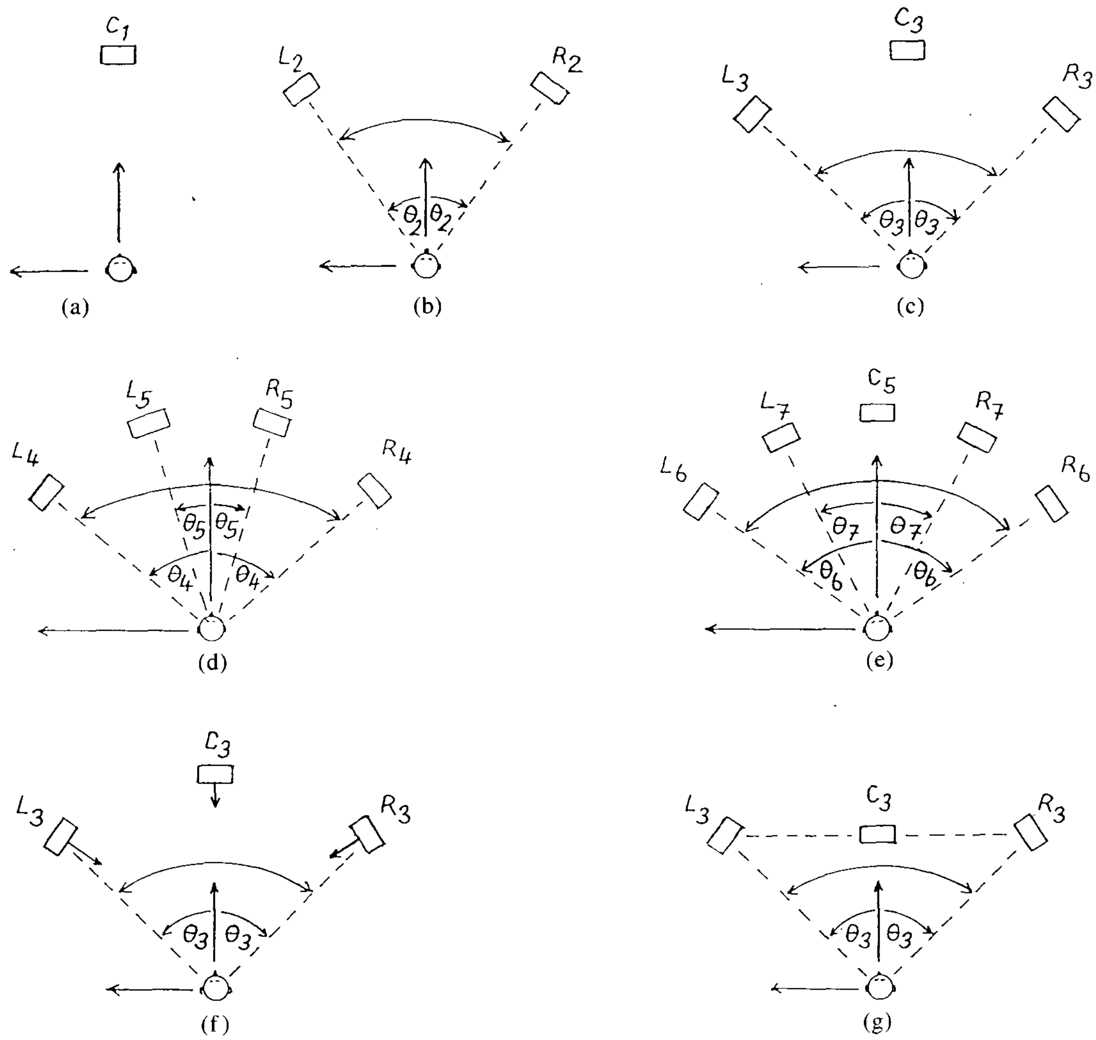

Fig. 1. Loudspeaker arrangements for frontal-stage stereo. (a)–(f) All loudspeakers are equally distant from central listener shown. (a)–(e) All loudspeakers face that listener. (f), (g) Toed in outer loudspeakers.

serve the total energy reproduced into the room. The obvious reason for this is that if the total stereo energy is preserved, 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 (for example, see Mershon and Bowers [12]). If a decoder is not energy preserving to a significant degree, it will significantly degrade the sense of sound-source distance in those recordings having it. (such as UHJ recordings using a SoundField microphone).

The second reason is a little more complicated to explain, but is related to the fact that there is no consensus as to the ideal type of stereo recording technique. The reason for this lack of consensus 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 concerning the aspects that 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 colorations. 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 (for example, see [13], especially p. 86, col. 1).

Thus in general it is found that energy-preserving decoders suffer from far less audible coloration 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, and it has to do with total stage width. It is generally found that most attempts at matrixing n_1-loudspeaker 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 center loudspeaker system—and Klipsch [3]–[5] 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 localization 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 it 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

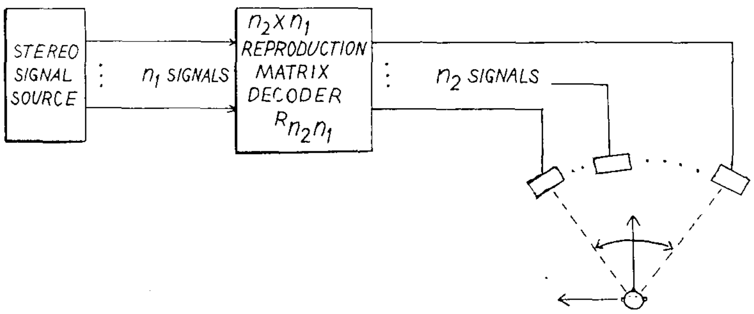

Fig. 2. Schematic of n_2-loudspeaker stereo reproduction via decoding matrix from n_1-loudspeaker stereo source.

