Hierarchical Transmission System for Multispeaker Stereo*

MICHAEL A. GERZON**, *AES Fellow*

Oxford OX2 6DJ, UK

A matrix method is described for encoding and decoding n-loudspeaker stereo to and from transmission channels, suitable for broadcasting and recording applications. The transmission signals for different numbers n of loudspeakers are mutually compatible such that listeners receiving n_1-loudspeaker stereo transmissions via n_2 loudspeakers will obtain psychoacoustically optimum results. This allows frontal-stage stereo signals originated for different numbers of loudspeakers to be mixed together freely, simplifying the operational aspects of multiloudspeaker stereo and HDTV stereo systems. Such hierarchical transmission systems are based on psychoacoustic decoders given in the companion paper "Optimum Reproduction Matrices for Multispeaker Stereo."

0 INTRODUCTION

There has been some controversy about how many loudspeakers and channels are needed to provide stable frontal-stage stereo images to match visual images for high-definition television (HDTV). While it is generally agreed that two-loudspeaker stereo provides inadequate image stability, some have argued that three-loudspeaker stereo is practically adequate, whereas others have argued the case for using four or more loudspeakers to cover the frontal stage [1]–[6]. The film industry got away with the use of three front-stage loudspeakers for wide-screen film, although as early as the 1950s, many argued that up to five loudspeakers were essential for good imaging [7].

In this paper we argue the case for not making a decision on these matters, recognizing that there is a tradeoff between the best quality of image resolution and the convenience or inconvenience of using few or many loudspeakers. We present an alternative approach that permits the broadcaster to choose the number of front-stage transmission channels that are operationally convenient, and the listener to decode any transmission to the listener's choice of numbers of loudspeakers. This is done by using a hierarchical approach to transmission, whereby each added transmission channel carries finer detail information about directional resolution that can be used or discarded at any stage.

Hierarchical transmission systems are not new; the UMX system of Cooper and Shiga [8], [9] provided a hierarchical system of surround-sound transmission and reception using any number of channels greater than one, and the author [10] and Gibson et al. [11] described hierarchical approaches to the transmission of periphony, that is, with-height full-sphere surround sound. However, these previous surround-sound hierarchies are not suitable for systems such as HDTV where one requires a very high directional resolution throughout a limited sector of directions across a screen without having to have a corresponding directional resolution (with a correspondingly impractical number of loudspeakers) in all other directions.

The problem with designing true hierarchical systems for frontal-stage multiloudspeaker stereo has been the lack of an adequate understanding of the psychoacoustics of such systems. A companion paper [12] has provided the solution to this problem by deriving the optimum matrix decoders for reproducing n_1-loudspeaker stereo signals via a greater number n_2 of loudspeakers. The derivation of these decoders was mathematically complicated, but having derived them, it becomes possible to design optimized hierarchical sys-

*Manuscript received 1991 February 23; 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.

tems for transmitting multiloudspeaker stereo via any available number of channels.

This paper discusses the desirable properties of hierarchical transmission systems, gives the general mathematical procedure for deriving a hierarchical system based on known $n_2 \times n_1$ matrix decoders, and derives equations for a hierarchical system for transmitting two-, three-, four-, and five-loudspeaker stereo via two, three, four and five transmission channels, and for reproducing these channels via two, three, four, and five stereo loudspeakers.

1 HIERARCHICAL SYSTEMS

In this section we set out the basic ideas behind hierarchical systems of encoding and decoding, which is an extension of the idea of mono–stereo compatibility. Fig. 1 shows typical loudspeaker layouts for use with frontal stage stereo systems using from one to five loudspeakers. All layouts have all loudspeakers at the same distance from the listener, and all loudspeakers facing the listener at an ideal stereo seat. For numerical subscripts p, central loudspeakers are denoted by C_p, loudspeakers to the left of center are denoted by L_p, and their right mirror-image counterparts by R_p; the loudspeakers L_p and R_p are disposed in directions at an angle θ_p to the respective left and right of due front. The numerical subscripts p are as indicated in Fig. 1.

If stereo signals are created for reproduction via n loudspeakers, then they must be encoded for transmission or recording via n transmission channels (and we treat a channel on tape, disk, film, or a digital recording, storage, or playback medium as a "transmission" channel). The first requirement of a hierarchical system of encoding is that one should have transmission channel signals $T_1, T_2, T_3, \ldots$ such that mono signals are coded just into T_1, two-loudspeaker stereo signals into T_1 and T_2, three-loudspeaker stereo signals into T_1 to T_3, and so forth, as shown in Fig. 2, where we denote T_1 by M, T_2 by S, and T_3 by T for consistency with previous notations used elsewhere. Thus M is a mono signal, S a stereo-side signal, T a third-channel signal, and T_i an ith transmission signal. The encoding hierarchy used in Fig. 2 encodes n-loudspeaker stereo (for n = 1, 2, 3, 4, and 5) into n transmission channels such that the first i of these channels are also used for i-loudspeaker stereo in, hopefully, a compatible way. The $n \times n$ encoding matrices used to encode n-loudspeaker stereo into the n transmission channels are denoted by E_{nn}.

Fig. 3 shows the corresponding inverse system of decoding the transmission channels for reproduction via any number of loudspeakers. A transmission decoding matrix D_{nn} is used to decode the n transmission channels T_1 to T_n into signals to feed n-loudspeaker stereo layouts.

Fig. 4 shows schematically the effect of encoding n_1-loudspeaker stereo into m transmission channels T_1 to T_m and of decoding these into n_2 loudspeaker feed signals. The overall encoding–transmission–decoding system (shown by dashed box) converts the n_1-loudspeaker stereo signals at the input into n_2-loudspeaker stereo signals at the output and constitutes what we term a reproduction decoder for converting stereo for one number of loudspeakers to stereo for another.

Obviously we wish that for n-loudspeaker inputs, n-channel transmission, and n-loudspeaker reproduction the output signals should be identical to the input signals. Thus D_{nn} should be the matrix inverse of E_{nn}, that is,

$$D_{nn} = E_{nn}^{-1} . \tag{1}$$

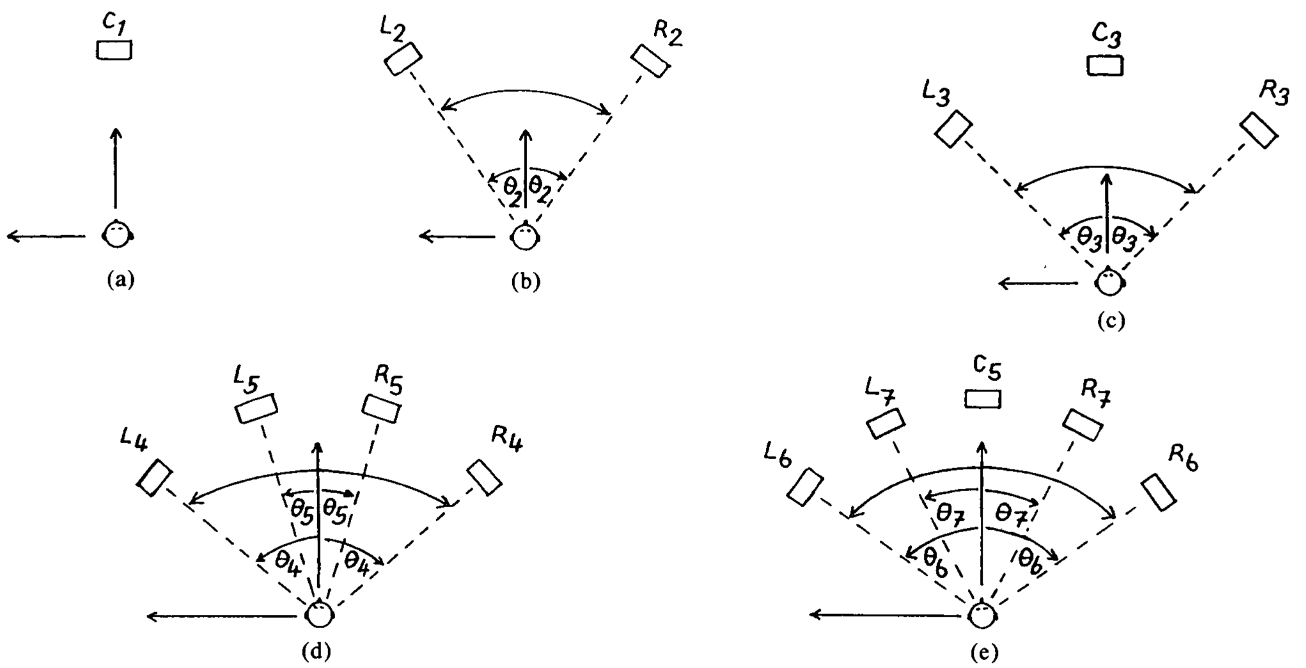

Fig. 1. Loudspeaker layouts for front-stage stereo using one to five loudspeakers, including angles and loudspeaker symbols.