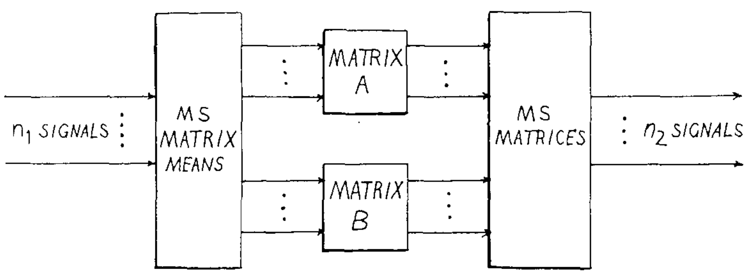

Fig. 3. $n_2 \times n_1$ matrix reproduction decoder using input and output sum-and-difference processing, matrix A handling sum signals and matrix B handling difference signals.

they have complex-valued coefficients (for example, 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 either are totally frequency independent or 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 analyze 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 Fig. 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 preceded 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

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

where ϕ is an angle parameter, although by replacing ϕ by $90° - \phi$ 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

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

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

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

has a matrix of the form

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

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

In practice, the parameterization of Eq. (8) is rarely the most convenient, and in general, an $n \times n$ matrix with real coefficients is orthogonal if and only if 1) all its columns are vectors of unit length, that is, the sum of squares of its entries is 1, and 2) any two columns have a zero inner product, that is, the sum of the products of the corresponding entries of the two columns is 0. This characterization 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 Fig. 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

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

and the matrix B equation is simply

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

where the angle parameter ϕ lies between 0° and 90°.

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

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

and the matrix B equation is of the form

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

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

The form of an energy-preserving 5×4 decoding matrix is more complicated, and we give a parameterization 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

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

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

$$\lambda = (a^2 + b^2)^{1/2}$$

$$\mu_1 = \frac{b}{\lambda}, \qquad \mu_2 = \frac{a}{\lambda} \tag{12}$$

$$\nu_1 = \frac{ac}{\lambda}, \qquad \nu_2 = \frac{bc}{\lambda} .$$

The matrix B equations have the form

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

where ϕ_5 is an angle parameter that is likely to be well within the range of $-45°$ to $+45°$.

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

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

and the matrix B decoding equation has the form

$$\begin{pmatrix} S_4 \\ S_5 \end{pmatrix} = \begin{pmatrix} \cos \phi_D \\ \sin \phi_D \end{pmatrix} [S_2] \tag{13b}$$

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

Figs. 4–6 show the forms of energy-preserving 3×2, 4×2, and 4×3 matrix decoders satisfying the preceding equations, where the decoders with two-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-preserving matrix decoders are implemented similarly as in Fig. 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 1, and the matrix B equations are a 2×1 matrix whose entries are positive and whose squares add up to 1. 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 forms of these matrix A and matrix B equations are given in Sec. 3.

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 two 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 $n = 4$ or more, and 4×4 rotation matrices have a quite complicated general form involving six free parameters. Fortunately the case with $n_2 > 6$ rarely needs to be considered in practice for front-stage stereo.

3 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-loudspeaker stereo input signals and giving outputs for reproduction via an n_2-loudspeaker 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 com-

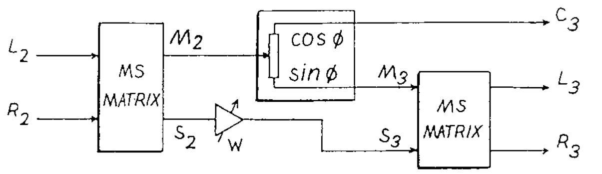

Fig. 4. Energy-preserving 3×2 matrix decoder network with additional width control gain w.

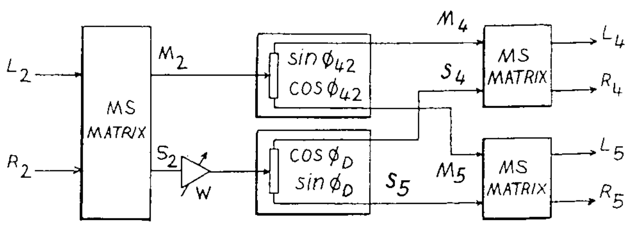

Fig. 5. Energy-preserving 4×2 matrix decoder network with additional width control gain w.

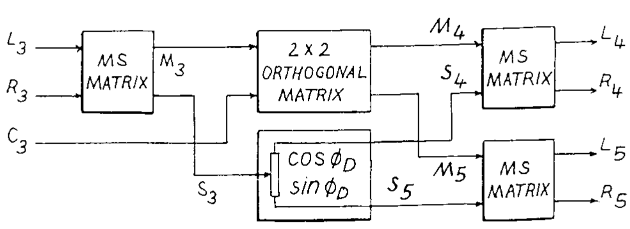

Fig. 6. Energy-preserving 4×3 matrix decoder network.

posite 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 the matrix $R_{n_3 n_1}$ and the $n_2 \times n_3$ matrix decoder is described by the matrix $R_{n_2 n_3}$, then the $n_2 \times n_1$ composite decoder of Fig. 7 is described by the product matrix

$$R_{n_2 n_1} = R_{n_2 n_3} R_{n_3 n_1} \ . \tag{14}$$

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

$$\phi_{42} = \phi - \phi_3 \ . \tag{15}$$

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 Eqs. (10) with the 5×4 matrix decoder of Eqs. (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 Eqs. (13) with a 5×4 matrix decoder as in Eqs. (11)—although a given decoder matrix can generally be expressed as a product of two component matrices in more than 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 as to substantially preserve a given aspect of the stereo localization 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. Fig. 8 shows how signals intended for n_1-loudspeaker 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 Fig. 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 Fig. 8.

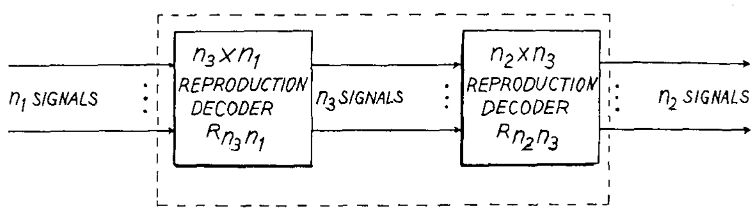

Fig. 7. Composite matrix reproduction decoder formed from series connection of two component matrix reproduction decoders.

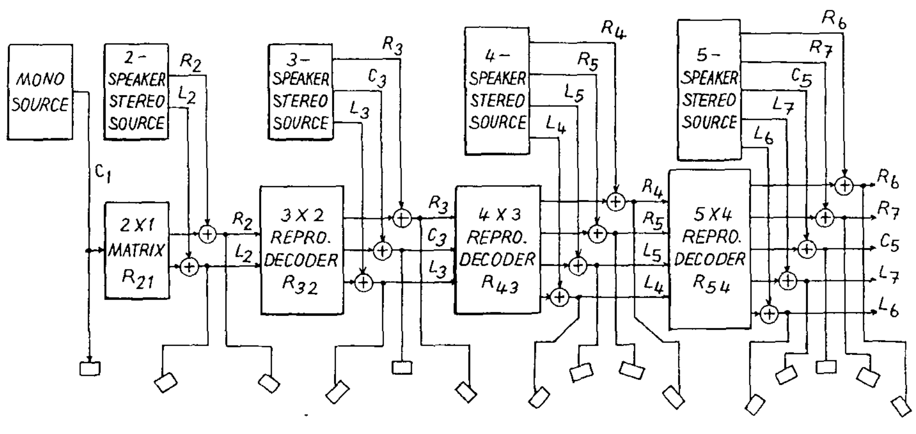

Fig. 8. Schematic showing how stereo signals for any number of loudspeakers can be mixed with those intended for and reproduced via any larger number of loudspeakers using component $(n + 1) \times n$ matrix reproduction decoders.

The possibility of this kind of hierarchy of matrix reproduction decoders for n-loudspeaker 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 $n_2 = n_1 + 1$, and implement the other cases as composite decoders.

4 DIRECTIONAL PSYCHOACOUSTICS

A summary of a theoretical framework for the perception of a 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 [9], is quite a good predictor for frequencies to at least 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 [10], where it 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 [11]. These previous descriptions were applied to Ambisonic surround sound and gave little explicit information about the applications of the theory to noncentral listeners, whereas in the present stereo application it is the intention to stabilize the *relative* position of stereo images with respect to the frontal stereo loudspeaker layout for listeners well away from the ideal stereo seat. (In Ambisonics [14] the intention is more to stabilize the *absolute* direction of images.)

Fig. 9 shows a stereo loudspeaker layout disposed on a sector of a circle centered at an ideal stereo seat position, with all loudspeakers 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 $i = 1, \ldots, n$, the ith loudspeaker is disposed in a direction measured by the angle $\hat{\theta}_i$ anticlockwise from the x axis, that is, from the x axis toward the y axis, so that $\hat{\theta}_i$ is positive for loudspeakers to the left of center and negative for loudspeakers to the right of center.

Suppose that a sound is fed to all of the n loudspeakers

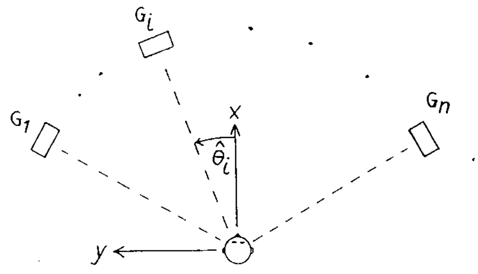

Fig. 9. Rectangular axes x, y and polar angles $\hat{\theta}_i$ of loudspeakers fed with indicated gains G_i used in psychoacoustic localization theories.

of Fig. 9 with gain G_i, which may be real or complex, to the ith 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 the sound intensity vector, and relate these to the resulting stereo illusion. This gives rise to two subtheories of localization. One, termed the velocity theory, relates the localization 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 localize 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 subtheory, which we term the energy vector theory, relates localization 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 (as might be the case with very noncentral 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 subtheories, 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 further worth mentioning that it is possible to demonstrate that the ears use different localization 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 [15], [16] is used, whereby earlier sound arrivals tend to influence localization 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 localization.

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 localization at these frequencies, but it is no longer a reasonably accurate predictor of the actual localization direction.

With all these qualifications in mind, we state the two subtheories in a usable mathematical and computational form. For a certain listener in Fig. 9, the total pressure gain at the listener is

$$P = \sum_{i=1}^{n} G_i \tag{16}$$

and the total energy gain at the listener is

$$E = \sum_{i=1}^{n} |G_i|^2 . \tag{17}$$

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

$$e = (e_x, e_y) \tag{18}$$

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

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

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

Localization according to the energy vector localization theory is determined by the vector ratio e/E, which has length $r_\mathrm{E} \geqslant 0$ and angular direction θ_E given by

$$\left. \begin{aligned} r_\mathrm{E} \cos \theta_\mathrm{E} &= \frac{e_x}{E} \\ r_\mathrm{E} \sin \theta_\mathrm{E} &= \frac{e_y}{E} . \end{aligned} \right. \tag{20}$$

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

Thus, between them, θ_E and r_E provide good predictors both of image localization for centrally placed listeners and of image stability for other listeners. Above 3.5 kHz, θ_E is not an accurate predictor of localization, 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

$$v = (v_x, v_y) \tag{21}$$

where

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

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

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

$$\frac{v}{P} = \left(\frac{v_x}{P}, \frac{v_y}{P}\right) \tag{23}$$

with the complication that in general this vector has complex-valued entries. However, low-frequency interaural phase localization theories predict [9], [10] that only the real part

$$\left(\mathrm{Re}\left(\frac{v_x}{P}\right), \mathrm{Re}\left(\frac{v_y}{P}\right)\right) \tag{24}$$

of this vector contributes to the apparent localization, although it is found that the imaginary component

$$\left(\mathrm{Im}\left(\frac{v_x}{P}\right), \mathrm{Im}\left(\frac{v_y}{P}\right)\right) \tag{25}$$

causes image broadening and an unpleasant subjective sensation termed phasiness, and also contributes to localization at middle frequencies around 600 Hz. In practice, if the length of the phasiness vector of Eq. (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, Eq. (24), of the ratio of velocity gain to pressure gain can be used to predict localization. This vector has length $r_\mathrm{V} \geqslant 0$ and angular direction θ_V given by the equations

$$\left. \begin{aligned} r_\mathrm{V} \cos \theta_\mathrm{V} &= \mathrm{Re}\left(\frac{v_x}{P}\right) \\ r_\mathrm{V} \sin \theta_\mathrm{V} &= \mathrm{Re}\left(\frac{v_y}{P}\right) . \end{aligned} \right. \tag{26}$$

The velocity vector localization θ_V, also termed the Makita localization after the work of Makita [17], [18], 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, $r_\mathrm{V} = 1$, and departures of r_V from 1 cause instability of images as the listener rotates his or her head. $r_\mathrm{V} >$

1 increases apparent image width, and $r_V < 1$ reduces it, but in general, values of r_V close to 1, say in the range of 0.85 to 1.1, are found not to degrade image quality greatly, but much smaller (such as <0.7) or much larger (such as > 1.3) values of r_V are generally unacceptable.