However, besides replicating the input signals at the output when identical numbers of original loudspeaker channels, transmission channels, and final loudspeaker channels are used, we also wish for good results in other cases. In particular, increasing the number of transmission channels above the number n_1 of original loudspeaker channels should not change results, which means that for such n_1-loudspeaker stereo inputs, the additional transmission signals T_i, for $i > n_1$, should equal zero. Also, when the numbers n_1, m, and n_2 of original-loudspeaker, transmission, and final-loudspeaker channel signals, respectively, is such that

$$n_1 \leq m \leq n_2 \qquad (2)$$

so that no information is lost at any stage, we require that the resulting stereo effect should be either as similar as possible to or an improvement on the originally intended n_1-loudspeaker stereo effect.

In other words, when Eq. (2) is satisfied, the overall system shown in Fig. 4 should have the effect shown in Fig. 5, where an n_1-loudspeaker stereo signal is matrixed via a matrix reproduction decoder for reproduction via a larger number n_2 of loudspeakers, and this reproduction decoder should be such as to optimally preserve or improve the original stereo effect.

The problem of designing matrix reproduction decoders that convert signals intended for n_1-loudspeaker stereo reproduction into signals suitable for $n_2 > n_1$ loudspeaker stereo reproduction via the layouts of Figs. 1(b)–(e) with optimum psychoacoustic effects has recently been solved by the author [12]. This solution is mathematically very complicated to describe in general, but in this paper we need only refer to certain results and properties of these optimal reproduction decoders without needing the full theory of [12].

We introduce the following notations. A matrix reproduction decoder converting n_1-loudspeaker stereo signals into n_2 signals intended for reproduction via an n_2-loudspeaker stereo system is described by an $n_2 \times n_1$ array of matrix coefficients forming an $n_2 \times n_1$ reproduction matrix $R_{n_2n_1}$. A matrix encoder converting n_1 signals intended to feed an n_1-loudspeaker stereo layout into m transmission channel signals T_1 to T_m is similarly described by an $m \times n_1$ matrix denoted by E_{mn_1}, called the (transmission) encoding matrix, and a matrix transmission decoder converting m transmission channel signals into n_2 signals intended for feeding an n_2-loudspeaker stereo layout is similarly described by an $n_2 \times m$ transmission decoding matrix D_{n_2m}.

In particular, the effect of decoding an encoded signal is to form a reproduction decoder described by the product of the $n_2 \times m$ transmission decoding matrix with the $m \times n_1$ transmission encoding matrix, namely,

$$R_{n_2n_1} = D_{n_2m}E_{mn_1} \ . \qquad (3)$$

A system of encoding n_1-loudspeaker stereo signals for transmission via n transmission channels, and of decoding m transmission channels into n_2-loudspeaker stereo signals for all of a range of values of n_1, m, and n_2 is termed a hierarchical transmission system if:

1) Eq. (1) is satisfied, so that original signals are recovered when $n_1 = m = n_2$

2) All transmission signals T_i for $i > n_1$ are equal to zero.

3) The matrix reproduction decoder of Eq. (3) when Eq. (2) is satisfied is a psychoacoustically satisfactory decoder for reproducing the effect of an n_1-loudspeaker stereo signal via a larger number n_2 of loudspeakers.

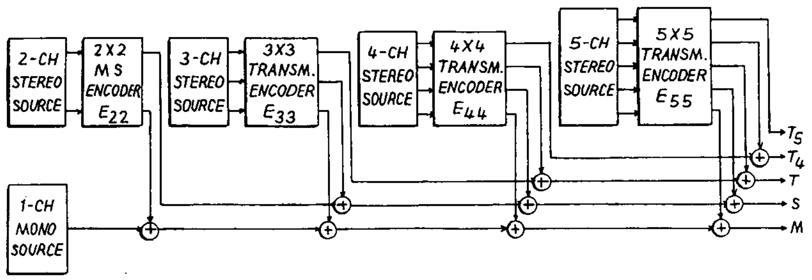

Fig. 2. Hierarchy of encoding transmission channel signals for one- to five-channel signals.

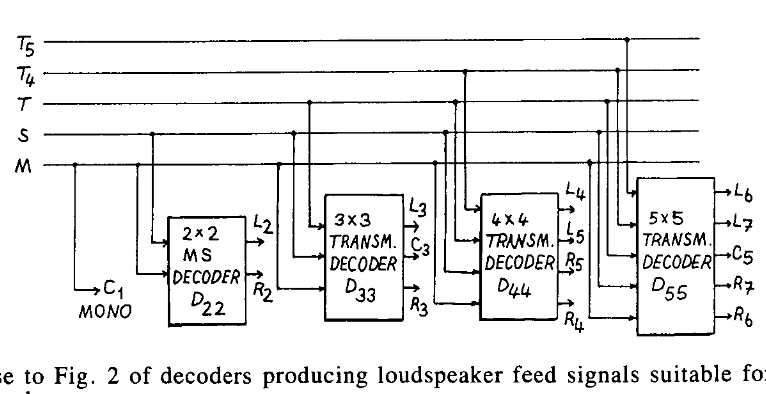

Fig. 3. Hierarchy inverse to Fig. 2 of decoders producing loudspeaker feed signals suitable for layouts of Figs. 1 from transmission channel signals.

Apart from the trivial case of sum-and-difference transmission of two-loudspeaker stereo, no hierarchical system of transmission for multiloudspeaker frontal-stage stereo has been described previously. However, hierarchical systems are nothing new for surround-sound systems, as mentioned in the Introduction.

This paper gives a systematic design procedure for designing hierarchical transmission systems based on the use of optimum reproduction decoders. The advantage of using a hierarchical system is seen in the schematic of Fig. 2, where stereo signals originated for different numbers of stereo loudspeakers can be mixed freely together and reproduced via any other number of loudspeakers.

The need for this ability to freely mix different types of stereo program together arises from the fact that, historically, stereo has developed using different numbers of loudspeakers in different areas, and it is often necessary to use stereo material originated in one area in another. For example, domestic audio and standard-definition stereo television have traditionally used two-channel stereo, and a huge back catalog of material exists in this form, which is difficult or even impossible to remix into three- or four-channel form. This back catalog includes not only commercial music and dramatic material, but large libraries of sound effects and library music intended for background use in stereo productions.

On the other hand, many films were prepared using three-, four-, or (in the 1950s) even five-loudspeaker frontal stereo forms. HDTV may use three- or four-loudspeaker stereo, and one then has the problem of mixing material from different sources together in a single production. It is not satisfactory to change transmission modes whenever material originated for different numbers of channels is used. There are two reasons.

First, such a change of mode requires that "flag" information indicating the number of channels be transmitted, and it is all too easy for flags to get lost in the complex production chain associated with HDTV, where the preparation of a program may involve material originated from many sources, including film, videotape, audio recordings, satellite relays, and rebroadcasts from both standard-definition and HDTV broadcasts from different parts of the world.

Second, one often needs to mix together sounds originated from different sources, such as a two-channel sound-effect recording with a four-loudspeaker stereo mix of actors' voices, or a four-channel "no-dialog" sound of a live event with a fewer number of dialog channels in a multilingual broadcast of a major event where the number of audio channels prevents the dialog from using the full number of channels in all the required languages. Thus the receiver, in a single reception mode, must be capable of handling all stereo source material at the same time within a mix.

Having standardized procedures for handling source material originated for different numbers of loudspeakers also means greatly reduced production prob-

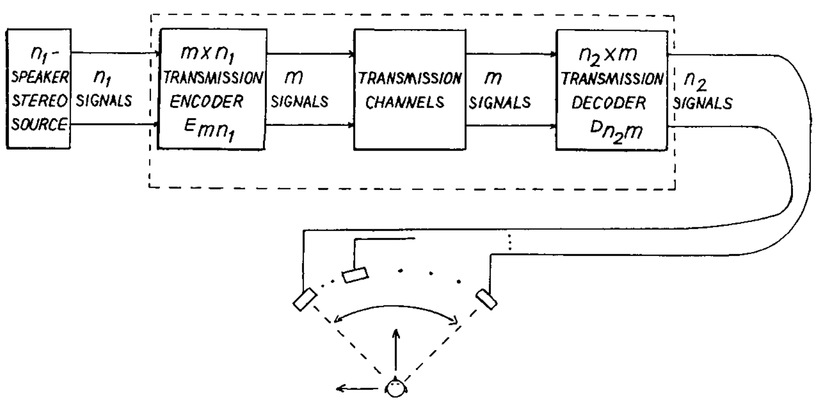

Fig. 4. Signal path via encoder, transmission channels, and decoder, constituting a matrix reproduction decoder (dashed box) converting n_1-loudspeaker stereo into n_2-loudspeaker stereo.