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

$$\left(\frac{G_i}{d_i}\right) e^{-j\omega\tau_i} \tag{27}$$

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

5 3 × 2 DECODERS

The preceding psychoacoustic theories are now applied to the reproduction of standard two-channel stereo signals L_2 and R_2 via a 3 × 2 energy-preserving matrix decoder of the form shown in Fig. 4 via three stereo loudspeakers L_3, C_3, and R_3, as shown in Fig. 10. We assume the use of a three-loudspeaker stereo layout for which $\theta_3 = 45°$, that is, which subtends a total angle of 90° at the listener, as shown in Fig. 1(c) or (f).

It is convenient to describe the stereo position of sounds in a two-channel stereo signal in terms of a panpot angle parameter θ, where $45° \geq \theta \geq -45°$, such that the L_2 signal has gain $\cos(45° - \theta)$ and the R_2 signal has gain $\cos(45° + \theta)$. This is the gain law of a constant-power or sine–cosine panpot of the kind discussed, for example, by Orban [19], and is such that the sound is panned to the left for $\theta = 45°$, at the center for $\theta = 0°$, and to the right for $\theta = -45°$.

Fig. 11 shows the psychoacoustic localization parameters r_V, θ_V, r_E, and θ_E computed for a sound with panpot angle θ when reproduced via a conventional two-loudspeaker stereo layout, such as that of Fig. 1(b), with $\theta_2 = 35°$, that is, 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 three-loudspeaker reproduction data. It will be seen that, as expected, the velocity and energy vector localization directions are the same for left, center, and right positions, but that at intermediate positions, the energy vector localization angle θ_E is wider than the velocity localization θ_V, being about twice as wide for near-center locations. This agrees with the well-established observation [20]–[22] that high-frequency two-loudspeaker stereo localization gives wider images than low-frequency localization. Also note the "detent" effect for energy vector localization whereby the localizations of sounds toward the edges of the stereo stage are pulled into the nearest loudspeaker—a phenomenon noted by Harwood [20] experimentally.

It will be seen that the values of r_V and r_E at the two edges of the stereo stage equal 1, denoting the expected good image stability of sounds emerging from only one loudspeaker, but that center-stage images have markedly reduced values of r_V and r_E, indicative of poor image stability.

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

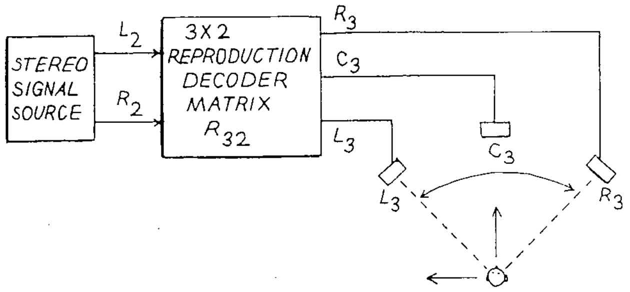

Fig. 10. Reproduction of two-channel stereo through three loudspeakers via matrix decoder.

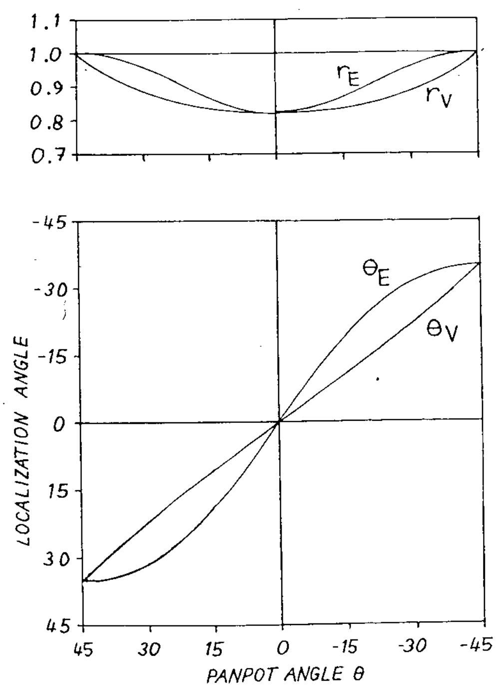

Fig. 11. Psychoacoustic localization parameters for two-loudspeaker stereo reproduction with $\theta_2 = 35°$.

a wide loudspeaker layout.

Fig. 13, with $\phi = 54.74°$, shows a somewhat narrower stage width of reproduction, with the relative width of the energy vector localization relative to the velocity localization even greater than for the two-loudspeaker stereo of Fig. 11. However, the center-stage values of r_V and r_E indicate that the image stability of central images is now only a little poorer than for Fig. 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 center-stage images. Fig. 14, with $\phi = 35.26°$, shows a virtually unchanged velocity vector localization di-

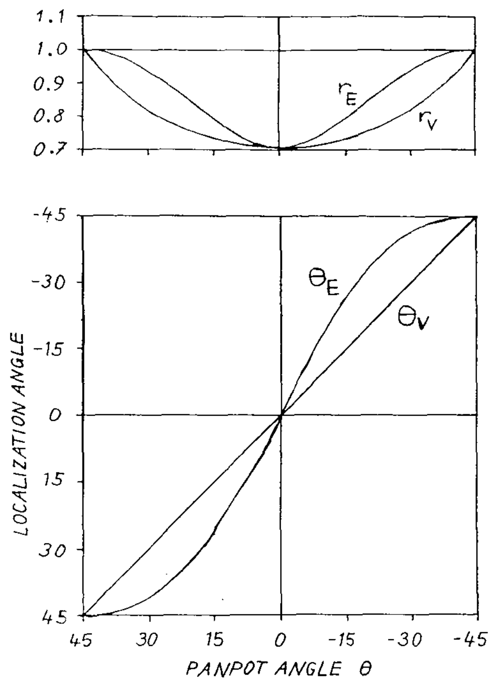

Fig. 12. Psychoacoustic localization parameters for two-loudspeaker stereo reproduction with $\theta_2 = 45°$, that is, for three-loudspeaker stereo reproduction with $\theta_3 = 45°$ via 3 × 2 energy-preserving matrix decoder with $\phi = 90°$.

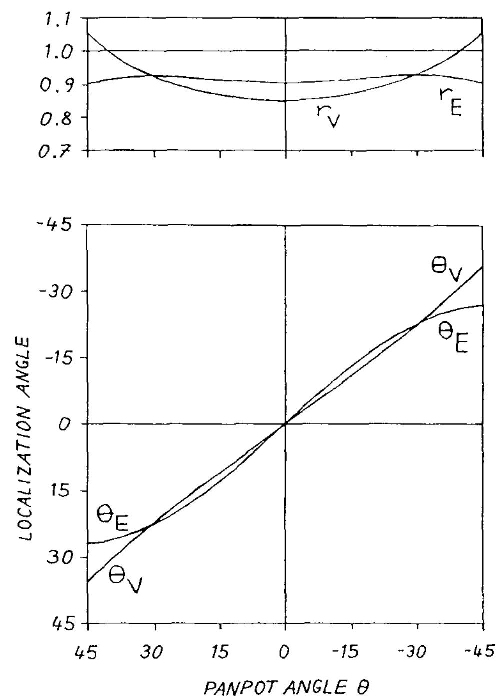

Fig. 14. Psychoacoustic localization parameters for three-loudspeaker stereo reproduction with $\theta_3 = 45°$ via 3 × 2 energy-preserving matrix decoder with $\phi = 35.26°$.

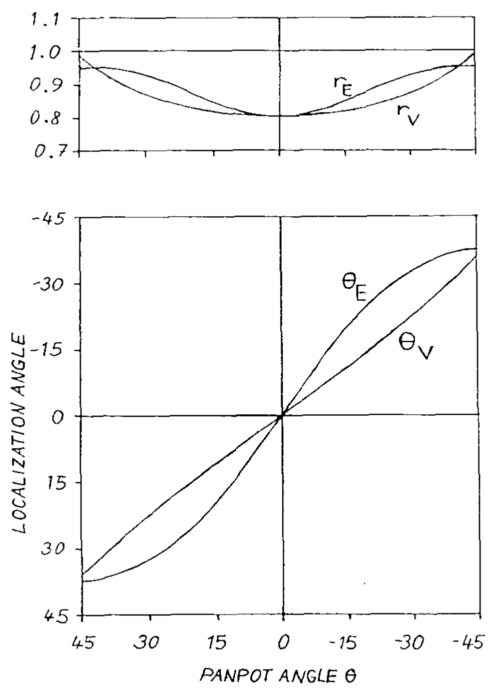

Fig. 13. Psychoacoustic localization parameters for three-loudspeaker stereo reproduction with $\theta_3 = 45°$ via 3 × 2 energy-preserving matrix decoder with $\phi = 54.74°$.

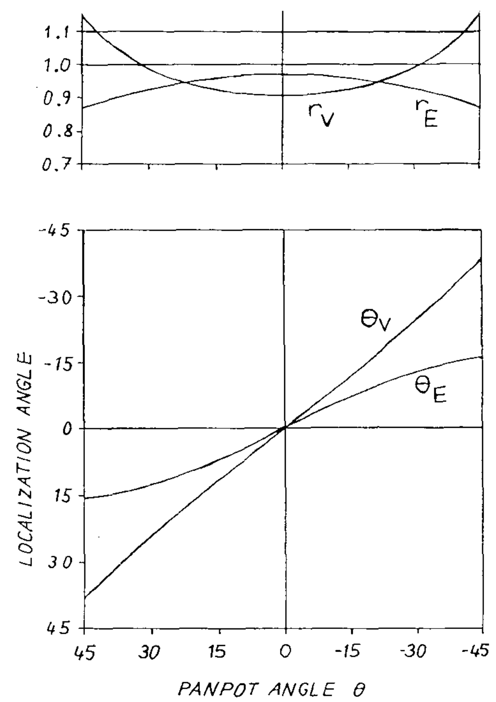

Fig. 15. Psychoacoustic localization parameters for three-loudspeaker stereo reproduction with $\theta_3 = 45°$ via 3 × 2 energy-preserving matrix decoder with $\phi = 19.47°$.

rection, but the energy vector localization 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 of 1 than even for Fig. 11, and the energy vector magnitude r_E is very nearly constant in value across the entire 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 Fig. 12 for center-stage images and is about 0.54 of its value in Fig. 11, giving much improved central image stability.

The main flaw with the localization behavior shown in Fig. 14 is the excessively narrow energy vector localization for extreme left and right images, which causes this decoder ($\phi = 35.26°$) to lose a sense of adequate width.

As ϕ is reduced further to 19.47°, as in Fig. 15, the velocity vector localization direction θ_V has hardly changed, but the energy vector localization 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 1, 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 stability becomes perfect, θ_V is not much changed, but $\theta_E = 0°$ 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 these theoretical predictions, and in particular that, for a central listener, $\phi = 35.26°$ 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 tradeoffs at different frequencies can be made.

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

If this band-splitting filter has outputs that sum to its inputs, then the difference S_2 channel processing in Fig. 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 three- or more-way band splitting 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 localization 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 $\phi = 35°$ 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, for both central and noncentral listeners, due to the failure of energy vector and Haas-effect localization models at these frequencies.

On most program 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 two-loudspeaker stereo by filtering out these frequencies. Therefore a good strategy for ameliorating the loss of stage width with $\phi = 35°$ 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 program 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 $\phi_H = 55°$ or larger do. However, it is also found that

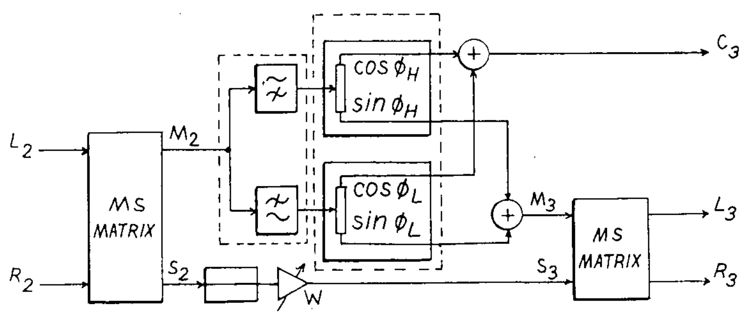

Fig. 16. Frequency-dependent version of 3×2 matrix decoder of Fig. 4, using a band-splitting filter network preceding the sine–cosine gains and a possible all-pass phase compensation in the difference channel.

ϕ_H much larger than 55° has a high-frequency "hole in the middle" with inadequate center-stage sounds. The value $\phi_H = 55°$, or thereabouts, gives both a wide stage and a good spread of sound among all three loudspeakers. Informal listening tests to full-range material confirm that a decoder of the form shown in Fig. 16 with $\phi_L = 35°$ and $\phi_H = 55°$ or thereabouts seems to be the most satisfactory 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 band-splitting 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 preceding values of ϕ_L and ϕ_H are consistently preferable to other choices. However, within limits, the characteristics of the band-splitting filter are not so critical.