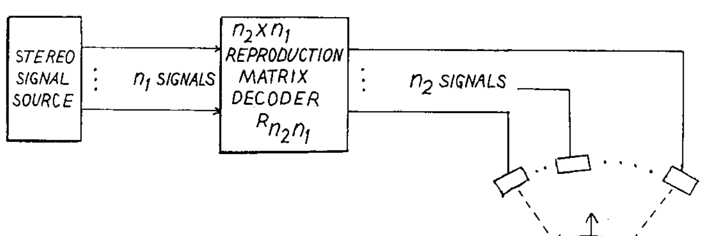

Fig. 5. Matrix reproduction decoder converting n_1-loudspeaker stereo signals into n_2-loudspeaker stereo signals.

lems and time spent on the audio side. In addition, the hierarchical approach means that, when program material is reconverted again and again to fit different numbers of loudspeakers down a long production, broadcast, and rebroadcast chain, there will be no degradation of results other than that caused by the smallest number n_B (termed the bottleneck number) of signal channels used in the intermediate chain. Thus if the bottleneck number is 3, the results will never be worse than if the original signal had been encoded into three transmission channels and decoded from those three channels, no matter how many intermediate stages of reconversion were actually used.

With the hierarchical approach, all production work can be done using the normal maximum number m of transmission signals T_1 to T_m, and any incoming signal can be converted into transmission signals T_1 to T_n with $n < m$ and mixed freely, as shown in Fig. 2. Otherwise no special production procedures are required unless one needs to alter the desired stereo effect of that signal.

2 COMPOSITE DECODERS

The possibility of forming hierarchical transmission systems depends on the construction of what I have termed composite decoders in [12]. The idea is that, if one has a matrix reproduction decoder, described by an $n_2 \times n_1$ matrix $R_{n_2n_1}$, for converting n_1-loudspeaker stereo signals into n_2-loudspeaker stereo signals, and also has a second matrix reproduction decoder, described by an $n_3 \times n_2$ matrix $R_{n_3n_2}$, for converting n_2-loudspeaker stereo into n_3-loudspeaker stereo, then one can connect the two decoders in series, as shown in Fig. 6, to form a matrix reproduction decoder, described by an $n_3 \times n_1$ matrix $R_{n_3n_1}$, for converting n_1-loudspeaker stereo into n_3-loudspeaker stereo, where

$$R_{n_3n_1} = R_{n_3n_2}R_{n_2n_1} \ . \tag{4}$$

Such a series connection of two matrix reproduction decoders, or a decoder whose input–output behavior is equivalent to such a series connection, is termed a composite decoder, and the two decoders of which it is formed, as shown in Fig. 6, are termed component decoders.

If the component decoders of a composite decoder have the individual properties of preserving or improving a specific psychoacoustic property of their incoming stereo sound, then the composite decoder also preserves or improves that specific psychoacoustic property. Thus to design an $n_3 \times n_1$ matrix reproduction decoder that preserves or improves specific aspects of the sound, one need only design $n_3 \times n_2$ and $n_2 \times n_1$ matrix reproduction decoders that have the same property. In particular, for $n_2 > n_1$ it is sufficient to be able to design $(n + 1) \times n$ matrix reproduction decoders having the desired properties for $n = n_1, n_1 + 1, \ldots, n_2 - 1$, and to connect them in series to form a composite decoder with the required property.

This approach of cascading $(n + 1) \times n$ matrix reproduction decoders, each preserving or improving stated psychoacoustic aspects of stereo sound, is illustrated in Fig. 7 up to the case $n = 4$, although the

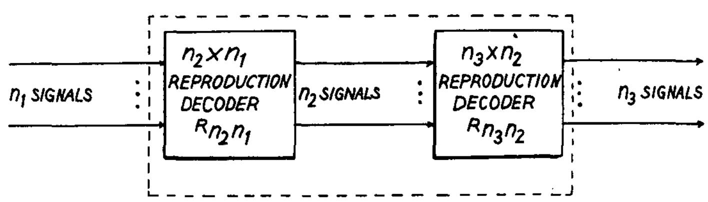

Fig. 6. Series connection of two matrix reproduction decoders to form composite matrix reproduction decoder for converting n_1-loudspeaker stereo into n_3-loudspeaker stereo (dashed box).

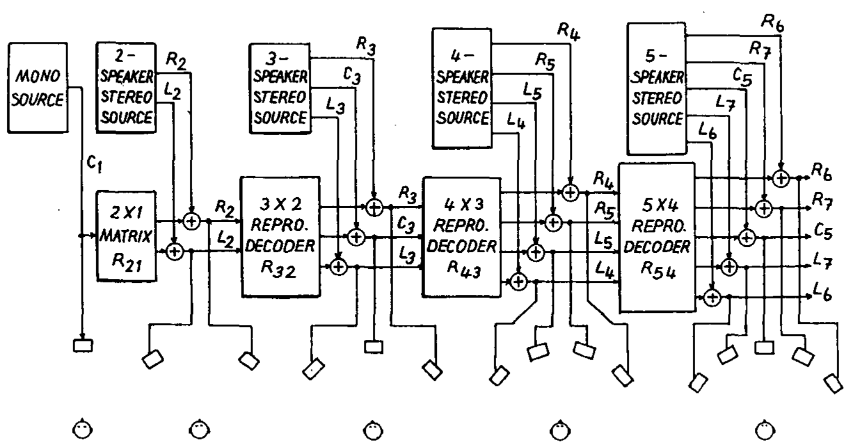

Fig. 7. Converting any number of stereo channels into any larger number by series-connected $(n + 1) \times n$ matrix reproduction decoders.

diagram extends to any number n. In this diagram stereo originated for any number n_1 of loudspeakers can be converted for reproduction via any larger number n_2 of loudspeakers, and mixed with signals intended for any other number n_3 of loudspeakers, simply by cascading $(n + 1) \times n$ matrix reproduction decoders $R_{(n+1)\,n}$ and mixing stereo source signals at each stage of the process.

Thus if one can find, say, psychoacoustically desirable 3×2, 4×3, and 5×4 matrix reproduction decoders that preserve or improve the stereo effect, then one can cascade them as in Fig. 7 to form composite decoders that are also psychoacoustically desirable in the same way.

The derivation of psychoacoustically desirable $(n + 1) \times n$ matrix reproduction decoders is detailed in [12], and the theory cannot easily be summarized. Although the optimum matrix depends slightly on the actual angles θ_p of the loudspeaker layouts in Fig. 1(b)–(e), this dependence is surprisingly small, so that we can, for the purposes of designing a transmission hierarchy, assume standardized loudspeaker layouts without causing any great error if the layouts are rather different in angular disposition from that assumed.

For reasons described in [12], the standardized layouts assumed in Fig. 1(b)–(e) have

$$\begin{aligned} &\theta_2 = 35°, \quad \theta_3 = 45°, \quad \theta_4 = 50°, \\ &\theta_5 = \tfrac{1}{3}\theta_4, \quad \theta_6 = 54°, \quad \theta_7 = \tfrac{1}{2}\theta_6 \,. \end{aligned} \tag{5}$$

The design procedure of [12] leads to $(n + 1) \times n$ matrix reproduction decoders that preserve the total signal energy of incoming signals and that, via the above loudspeaker layouts, substantially preserve the reproduced angular disposition of sounds, with the exception of the 3×2 decoder, which is designed to reduce anomalies in the normal reproduced effect of two-loudspeaker stereo and to improve the stability of center-stage images as the listener position changes in a room.

These $(n + 1) \times n$ matrix reproduction decoders satisfy the following equations, as shown in [12].

1) 3×2 matrix reproduction decoder with $\phi = 45°$:

$$\begin{pmatrix} L_3 \\ C_3 \\ R_3 \end{pmatrix} = \begin{pmatrix} 0.8536 & -0.1464 \\ 0.5000 & 0.5000 \\ -0.1464 & 0.8536 \end{pmatrix} \begin{pmatrix} L_2 \\ R_2 \end{pmatrix} \tag{6}$$

2) 4×3 "preservation" matrix reproduction decoder:

$$\begin{pmatrix} L_4 \\ L_5 \\ R_5 \\ R_4 \end{pmatrix} = \begin{pmatrix} 0.9303 & -0.1297 & 0.0527 \\ 0.3314 & 0.6951 & -0.1479 \\ -0.1479 & 0.6951 & 0.3314 \\ 0.0527 & -0.1297 & 0.9303 \end{pmatrix} \begin{pmatrix} L_3 \\ C_3 \\ R_3 \end{pmatrix} \tag{7}$$

3) 5×4 "preservation" matrix reproduction decoder:

$$\begin{pmatrix} L_6 \\ L_7 \\ C_5 \\ R_7 \\ R_6 \end{pmatrix} = \begin{pmatrix} 0.9535 & -0.1084 & 0.0590 & -0.0324 \\ 0.2533 & 0.7870 & -0.1989 & 0.0859 \\ -0.1349 & 0.5708 & 0.5708 & -0.1349 \\ 0.0859 & -0.1989 & 0.7870 & 0.2533 \\ -0.0324 & 0.0590 & -0.1084 & 0.9535 \end{pmatrix} \begin{pmatrix} L_4 \\ L_5 \\ R_5 \\ R_4 \end{pmatrix} . \tag{8}$$

The composite decoders formed by placing these decoders in series, as in Fig. 7, satisfy the following equations, whose matrices are obtained by taking the matrix product of these matrices.