We have found that the transition frequency of the band-splitting filter should not be below about 5 kHz, but that otherwise the transition frequency is not very critical. Also, the crossover between the two frequency ranges should not be too sharp since the ears object to sudden changes of behavior with frequency, but otherwise the crossover 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 center loudspeaker has less bass power handling than the outer loudspeakers, then it is convenient to choose ϕ near 90° at these frequencies, whereas if only the center loudspeaker can handle deep bass, a value of ϕ approaching 0° at very low frequencies is more apt. If all three loudspeakers 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 loudspeakers, maximizing bass power handling and giving maximum reinforcement of bass from the three loudspeakers.

If a frequency-independent decoder has to be used, we have found that $\phi = 45°$ 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 center loudspeaker proposal [2]–[5], while being audibly inferior to an optimized frequency-dependent decoder, for both central and noncentral listeners.

In applications where central image stability is critical, such as stereo television 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, which are important for atmosphere in dramatic programs, news material, and sports broadcasts.

6 PRESERVATION DECODERS

While the 3 × 2 matrix decoders are optimally designed to improve on the inherent limitations of two-loudspeaker reproduction, the degree of improvement possible by reproducing three- or four-loudspeaker stereo via more loudspeakers is somewhat reduced. In this section we present or main results, which are the determination of what we term preservation decoders, that is, $n_2 \times n_1$ energy-preserving matrix decoders that substantially preserve the values of the stereo localization parameters (apart from the effects of any overall change of angular stage width caused by the overall stage width of the loudspeaker layout).

Because of the possibility of forming composite decoders, indicated in Fig. 8, 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 loudspeakers in the layouts assumed, shown in Figs. 1(b)–(e). However, using the design procedures described in this section, 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:

$$\begin{aligned} &\theta_2 = 35°, \qquad \theta_3 = 45°, \qquad \theta_4 = 50°, \\ &\theta_5 = 1/3\,\theta_4, \qquad \theta_6 = 54°, \qquad \theta_7 = 1/2\,\theta_6 . \end{aligned} \tag{28}$$

These reference values have been chosen because 1) they are such that the angle between any adjacent pair of loudspeakers in a layout is the same as between other adjacent pairs in the same layout, and 2) 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 be otherwise.

Unlike in the two-channel case, there is no standard panpot law for three or more channel stereo (see the discussion in [23]), so that one cannot optimize 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 loudspeaker gains G_i of Fig. 9 can be chosen over quite 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, and to compute their localization 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 localization parameters via the matrix decoder are broadly similar or identical to the original localization parameters.

This procedure is quite complicated in that the mathematical equations for the localization 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 localization parameter to be exactly preserved. For example, if r_E is close to 1 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 prioritize localization parameters in deciding exactly what constitutes a good approximation to "preserving" them. In practice we do this by concentrating on the localization angles θ_V and θ_E and attempting to preserve these.

More precisely, we choose the following n-loudspeaker "stereo test signal" loudspeaker gains G_i: the cases where only one loudspeaker is excited with $G_k = 1$ and $G_i = 0$ for $i \neq k$, and the cases where a pair of loudspeakers is excited with $G_k = G_l = 1$ and $G_i = 0$ when $k \neq i \neq l$, that is, with equal in-phase gains. This yields $\frac{1}{2}n(n - 1)$ 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 kth loudspeaker, $\theta_V = \theta_E = \hat{\theta}_k$ and $r_V = r_E = 1$, and for an equal feed to the kth and lth loudspeakers,

$$\begin{aligned} \theta_V &= \theta_E = 1/2(\hat{\theta}_k + \hat{\theta}_l) \\ r_V &= r_E = \cos 1/2(\hat{\theta}_k - \hat{\theta}_l) \ . \end{aligned} \tag{29}$$

Because of left–right symmetry in the loudspeaker layouts and decoder equations, we need only consider one of these signals and not its mirror image. If the localization 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

$$\theta_V' = \theta_E' \tag{30}$$

for all $\frac{1}{2}n(n + 1)$ stereo test signal gains considered. For left–right symmetric test signals, this is automatically true since symmetry implies $\theta_V = \theta_E = \theta_V' = \theta_E' = 0°$, and if Eq. (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 Eqs. (30), and it can be shown that the number of Eqs. (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 Eq. (30) is satisfied for (L_2, R_2) gains (1, 0); for the 4×3 preservation decoder, we must find those values of the two decoder parameters ϕ_3 and ϕ_D for which Eqs. (30) are 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 four free decoder parameters (a, b, c) with $a^2 + b^2 + c^2 = 1$, ϕ_4, and ϕ_5 for which Eqs. (30) are 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 Eqs. (30) are 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 of simultaneous equations [Eqs. (30)] is highly nonlinear, and so must be solved by numerical methods, such as 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 search of plausible values must be undertaken.

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

$$\phi = 50.36° \tag{9'}$$

for the 3×2 decoder of Eqs. (9),

$$\phi_3 = 10.57°, \qquad \phi_D = 28.64° \tag{10'}$$

for the 4×3 decoder of Eqs. (10) and

$$\begin{aligned} (a, b, c) &= (0.6164, 0.6558, 0.4359), \\ \phi_4 &= 51.64°, \qquad \phi_5 = 9.64° \end{aligned} \tag{11'}$$

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

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

It can be seen from these tables that the values of θ_V and θ_E are preserved to within about 1°, 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, such as for central images in the 3×2 and 5×4 decoder cases—a useful if marginal improvement.

Investigations of the before-and-after localization parameters of these decoders with signal gains according to various three- and four-channel panpot laws have indicated that these decoders maintain the values of localization 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 localization before and after decoding seem invariably to be below 2° 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 loudspeaker layout—something that is inevitable given that any crosstalk 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. The Appendix lists the matrix equations of these preservation decoders in direct terms, that is, in terms of the left–right form of the loudspeaker feed signals.

For loudspeaker layouts with angles θ_p other than the standard reference values θ_p given in Eqs. (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 loudspeaker layouts with various indicated values of θ_3, θ_4, and θ_5. It can be seen that the values of these preservation decoder parameters do not vary enormously with the precise angular disposition of the loudspeaker layout. However, detailed computations of the localization parameters indicate that the best behaved values of r_E are obtained if θ_5 is near $\frac{1}{3}\theta_4$. 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 center-stage sounds. Space precludes full details of these results here, but nevertheless it is expected that preservation decoders should prove useful for most reasonable multi-loudspeaker stereo layouts.

Table 1. Gains of two-loudspeaker stereo test signals and their localization parameters, direct via two loudspeakers with $\theta_2 = 35°$, and via 3 × 2 preservation decoder via three loudspeakers with $\theta_3 = 45°$.