1) 4×2 composite matrix reproduction decoder

$$\begin{pmatrix} L_4 \\ L_5 \\ R_5 \\ R_4 \end{pmatrix} = \begin{pmatrix} 0.7215 & -0.1561 \\ 0.6521 & 0.1728 \\ 0.1728 & 0.6521 \\ -0.1561 & 0.7215 \end{pmatrix} \begin{pmatrix} L_2 \\ R_2 \end{pmatrix} \tag{9}$$

2) 5×2 composite matrix reproduction decoder:

$$\begin{pmatrix} L_6 \\ L_7 \\ C_5 \\ R_7 \\ R_6 \end{pmatrix} = \begin{pmatrix} 0.6325 & -0.1525 \\ 0.6482 & 0.0287 \\ 0.3945 & 0.3945 \\ 0.0287 & 0.6482 \\ -0.1525 & 0.6325 \end{pmatrix} \begin{pmatrix} L_2 \\ R_2 \end{pmatrix} \tag{10}$$

3) 5×3 composite "preservation" matrix reproduction decoder:

$$\begin{pmatrix} L_6 \\ L_7 \\ C_5 \\ R_7 \\ R_6 \end{pmatrix} = \begin{pmatrix} 0.8407 & -0.1538 & 0.0557 \\ 0.5304 & 0.3648 & -0.0891 \\ -0.0279 & 0.8285 & -0.0279 \\ -0.0891 & 0.3648 & 0.5304 \\ 0.0557 & -0.1538 & 0.8407 \end{pmatrix} \begin{pmatrix} L_3 \\ C_3 \\ R_3 \end{pmatrix} \tag{11}$$

These particular matrix decoders derived via the schematic of Fig. 7 are special cases of more general decoders to be described later and discussed in [12]. However, they are close to the desirable reproduction decoder matrices $R_{n_2 n_1}$ for $n_2 > n_1$, and other practical proposals will have matrix coefficients differing from the above by only small amounts. We desire to construct a hierarchical encoding and decoding system, as shown in Figs. 2 and 3, so that the overall encoding–decoding matrices of the system shown in Fig. 4 constitute reproduction decoder matrices $R_{n_2n_1}$ such as given by Eqs. (6)–(11).

3 CONSTRUCTION OF TRANSMISSION HIERARCHIES

Given desired reproduction decoder matrices $R_{(n+1)\,n}$ for use as in Fig. 7 for constructing composite matrix reproduction decoders $R_{n_2n_1}$, we seek to construct encoder matrices E_{nn} and decoder matrices D_{nn} satisfying Eq. (1) such that transmission and reception via Figs. 2 and 3, with any noncoded channels set to zero, results in an overall matrix reproduction decoder $R_{n_2n_1}$. For $n_2 > n_1$, Eqs. (1) and (3) between them imply that

$$D_{n_2n_1} = R_{n_2n_1}D_{n_1n_1} \tag{12}$$

which, along with Eq. (4), implies that for any $m \leq n_2 < n_4$,

$$D_{n_4m} = R_{n_4n_2}D_{n_2m} \tag{13}$$

so that a transmission decoder followed by a matrix reproduction decoder, as shown in Fig. 8, is another transmission decoder (termed, naturally, a composite transmission decoder).

In a similar way, using Eqs. (1) and (3), we find that

$$E_{n_2n_1} = E_{n_2n_2}R_{n_2n_1} \tag{14}$$

for $n_2 > n_1$, which, along with Eq. (4), implies more generally that for any $n_1 < n_3 \leq m$,

$$E_{mn_1} = E_{mn_3}R_{n_3n_1} \tag{15}$$

so that a matrix reproduction decoder followed by a transmission encoder, as shown in Fig. 9, is another transmission encoder (termed, naturally, a composite transmission encoder).

In particular, from Eq. (12) we have

$$D_{(n+1)\,n} = R_{(n+1)\,n}D_{nn} \tag{16}$$

and also, from Fig. 3, $D_{(n+1)\,n}$ is obtained from $D_{(n+1)\,(n+1)}$ simply by setting the channel signal T_{n+1} equal to zero so that $D_{(n+1)\,n}$ is simply the first n columns of the $(n+1) \times (n+1)$ matrix $D_{(n+1)\,(n+1)}$.

This leads to the following method of constructing the hierarchical system of matrix encoding and decoding for a predetermined choice of reproduction decoding matrices $R_{(n+1)\,n}$, illustrated in the flow diagram of Fig. 10.

Suppose that, for all n up to a value n', one has already determined the form of D_{nn} and its inverse E_{nn}. For example, for conventional mono,

$$E_{11} = D_{11} = [1] \tag{17}$$

that is, the mono signal is transmitted straight down the mono channel M, and for conventional stereo in MS, or sum-and-difference, transmission form,

$$\begin{pmatrix} M \\ S \end{pmatrix} = 2^{-1/2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} L_2 \\ R_2 \end{pmatrix} \tag{18}$$

and the inverse decoding equation is

$$\begin{pmatrix} L_2 \\ R_2 \end{pmatrix} = 2^{-1/2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} M \\ S \end{pmatrix} \tag{19}$$

where the gain factor $2^{-1/2} = 0.7071$ and the polarity of the S signal are convenient matters of convention.

Now suppose that, at a given stage of designing the encoding–decoding hierarchy for transmission, one has determined E_{nn} and D_{nn} for every $n \leq n'$. For example, we already have done this for $n' = 2$ in Eqs. (18) and (19). Then, from the foregoing results, construct the

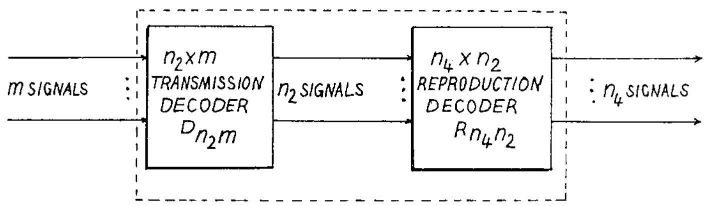

Fig. 8. Composite transmission decoder formed by series connection of transmission decoder and matrix reproduction decoder.

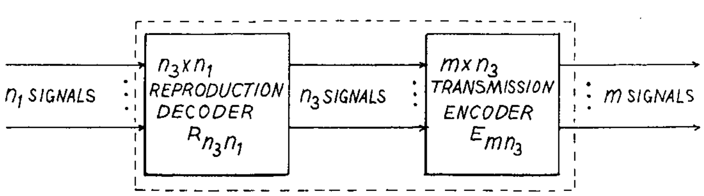

Fig. 9. Composite transmission encoder formed by series connection of matrix reproduction decoder and transmission encoder.

$(n' + 1) \times (n' + 1)$ matrix $D_{(n'+1)(n'+1)}$ to be a matrix whose first n' columns form the $(n' + 1) \times n'$ matrix

$$R_{(n'+1)n'} D_{n'n'}$$

of Eq. (16), and whose remaining column is chosen "arbitrarily" (that is, according to any criterion the designer wishes) to be linearly independent of the first n' columns, that is, not to be any linear combination of the first n' columns. This linear independence is necessary if the inverse matrix

$$E_{(n'+1)(n'+1)} = D_{(n'+1)(n'+1)}^{-1}$$

describing the encoder, is to exist. Having computed the encoder as the inverse of the decoder, n' is increased by 1 and the whole process is repeated until the desired maximum value of n' is achieved.

Of course, in practice the last column of the $(n + 1) \times (n + 1)$ decoding matrix $D_{(n+1)(n+1)}$ is not chosen in a purely arbitrary fashion. One obvious requirement is that of left–right symmetry, which in this context means that if the left signals to be encoded and decoded are interchanged with their right counterparts, then T_{n+1} should remain unchanged if n is even, and it should have its polarity inverted if n is odd.

In the case that the reproduction decoder matrices $R_{n_2 n_1}$ preserve the total energy of stereo signals passing through them, as is true of the decoders described in [12] and in Eqs. (6)–(11), it can be shown that the columns of the matrix $R_{n_2 n_1}$ are all of unit length and orthogonal (see [12]). In this case one can "complete" the $(n + 1) \times n$ matrix $R_{(n+1)n}$ by turning it into an $(n + 1) \times (n + 1)$ matrix by adding a last column that is also of unit length and orthogonal to the known n columns. The resulting $(n + 1) \times (n + 1)$ matrix converts the n-loudspeaker stereo loudspeaker feed signals plus an $(n + 1)$th signal T_{n+1} into $n + 1$ loudspeaker feed signals; the additional signal T_{n+1} is a suitable choice for the $(n + 1)$th transmission signal, and the resulting encoding and decoding matrices $E_{(n+1)(n+1)}$ and $D_{(n+1)(n+1)}$ can be shown to be orthogonal matrices. This choice is the same as that making the last column of $D_{(n+1)(n+1)}$ to be a unit-length vector orthogonal to its other columns.