Gains		Two-loudspeaker parameters				Three-loudspeaker 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 2. Gains of three-loudspeaker stereo test signals and their localization parameters, direct via three loudspeakers with $\theta_3 = 45°$, and via a 4 × 3 preservation decoder via four loudspeakers with $\theta_4 = 50°$ and $\theta_5 = \frac{1}{3}\theta_4$.

Gains			Three-loudspeaker parameters				Four-loudspeaker 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 3. Gains of four-loudspeaker stereo test signals and their localization parameters direct via four loudspeakers with $\theta_4 = 50°$ and $\theta_5 = \frac{1}{3}\theta_4$, and via a 5 × 4 preservation decoder via five loudspeakers with $\theta_6 = 54°$ and $\theta_7 = \frac{1}{2}\theta_6$.

Gains				Four-loudspeaker parameters				Five-loudspeaker 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 4. Gains of three-loudspeaker stereo test signals and their localization parameters, direct via three loudspeakers with $\theta_3 = 45°$, and via a 5 × 3 composite preservation decoder via five loudspeakers with $\theta_6 = 54°$ and $\theta_7 = \frac{1}{2}\theta_6$.

Gains			Three-loudspeaker parameters				Five-loudspeaker 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

7 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 loudspeakers, whereas the use of a greater number of loudspeakers should allow some improvement of at least some of the localization 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 1. Second, preservation decoders were designed assuming a localization theory that becomes inaccurate at high frequencies, particularly above 5 kHz.

Energy-preserving matrix decoders that "improve" the localization parameters, particularly r_E, at low and middle frequencies, and that have modified behavior above about 5 kHz to take into account the modified psychoacoustics of localization 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 tradeoff among various psychoacoustic requirements.

However, it is found that since the inherent defects of well-made three- and four-loudspeaker stereo sound are markedly smaller than for two-loudspeaker stereo sound, the room for improvement in three- and four-loudspeaker material is smaller than in the two-loudspeaker case. Thus the decoder parameters for improvement decoders with three- or four-loudspeaker 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 a preservation decoder. Thus, for example, a 4×2 improvement decoder can be formed by cascading the 3×2 improvement decoder with $\phi = 35°$ below 5 kHz and $\phi = 55°$ above 5 kHz with the 4×3 preservation decoder given above, with, say, $\phi_3 = 10.57°$ and $\phi_D = 28.64°$, giving a 4×2 decoder as in Eqs. (13) and Fig. 5, for which $\phi_{42} = \phi - \phi_3 = 25°$ approximately below 5 kHz and $\phi_{42} = 45°$ approximately above 5 kHz and $\phi_D = 28.6°$. 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°.

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

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

Table 5. Values of 3×2 preservation decoder parameter ϕ for various values of three-loudspeaker layout half-angle θ_3.

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

Table 6. Values of 4×3 preservation decoder parameters ϕ_3 and ϕ_D for various values of four-loudspeaker layout angles θ_4 and θ_5.

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

8 DELAY COMPENSATION

The preceding work has been applied to stereo layouts where all loudspeakers are equally distant from a listener at the ideal stereo seat. These decoders can be used with other stereo loudspeaker layouts, such as the one shown in Fig. 1(g), 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.

Fig. 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 the loudspeakers that are closest to the central listener, so that the path-length differences in the acoustic field are compensated. Thus if a central loudspeaker is 0.5 m closer than the outer loudspeakers, 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 loudspeakers 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 loudspeakers 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, as shown in Fig. 1(f). This has been realized since the 1950s [24] and is based on using the loudspeakers' polar diagrams to help compensate for the Haas effect. The optimum degree of toeing in is very dependent on the actual characteristics of the loudspeakers used.

9 CONCLUSIONS

This paper has presented detailed results on the optimum psychoacoustic design of matrix decoders for reproducing n_1-loudspeaker 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 localization, 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 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 loudspeaker layout dispositions, and adjustments are necessary only for the most critical applications.

This paper has, in particular, provided a detailed optimization of a matrix decoder for reproducing two-channel stereo via three 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 it gives most listeners the illusion of a continuous sound stage between the two loudspeakers. It has been found preferable both to traditional two-loudspeaker stereo and to Bell/Klipsch

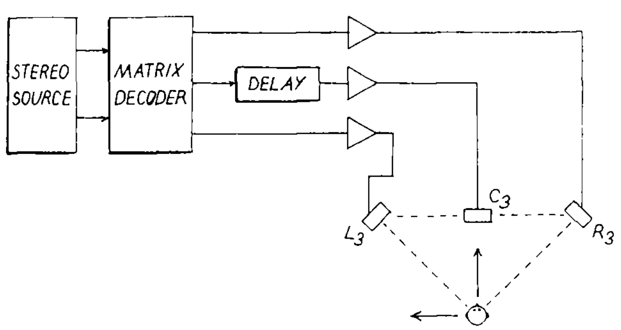

Fig. 17. Use of delay compensation to compensate for differing loudspeaker distances.

bridged center loudspeaker stereo. It is suitable for a wide variety of applications, including high-quality music reproduction, television stereo reproduction, and other applications such as portable reproducers with integral loudspeaker systems and in-car use.

The other more general $n_2 \times n_1$ decoders, tabulated in the Appendix for particular loudspeaker layouts, provide a detailed solution to the problem of feeding stereo source material to a larger number of loudspeakers for all numbers of loudspeakers up to five, so that, for example, three-channel stereo films can be reproduced via four or five front-stage loudspeakers. This may be particularly useful in connection with HDTV stereo systems, in that it makes the choice of the 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 optimum design of a matrix transmission system for recording and transmitting multichannel stereo, and to the optimization of three- and four-channel panpots, but the details of this have been published elsewhere [25], [26].

It is believed that this paper provides a systematic basis for handling multiloudspeaker 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 multiloudspeaker stereo systems, such as have been used previously in cinema or HDTV applications, and is particularly optimized toward domestic-scale applications.

10 PATENT NOTE

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

11 ACKNOWLEDGMENT

The author wishes to acknowledge the encouragement given to him in the course of this work by Peter Craven, Roger Furness, and John S. Wright.

12 REFERENCES

[1] J. Eargle (Ed.), *Stereophonic Techniques* (Audio Engineering Society, New York, 1986).

[2] J. C. Steinberg and W. B. Snow, "Auditory Perspective—Physical Factors," *Elect. Eng.*, vol. 53, pp. 12–15 (1934 Jan.), reprinted in [1].