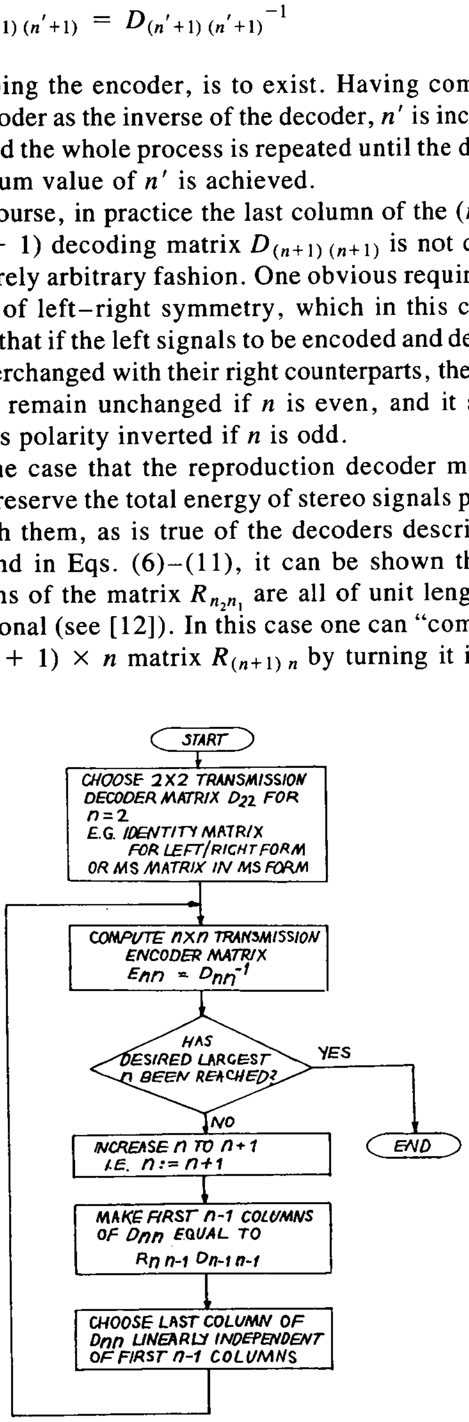

Fig. 10. Flow diagram of hierarchical system of encoding and decoding, given desired $(n + 1) \times n$ matrix reproduction decoders $R_{(n+1)n}$ in Fig. 7.

While there is no logical reason why the $n \times n$ encoding and decoding matrices of a transmission hierarchy *should* be orthogonal, we have found that in practice there is little if any advantage in departing from orthogonality.

The detailed examination of the general construction of a hierarchy according to Fig. 10 will be postponed until we have examined simple special cases that illuminate the mathematical density of these constructions.

4 MS MATRICES

It enormously simplies the description of matrixing of signals with left–right symmetry if the signals are expressed in sum-and-difference, or MS, form. MS, or sum-and-difference, concepts are far from new in stereo. They were introduced as long ago as 1931 by Blumlein [13], and are familiar both in connection with MS microphone technique [14] and in the Zenith–GE system of FM stereo multiplex broadcasting, as well as a production technique used in stereo television.

A pair L_p and R_p of mirror-symmetric left and right loudspeaker signals can be expressed in MS form as the signals

$$M_p = 2^{-1/2}(L_p + R_p) \tag{20a}$$

$$S_p = 2^{-1/2}(L_p - R_p) \tag{20b}$$

and the inverse equations are

$$L_p = 2^{-1/2}(M_p + S_p) \tag{21a}$$

$$R_p = 2^{-1/2}(M_p - S_p) . \tag{21b}$$

This describes an identical matrix. Thus the effect of two cascaded MS matrices is to restore the original signals. Loudspeaker feed signals expressed in terms of signals M_p, S_p, and C_p are said to be in MS form, those expressed in terms of L_p, R_p, and C_p are said to be in direct or left–right form.

It is easily shown that

$$L_p^2 + R_p^2 = M_p^2 + S_p^2 \tag{22}$$

so that the total signal energy in MS form is the same as that in left–right form.

We conveniently term signals of the form M_p or C_p as sum signals, and those of the form S_p as difference signals. Sum signals do not change if left and right inputs are interchanged, whereas difference signals are inverted in polarity.

These equations can, for a chosen value of the parameter $\phi = \phi'$, be converted into three-channel decoding equations suitable for use with Fig. 11 by adding an extra column orthogonal to the other two to the 3×2 matrix of Eq. (24). Thus,

$$\begin{pmatrix} L_3 \\ R_3 \\ C_3 \end{pmatrix} = \begin{pmatrix} \tfrac{1}{2}(\sin\phi' + 1) & \tfrac{1}{2}(\sin\phi' - 1) & 2^{-1/2}\cos\phi' \\ \tfrac{1}{2}(\sin\phi' - 1) & \tfrac{1}{2}(\sin\phi' + 1) & 2^{-1/2}\cos\phi' \\ 2^{-1/2}\cos\phi' & 2^{-1/2}\cos\phi' & -\sin\phi' \end{pmatrix} \begin{pmatrix} L \\ R \\ T \end{pmatrix} . \tag{25}$$

This 3×3 matrix transmission decoding equation is not only compatible with two-channel reception with $T = 0$, but is such that the inverse encoding equation

$$\begin{pmatrix} L \\ R \\ T \end{pmatrix} = \begin{pmatrix} \tfrac{1}{2}(\sin\phi' + 1) & \tfrac{1}{2}(\sin\phi' - 1) & 2^{-1/2}\cos\phi' \\ \tfrac{1}{2}(\sin\phi' - 1) & \tfrac{1}{2}(\sin\phi' + 1) & 2^{-1/2}\cos\phi' \\ 2^{-1/2}\cos\phi' & 2^{-1/2}\cos\phi' & -\sin\phi' \end{pmatrix} \begin{pmatrix} L_3 \\ R_3 \\ C_3 \end{pmatrix} \tag{26}$$

is described by exactly the same matrix as the decoding equation, Eq. (25). This pattern does not generalize to the four- or more channel case but means that the transmission encoder E_{33} and the transmission decoder D_{33} can be implemented by the same matrix circuit.

In MS form, as shown in Fig. 12, the encoding equations are

$$\begin{pmatrix} M \\ T \end{pmatrix} = \begin{pmatrix} \sin\phi' & \cos\phi' \\ \cos\phi' & -\sin\phi' \end{pmatrix} \begin{pmatrix} M_3 \\ C_3 \end{pmatrix} \tag{27a}$$

$$S = S_3 \tag{27b}$$

and the inverse transmission decoding equations are

$$\begin{pmatrix} M_3 \\ C_3 \end{pmatrix} = \begin{pmatrix} \sin\phi' & \cos\phi' \\ \cos\phi' & -\sin\phi' \end{pmatrix} \begin{pmatrix} M \\ T \end{pmatrix} \tag{28a}$$

$$S_3 = S . \tag{28b}$$

It will be seen that these equations have the simplest form in MS form rather than left–right form, something that remains true for the 4×4 and 5×5 and higher-order cases.

A problem with this transmission hierarchy is that the parameter ϕ' for the encoding cannot be made frequency-dependent without making the stereo baseband signals L and R frequency-dependent, which would degrade compatibility with two-loudspeaker reception. Thus one is forced to use a fixed value of ϕ', say 45°, whereas subjectively ideal 3×2 matrix reproduction decoders have ϕ vary between 35 and 55° with frequency as we have noted; see [12]. However, this is a less serious problem than it seems. If the receiver knows that only two transmission channels are being received, a frequency-dependent 3×2 decoder can be used. If three channels are being transmitted and received, then the 3×3 encoding matrix can be preceded by a 3×2 frequency-dependent matrix reproduction decoder

5 THREE-CHANNEL HIERARCHY

Figs. 11 and 12 show a three-channel hierarchical encoding and decoding system, where Fig. 11 illustrates the case where there are two basic channels L and R normally used for transmitting conventional two-channel stereo left and right, and T is an additional third channel. In Fig. 12 the two basic channels used for two-channel stereo are signals M and S in MS form.

The hierarchy needs to be such that the three-channel decoder D_{33} provides good three-loudspeaker reproduction of transmitted two-channel signals. As noted in some detail in [12], the general form of a left–right symmetric energy-preserving 3×2 decoder has the form

$$\begin{pmatrix} M_3 \\ C_3 \end{pmatrix} = \begin{pmatrix} \sin\phi \\ \cos\phi \end{pmatrix} [M_2] \tag{23a}$$

$$S_3 = S_2 \tag{23b}$$

where the angle parameter ϕ lies between 0 and 90°. The best value subjectively of the parameter ϕ is generally in the region of 35 to 55°, with no value giving ideal results. The psychoacoustic tradeoffs for different values of ϕ are detailed in [12]. For a frequency-independent value of ϕ, it is generally best around 45°, whereas if ϕ is permitted to depend on frequency, it has been found that $\phi = 35°$ below 5 kHz and $\phi = 55°$ above 5 kHz tend to give the best results.

Fig. 13 shows the form of a 3×2 decoder implementing Eqs. (23) in left–right form. The left–right form of the 3×2 reproduction decoding equations is

$$\begin{pmatrix} L_3 \\ R_3 \\ C_3 \end{pmatrix} = \begin{pmatrix} \tfrac{1}{2}(\sin\phi + 1) & \tfrac{1}{2}(\sin\phi - 1) \\ \tfrac{1}{2}(\sin\phi - 1) & \tfrac{1}{2}(\sin\phi + 1) \\ 2^{-1/2}\cos\phi & 2^{-1/2}\cos\phi \end{pmatrix} \begin{pmatrix} L_2 \\ R_2 \end{pmatrix} . \tag{24}$$

with angle parameter ϕ varying with frequency, as shown in Fig. 9. This will result in a three-channel receiver receiving outputs identical to those of the frequency-dependent 3×2 decoder used at the transmitter. It may be thought that the use of such a frequency-dependent reproduction decoder at the encoding stage will result in a frequency-dependent baseband two-loudspeaker signal L and R, but this frequency dependence is actually negligible, since by Eqs. (23a) and (27a) we have

$$\begin{aligned} M &= (\sin \phi \sin \phi' + \cos \phi \cos \phi')M_2 \\ &= [\cos(\phi - \phi')]M_2 \end{aligned} \tag{29a}$$

and from Eqs. (23b) and (27b),

$$S = S_2 \tag{29b}$$

so that provided that $|\phi - \phi'|$ is less than, say, 15°, the gain alteration $\cos(\phi - \phi')$ of the M_2 signal is negligible. For a transmission parameter $\phi' = 45°$ and a frequency-dependent reproduction decoder with ϕ varying between 35 and 55°, the gain of M_2 varies between 1 and $\cos 10° = 0.9848$, which causes a stereo crosstalk of less than −42 dB.

Thus the use of a 3×3 transmission system according to Eqs. (25)–(28) with a fixed value of ϕ' is compatible with the use in a three-loudspeaker mix of optimally 3×2 decoded two-channel program sources without significant effect on stereo compatibility. The transmitted third channel T when a frequency-dependent 3×2 reproduction decoder is used before the 3×3 transmission encoder is given from Eqs. (23a) and (27a) by

$$T = \sin(\phi - \phi')M_2 \ . \tag{29c}$$

In order to avoid any disturbance of compatibility, we can replace the $\cos(\phi - \phi')$ factor in Eq. (29a) by 1, dividing the sine factor in Eq. (29c) by it so that the balance between M and T is not altered. This results in an approximate three-channel frequency-dependent transmission of two-loudspeaker stereo given by the equations

$$\begin{aligned} M &= M_2 \\ S &= S_2 \\ T &= [\tan(\phi - \phi')]M_2 \end{aligned} \tag{30}$$

which can be implemented as shown in Fig. 14, where the gain has a value $\tan 10° = 0.176$ and the all-pass

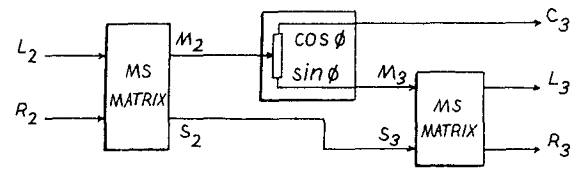

Fig. 13. Energy-preserving 3×2 matrix reproduction decoder using sine–cosine gain between MS matrices.

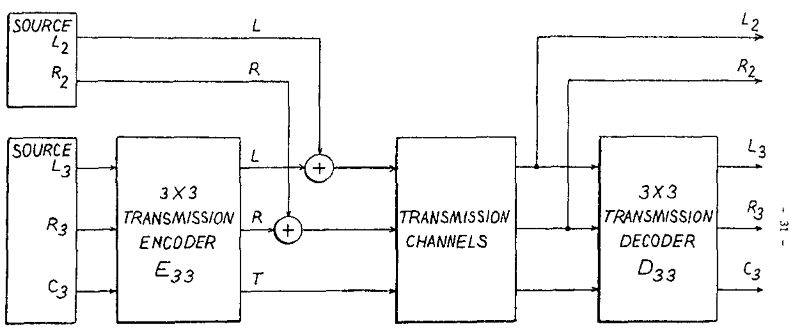

Fig. 11. Three-channel transmission encoding–decoding hierarchy using left–right transmission form.

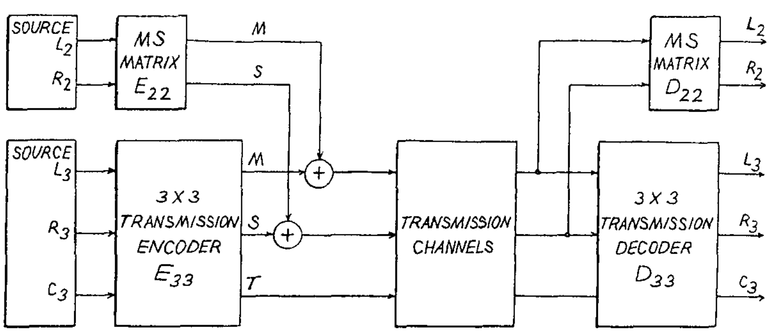

Fig. 12. Three-channel transmission encoding–decoding hierarchy using MS transmission form.

network has gain -1 at frequencies below 5 kHz and $+1$ above 5 kHz. With the transmission decoder D_{33} for $\phi' = 45°$, this implements an optimized 3×2 decoded result from a two-channel input.

These 3×3 transmission encoding and decoding equations have a reasonable compatibility when three-channel signals are received via two loudspeakers, with a slight out-of-phase crosstalk on extreme left and right sounds (about -15.3 dB for $\phi' = 45°$) and a slight boost of edge images as compared to middle images in stereo—and a cut of -3 dB of edge-of-stage sounds in mono, similar to that occurring with mono reception of two-loudspeaker stereo. The extra width of two-loudspeaker reception of three-channel transmissions reflects the slightly greater width reported in [12] of three-loudspeaker presentation of three-channel material as compared to three-loudspeaker presentation of optimally decoded two-loudspeaker material.

Attempts to find a better two-loudspeaker presentation of three-channel material by using a nonorthogonal hierarchy do not work. A nonorthogonal three-channel transmission hierarchy is obtained by replacing Eq. (28a) by

$$\begin{pmatrix} M_3 \\ C_3 \end{pmatrix} = \begin{pmatrix} \sin\phi' & \cos\phi'' \\ \cos\phi' & -\sin\phi'' \end{pmatrix} \begin{pmatrix} M \\ T \end{pmatrix} \tag{31}$$

in the decoding equations with the inverse encoding equations having the form

$$\begin{pmatrix} M \\ T \end{pmatrix} = \frac{1}{\cos(\phi' - \phi'')} \begin{pmatrix} \sin\phi'' & \cos\phi'' \\ \cos\phi' & -\sin\phi' \end{pmatrix} \begin{pmatrix} M_3 \\ C_3 \end{pmatrix} \tag{32}$$

with $S_3 = S$ as before. The extra parameter ϕ'' does not affect the three-loudspeaker decoding of the two baseband channels M and S or L and R, but it does affect two-loudspeaker reception of three-loudspeaker transmissions. Making ϕ'' differ from ϕ', however, does not offer any overall improvement in 3-to-2 compatibility, although we omit the details here.

6 GENERAL ORTHOGONAL HIERARCHY

The preceding three-channel case can be extended to the general case. In this section we confine attention to signals expressed in MS form in order to keep the math as simple as possible. In [12] we showed that the general energy-preserving 4×3 matrix reproduction decoder satisfied the equations

$$\begin{pmatrix} M_4 \\ M_5 \end{pmatrix} = \begin{pmatrix} \cos\phi_3 & -\sin\phi_3 \\ \sin\phi_3 & \cos\phi_3 \end{pmatrix} \begin{pmatrix} M_3 \\ C_3 \end{pmatrix} \tag{33a}$$

$$\begin{pmatrix} S_4 \\ S_5 \end{pmatrix} = \begin{pmatrix} \cos\phi_D \\ \sin\phi_D \end{pmatrix} (S_3) \tag{33b}$$

where ϕ_3 and ϕ_D are angle parameters which, for the decoder of Eq. (7), were $\phi_3 = 10.57°$ and $\phi_D = 28.64°$, values which were shown in [12] to preserve stereo effect.