[3] P. W. Klipsch, "Stereophonic Sound with Two Tracks, Three Channels by Means of a Phantom Circuit (2PH3)," *J. Audio Eng. Soc.*, vol. 6, pp. 118–123

(1958 Apr.); reprinted in [1].

[4] P. W. Klipsch, "Stereo Geometry Tests," *IRE Trans. Audio*, vol. AU-10, pp. 174–176 (1962 Nov./Dec.); reprinted in [1].

[5] P. W. Klipsch, "Signal Mutuality in Stereo Systems," *IRE Trans. Audio*, vol. AU-8, pp. 168–173 (1960 Sept./Oct.); reprinted in [1].

[6] D. J. Meares, "High Quality Sound for High-Definition Television," *Proc. Audio Eng. Soc. 10th Int. Conf.* (1991 Sept.), pp. 163–177.

[7] S. Komiyama, "Subjective Evaluation of Angular Displacement between Picture and Sound Directions for HDTV Sound Systems," *J. Audio Eng. Soc.*, vol. 37, pp. 210–214 (1989 Apr.).

[8] G. Theile, "HDTV Sound Systems: How Many Channels?" *Proc. Audio Eng. Soc. 10th Int. Conf.* (1991 Sept.), pp. 147–162.

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

[10] M. A. Gerzon, "General Metatheory of Auditory Localisation," presented at the 92nd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 40, p. 447 (1992 May), preprint 3306 (originally submitted to EBU and FCC proceedings on quadraphonic broadcasting during 1977–80).

[11] M. A. Gerzon, "Sound Encoding Specification," presented at the 56th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 25, pp. 332, 335 (1977 May), preprint 1199.

[12] D. H. Mershon and J. N. Bowers, "Absolute and Relative Cues for the Auditory Perception of Egocentric Distance," *Perception*, vol. 8, pp. 311–322 (1979).

[13] M. A. Gerzon, "Fixing It Outside the Mix," *Studio Sound*, vol. 32, pp. 78–93 (1990 Sept.).

[14] M. A. Gerzon, "Ambisonics in Multichannel Broadcasting and Video," *J. Audio Eng. Soc.*, vol. 33, pp. 859–871 (1985 Nov.).

[15] H. Wallach, E. B. Newman, and M. R. Rosenzweig, "The Precedence Effect in Sound Localization," *Am. J. Psychol.*, vol. 52, pp. 315–336 (1949).

[16] H. Haas, "The Influence of a Single Echo on the Audibility of Speech," *J. Audio Eng. Soc.*, vol. 20, pp. 145–159 (1972 Mar.).

[17] Y. Makita, "On the Directional Localization of Sound in the Stereophonic Sound Field," *E.B.U. Rev.*, pt. A, no. 73, pp. 102–108 (1962).

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

[19] R. Orban, "Notes on Panpots," *J. Audio Eng. Soc. (Project Notes/Engineering Briefs)*, vol. 19, pp. 936–939 (1971 Dec.).

[20] H. D. Harwood, "Stereophonic Image Sharpness," *Wireless World*, vol. 74, pp. 207–211 (1968).

[21] J. S. Bower, "The Subjective Effects of Interchannel Phase-Shifts on the Stereophonic Image Localization of Narrowband Audio Signals," Rep. BBC RD 1975/28, BBC Res. Dep. (1975 Sept.).

[22] H. A. M. Clark, G. F. Dutton, and P. B. Vanderlyn, "The 'Stereosonic' Recording and Reproducing System," *J. Audio Eng. Soc.*, vol. 6, pp. 102–117 (1958 Apr.); reprinted in [1].

[23] M. A. Gerzon, "Three Channels. The Future of Stereo?" *Studio Sound*, vol. 32, pp. 112–125 (1990 June).

[24] W. B. Snow, "Basic Principles of Stereophonic Sound," *J. SMPTE*, vol. 61, pp. 567–589 (1953 Nov.); reprinted in [1].

[25] M. A. Gerzon, "Hierarchical Transmission System for Multispeaker Stereo," to be published *J. Audio Eng. Soc.*, vol. 40 (1992 Sept.).

[26] M. A. Gerzon, "Panpot Laws for Multispeaker Stereo," presented at the 92nd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 40, p. 447 (1992 May), preprint 3309.

APPENDIX A
EXPLICIT EQUATIONS FOR PRESERVATION DECODERS

The preservation decoder equations for the loudspeaker layouts of Fig. 1(b)–(e) with angles θ_p as in Eqs. (28), computed from Eqs. (9a), (9b), (9′), (10a), (10b), (10′), (11a), (11b), (11′), and (12), along with the MS equations [Eqs. (1–3)], are given here in terms of the loudspeaker 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:

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

THE AUTHOR

Michael A. Gerzon was born in Birmingham, England, in 1945 and received an M.A. degree in mathematics at Oxford University in 1967, where he did postgraduate work in axiomatic quantum theory.

His interest in audio stemmed from interests in music, sensory perception, and information theory. Since 1967 he has been active in sound recording live music, recording artists as diverse as Emma Kirkby, Michael Tippett, Pere Ubu, and Anthony Braxton, and has recorded music for around 15 LP and CD releases.

Arising from this interest, since 1971, he has earned his living from consultancy work in audio and signal processing. He was the main inventor of Ambisonic surround sound technology, working in the 1970s and early 1980s with the British National Research Development Corporation. With Peter Craven, he co-invented the sound field microphone. He also developed mathematical models for human directional psychoacoustics for use in the design of directional sound reproduction systems, and was made a Fellow of the AES in 1978 for this work. He was awarded the AES gold medal in 1991 for his work on Ambisonics.

He has published over 80 articles and papers in the field of audio and signal processing, including numerous papers on stereo and surround-sound systems. He has so far been granted 8 British patents and corresponding applications in the USA, Japan, and other countries. He has published papers on practical and theoretical aspects of linear and nonlinear signal processing and systems theory, digital reverberation, data compression, and spectral analysis.

He is currently working on digital signal processing algorithms with Audio Animation Inc., and has done consultancy work with Peter Craven and B & W Loudspeakers on digital room equalization. He is also currently working with Trifield Productions Ltd. on multispeaker stereo and Ambisonic surround-sound systems for use with TV, HDTV, and cinema applications.

His technical interests include image data compression systems, which he has been working on since 1965, and the use of advanced mathematical methods, including nonlinear operator theory and *-algebra techniques in system design. Other interests include writing poetry, the history of contemporary musics, and the mathematical and conceptual foundations of theoretical physics.