Combining Eqs. (33) with Eqs. (28) gives

$$\begin{pmatrix} M_4 \\ M_5 \end{pmatrix} = \begin{pmatrix} \sin(\phi' - \phi_3) & \cos(\phi' - \phi_3) \\ \cos(\phi' - \phi_3) & -\sin(\phi' - \phi_3) \end{pmatrix} \begin{pmatrix} M \\ T \end{pmatrix} \tag{34a}$$

$$\begin{pmatrix} S_4 \\ S_5 \end{pmatrix} = \begin{pmatrix} \cos\phi_D \\ \sin\phi_D \end{pmatrix} (S) . \tag{34b}$$

Adding a fourth channel to these equations while ensuring orthogonality of the matrices changes Eq. (34b) into

$$\begin{pmatrix} S_4 \\ S_5 \end{pmatrix} = \begin{pmatrix} \cos\phi_D & \sin\phi_D \\ \sin\phi_D & -\cos\phi_D \end{pmatrix} \begin{pmatrix} S \\ T_4 \end{pmatrix} \tag{34c}$$

and Eqs. (34a) and (34c) are the orthogonal matrix 4×4 transmission decoding equations for a four-channel hierarchy, where Eqs. (27) and (28) are the three-channel decoding and encoding equations. The inverse 4×4 transmission encoding equations to Eqs. (34a) and (34c) are given by

$$\begin{pmatrix} M \\ T \end{pmatrix} = \begin{pmatrix} \sin(\phi' - \phi_3) & \cos(\phi' - \phi_3) \\ \cos(\phi' - \phi_3) & -\sin(\phi' - \phi_3) \end{pmatrix} \begin{pmatrix} M_4 \\ M_5 \end{pmatrix} \tag{35a}$$

$$\begin{pmatrix} S \\ T_4 \end{pmatrix} = \begin{pmatrix} \cos\phi_D & \sin\phi_D \\ \sin\phi_D & -\cos\phi_D \end{pmatrix} \begin{pmatrix} S_4 \\ T_4 \end{pmatrix} . \tag{35b}$$

It will be noted that, in MS form, the matrix encoding and decoding equations for four-loudspeaker stereo are identical; this is not true in left–right form.

The details for the five-channel hierarchy case are more complicated, and we only summarize the results here. In [12], the 5×4 energy-preserving reproduction decoder was noted to have equations in MS form,

$$\begin{pmatrix} M_6 \\ M_7 \\ C_5 \end{pmatrix} = 2^{-1/2} \begin{pmatrix} a + \mu_1 \cos\phi_4 + \nu_1 \sin\phi_4 & a - \mu_1 \cos\phi_4 - \nu_1 \sin\phi_4 \\ b - \mu_2 \cos\phi_4 + \nu_2 \sin\phi_4 & b + \mu_2 \cos\phi_4 - \nu_2 \sin\phi_4 \\ c - \lambda \sin\phi_4 & c + \lambda \sin\phi_4 \end{pmatrix} \begin{pmatrix} M_4 \\ M_5 \end{pmatrix} \tag{36a}$$

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

$$\lambda = (a_2 + b_2)^{1/2}, \quad \mu_1 = \frac{b}{\lambda}, \quad \mu_2 = \frac{a}{\lambda},$$

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

and

$$\begin{pmatrix} S_6 \\ S_7 \end{pmatrix} = \begin{pmatrix} \cos \phi_5 & -\sin \phi_5 \\ \sin \phi_5 & \cos \phi_5 \end{pmatrix} \begin{pmatrix} S_4 \\ S_5 \end{pmatrix} \tag{36b}$$

ϕ_5 being another angle parameter. For a 5×4 decoder that preserves stereo effect it was shown in [12] that $\phi_4 = 51.64°$, $\phi_5 = 9.64°$, and $(a, b, c) = (0.6164, 0.6558, 0.4359)$, which is the basis of the 5×4 reproduction decoder matrix of Eq. (8). The orthogonal five-channel hierarchy derived from these equations by adding an extra column orthogonal to the others in Eq. (36a) can be shown to have the following form:

$$\begin{pmatrix} M_6 \\ M_7 \\ C_5 \end{pmatrix} = A \begin{pmatrix} \sin(\phi' - \phi_3) & \cos(\phi' - \phi_3) & 0 \\ \cos(\phi' - \phi_3) & -\sin(\phi' - \phi_3) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} M \\ T \\ T_5 \end{pmatrix} \tag{37a}$$

where the orthogonal 3×3 matrix A has the form

$$A = 2^{-1/2} \begin{pmatrix} a + \mu_1 \cos \phi_4 + \nu_1 \sin \phi_4 & a - \mu_1 \cos \phi_4 - \nu_1 \sin \phi_4 & 2^{1/2}[\nu_1 \cos \phi_4 - \mu_1 \sin \phi_4] \\ b - \mu_2 \cos \phi_4 + \nu_2 \sin \phi_4 & b + \mu_2 \cos \phi_4 - \nu_2 \sin \phi_4 & 2^{1/2}[\nu_2 \cos \phi_4 + \mu_2 \sin \phi_4] \\ c - \lambda \sin \phi_4 & c + \lambda \sin \phi_4 & -2^{1/2} \lambda \cos \phi_4 \end{pmatrix}$$

where the parameters a, b, c, ϕ_4, λ, μ_1, μ_2, ν_1, and ν_2 are as in Eq. (36a) and

$$\begin{pmatrix} S_6 \\ S_7 \end{pmatrix} = \begin{pmatrix} \cos(\phi_D + \phi_5) & \sin(\phi_D + \phi_5) \\ \sin(\phi_D + \phi_5) & -\cos(\phi_D + \phi_5) \end{pmatrix} \begin{pmatrix} S \\ T_4 \end{pmatrix} . \tag{37b}$$

The encoding equations inverse to the transmission decoding equations, Eqs. (37), have the form

$$\begin{pmatrix} M \\ T \\ T_5 \end{pmatrix} = \begin{pmatrix} \sin(\phi' - \phi_3) & \cos(\phi' - \phi_3) & 0 \\ \cos(\phi' - \phi_3) & -\sin(\phi' - \phi_3) & 0 \\ 0 & 0 & 1 \end{pmatrix} A^T \begin{pmatrix} M_6 \\ M_7 \\ C_5 \end{pmatrix} \tag{38a}$$

where A^T is the 3×3 matrix transpose (a_{ji}) of the 3×3 matrix $A = (a_{ij})$ and

$$\begin{pmatrix} S \\ T_4 \end{pmatrix} = \begin{pmatrix} \cos(\phi_D + \phi_5) & \sin(\phi_D + \phi_5) \\ \sin(\phi_D + \phi_5) & -\cos(\phi_D + \phi_5) \end{pmatrix} \begin{pmatrix} S_6 \\ S_7 \end{pmatrix} . \tag{38b}$$

Eqs. (27), (28), (34a), (34c), (35), (37), and (38) together fully define a transmission encoding and decoding hierarchy, giving optimum psychoacoustic reproduction for n_1-loudspeaker stereo signals received via n_2-loudspeakers according to [12], when the values of the various parameters suggested are used.

7 CONCLUSIONS

In this paper we have constructed a hierarchical system for the transmission and reception of multiloudspeaker frontal-stage stereo using any number of loudspeakers up to five such that any numbers of transmission channels from two to five can be used in a mutually compatible way, and in particular such that reception via a larger number of channels than the number of loudspeakers the program was originated for gives the best possible psychoacoustic reproduction of the originally intended effect. The proposed hierarchy, whose explicit equations are given in the Appendix based on the values (5) for the loudspeaker layout angles, varies only slightly for different choices of loudspeaker layout, as shown in [12].

We have optimized upward compatibility, but did not investigate downward compatibility in detail. The downward compatibility of the hierarchy is not perfect, but is generally acceptable. Investigations not detailed here show that departing from the orthogonal matrix hierarchy does not improve downward compatibility, so that we conclude that it is impossible to design a hierarchy with perfect upward and downward compatibility. Downward compatibility is less well defined than upward compatibility, since there are many ways of panning multiloudspeaker stereo positions (see [15], [16]), and each method gives different downward compatibility performance. Given that listeners using a larger number of loudspeakers will be used to better results than those using few loudspeakers, we feel that excellent upward compatibility is important.

We hope to discuss downward compatibility in more

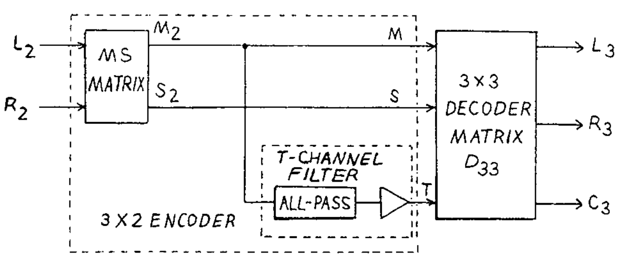

Fig. 14. Frequency-dependent 3×2 reproduction decoder consisting of 3×2 transmission encoder followed by 3×3 transmission decoder.

detail in a future paper. However, the basic presentation of reception of n_1-loudspeaker stereo via fewer loudspeakers is still quite good by the nonoptimized standards hitherto accepted in multiloudspeaker stereo. Moreover, the hierarchical approach given here ensures that repeated encoding and decoding via different numbers of transmission and loudspeaker channels will never suffer more degradation than that implied by the smallest "bottleneck" number in the chain, so that the proposed hierarchy is operationally robust.

The hierarchy proposed gives greater operational freedom in processing, reprocessing, and interfacing signals originated for different numbers of loudspeakers than approaches not based on a psychoacoustically optimized transmission hierarchy. For systems having other sound stages besides the front stage, such as proposed for HDTV [1], [3]–[5], two hierarchies can be used, one for a frontal stage and one for a rear stage. However, optimized surround sound might usefully take on board aspects of Ambisonic hierarchies [17], [18] in combination with the work of the present paper.

As far as we are aware, the work in the present paper is the first application of theoretical methods in directional psychoacoustics to the optimization of systems of transmission of frontal-stage multiloudspeaker stereo. This case is much more difficult than the surround-sound case [8]–[11], [17], but the mathematical complications lead to usable equations and methods, such as summarized in the Appendix.

9 REFERENCES

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

[2] 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.).

[3] G. Theile, "On the Performance of Two-Channel and Multi-Channel Stereophony," presented at the 80th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 38, p. 379 (1990 May), preprint 2887.

[4] G. Theile, "HDTV Sound Systems: How Many Channels?," presented at the AES/SMPTE Joint Conf. "Television Sound Today and Tomorrow," Detroit, MI (1991 Feb. 1–2).

[5] D. J. Meares, "High Definition Sound for High Definition Television," presented at the AES/SMPTE Joint Conf. "Television Sound Today and Tomorrow," Detroit, MI (1991 Feb. 1–2).

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

[7] W. B. Snow, "Basic Principles of Stereophonic Sound," *J. SMPTE*, vol. 61, pp. 567–589 (1953 Nov.); reprinted in [6].

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

[9] D. H. Cooper, "QFMX—Quadruplex FM Transmission Using the 4–4–4 QMX Matrix System," *J. Audio Eng. Soc.*, vol. 22, pp. 82–87 (1974 Mar.).

[10] M. A. Gerzon, "Periphony: With-Height Sound Reproduction," *J. Audio Eng. Soc.*, vol. 21, pp. 2–10 (1973 Jan./Feb.).

[11] J. J. Gibson, R. M. Christensen, and A. L. R. Limberg, "Compatible FM Broadcasting of Panoramic Sound," *J. Audio Eng. Soc.*, vol. 20, pp. 816–822 (1972 Dec.).

[12] M. A. Gerzon, "Optimum Reproduction Matrices for Multispeaker Stereo," *J. Audio Eng. Soc.*, vol. 40, pp. 571–589 (1992 July/Aug.).

[13] A. D. Blumlein, U. K. patent 394,325, filed 1931 Dec. 14; reprinted in [6].

[14] G. Bore and S. F. Temmer, " 'M-S' Stereophony and Compatibility," *Audio Mag.*, p. 19 (1958 Apr.); reprinted in [6].

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

[16] 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.

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

[18] M. A. Gerzon, "Hierarchical System of Surround Sound Transmission for HDTV," presented at the 92nd Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 40, p. 445 (1992 May), preprint 3339.

8 PATENT NOTE

Some of the methods reported in this paper are the subject of patent applications by the author.

APPENDIX A NUMERICAL ORTHOGONAL HIERARCHY EQUATIONS

The orthogonal matrix transmission hierarchy based on the $n_2 \times n_1$ matrix reproduction decoders listed in Eqs. (6)–(11) are given in explicit numerical form in this appendix. These equations express how the loudspeaker feed signals are related to the transmission signals L, R, T, T_4, and T_5.

1) Mono:

Encoding equation

$$L = R = 2^{-1/2} C_1 .$$

Decoding equation

$$C_1 = 2^{-1/2}(L + R) .$$

2) Two-loudspeaker stereo:

Encoding equations

$$L = L_2$$

$$R = R_2 \, .$$

Decoding equations

$$L_2 = L$$

$$R_2 = R \, .$$

3) Three-loudspeaker stereo:
Encoding equations

$$\begin{pmatrix} L \\ R \\ T \end{pmatrix} = \begin{pmatrix} 0.8536 & -0.1464 & 0.5000 \\ -0.1464 & 0.8536 & 0.5000 \\ 0.5000 & 0.5000 & -0.7071 \end{pmatrix} \begin{pmatrix} L_3 \\ R_3 \\ C_3 \end{pmatrix} .$$

Decoding equations

$$\begin{pmatrix} L_3 \\ R_3 \\ C_3 \end{pmatrix} = \begin{pmatrix} 0.8536 & -0.1464 & 0.5000 \\ -0.1464 & 0.8536 & 0.5000 \\ 0.5000 & 0.5000 & -0.7071 \end{pmatrix} \begin{pmatrix} L \\ R \\ T \end{pmatrix} .$$

4) Four-loudspeaker stereo:
Encoding equations

$$\begin{pmatrix} L \\ R \\ T \\ T_4 \end{pmatrix} = \begin{pmatrix} 0.7215 & -0.1561 & 0.6521 & 0.1728 \\ -0.1561 & 0.7215 & 0.1728 & 0.6521 \\ 0.5832 & 0.5832 & -0.3998 & -0.3998 \\ 0.3389 & -0.3389 & -0.6206 & 0.6206 \end{pmatrix} \begin{pmatrix} L_4 \\ R_4 \\ L_5 \\ R_5 \end{pmatrix} .$$

Decoding equations

$$\begin{pmatrix} L_4 \\ R_4 \\ L_5 \\ R_5 \end{pmatrix} = \begin{pmatrix} 0.7215 & -0.1561 & 0.5832 & 0.3389 \\ -0.1561 & 0.7215 & 0.5832 & -0.3389 \\ 0.6521 & 0.1728 & -0.3998 & -0.6206 \\ 0.1728 & 0.6521 & -0.3998 & 0.6206 \end{pmatrix} \begin{pmatrix} L \\ R \\ T \\ T_4 \end{pmatrix} .$$

5) Five-loudspeaker stereo:
Encoding equations

$$\begin{pmatrix} L \\ R \\ T \\ T_4 \\ T_5 \end{pmatrix} = \begin{pmatrix} 0.6325 & -0.1525 & 0.6482 & 0.0287 & 0.3945 \\ -0.1525 & 0.6325 & 0.0287 & 0.6482 & 0.3945 \\ 0.5570 & 0.5570 & -0.0373 & -0.0373 & -0.6138 \\ 0.4381 & -0.4381 & -0.5551 & 0.5551 & 0.0000 \\ -0.2730 & -0.2730 & 0.5191 & 0.5191 & -0.5585 \end{pmatrix} \begin{pmatrix} L_6 \\ R_6 \\ L_7 \\ R_7 \\ C_5 \end{pmatrix} .$$

Decoding equations

$$\begin{pmatrix} L_6 \\ R_6 \\ L_7 \\ R_7 \\ C_5 \end{pmatrix} = \begin{pmatrix} 0.6325 & -0.1525 & 0.5570 & 0.4381 & -0.2730 \\ -0.1525 & 0.6325 & 0.5570 & -0.4381 & -0.2730 \\ 0.6482 & 0.0287 & -0.0373 & -0.5551 & 0.5191 \\ 0.0287 & 0.6482 & -0.0373 & 0.5551 & 0.5191 \\ 0.3945 & 0.3945 & -0.6138 & 0.0000 & -0.5585 \end{pmatrix} \begin{pmatrix} L \\ R \\ T \\ T_4 \\ T_5 \end{pmatrix} .$$

The effect of encoding n_1-loudspeaker stereo into L, R, T, T_4, and T_5, putting unused transmission channels equal to zero and decoding to a greater number n_2 of loudspeakers, is given by the $n_2 \times n_1$ reproduction decoding equations (6)–(11) in the main text. For $n_2 < n_1$, the equations relating signals are the same as Eqs. (6)–(11), except that the transpose matrix equations are used, that is, the $n_2 \times n_1$ matrix m_{ij} is replaced by the $n_1 \times n_2$ matrix m_{ji}. This transpose property is a consequence of the equations using orthogonal matrices, and does not generalize to the nonorthogonal case.

The foregoing equations can be converted to MS transmission form using Eqs. (18) and (19) of the main text.

The biography for Michael Gerzon was published in the 1992 July/August issue of this *Journal*.