Discrete-Matrix Multichannel Stereo*

DUANE H. COOPER

University of Illinois, Urbana, Ill.

AND

TAKEO SHIGA

Nippon Columbia Company, Ltd., Kawasaki, Japan

Azimuthal harmonic synthesis generates a universe of matrices (UMX) with maximal input–output azimuthal correlations for two (BMX), three (TMX), four (QMX), etc., intermediary channels. The basic mono–stereo-compatible BMX accepts augmentation channels converting to TMX, QMX, etc. In quadrasonics, the psychoacoustically discrete TMX essentially equals the fully discrete QMX. The feasibility of multichannel disc, broadcast, and tape is profoundly enhanced by the sufficiency of using 2-kHz-wide augmentation channels.

INTRODUCTION: The multichannel matrix problem is the problem of devising a signal-mixing scheme whereby a number of signal sources, say 72, may be mixed to form a smaller number of transmission signals, say 4, to make a loudspeaker presentation, using yet another number of loudspeakers, say 6. This may be described as a 72–4–6 matrix problem, where, of course, any other numbers for the numbers of source–transmission–presentation channels could be specified. A popular example is the 4–2–4 matrix problem, in which the four sources are assumed to have been already prepared, perhaps by solving the 72–16–4 problem (72 instrumentalists mixed for recording on a 16-track tape machine and then remixed to the 4-channel quadrasonic format). Whether realizing any of these mixing schemes actually poses any problem must depend, of course, upon specific goals, in addition to the above, to be achieved by the process.

The 4–2–4 matrix problem has been extensively discussed [1]–[8]. It is at once obvious that the four presentation signals cannot be independent replicas of the corresponding four source signals, because linear combinations of two quantities cannot provide four independent quantities. Thus, a signal appearing in only one of the four source channels must appear to some extent in at least three of the four presentation channels. The ratios of its values in the various channels as a ratio of desired-to-undesired values are the channel separation figures, and the various schemes have differed in the patterning of these figures, with particular attention paid to their magnitudes and lesser attention paid to their phase. For one scheme, novel psychoacoustic theories have been advanced to explain the consequences of the undesired or side-effect signals [4]. Satisfactory localiza-

* Presented October 7, 1971, at the 41st Convention of the Audio Engineering Society, New York, under the title, "Multichannel Matrix Encoding."

tions may be obtained, however, with the use of gain-riding circuits called "logic" that have been devised to diminish, for one-at-a-time source signals, the appearance of these side effects [1], [4]. The avoidance of an audible "pumping" of more or less continuous signals by the time-varying appearance of other signals is an important problem in this "logic."

From the variety of schemes proposed, it is clear that there is a widespread disagreement as to the fundamental design principles, although a number of desirable overall properties have been listed [3]. It has been generally recognized, however, that the presentation devices (loudspeakers) are to be regarded as having more or less fixed locations with respect to the listener, and that the goal is to make an acoustical display that the ear will accept as a kind of an acoustical panorama in azimuth, corresponding as closely as possible to an original azimuthal distribution of sources. If, among these sources, there be included reverberation sources, then direct-to-reverberant ratios, hall sounds (ambiance), etc., will also be reproduced to satisfy various psychoacoustically desired qualities [9].

It is the purpose of this paper to develop matrix design principles whereby the azimuthal information represented by the original source distribution may be presented with minimal loss. The minimum loss is ensured by generating a source function of limited azimuthal harmonic content, which source function exhibits maximal correlation with the original source distribution. The limitation in azimuthal harmonic content provides a guarantee that a finite number of azimuthal samples will faithfully represent that source function. These azimuthal samples then provide the signals to excite the loudspeakers, using however many interpolated samples as may be needed to satisfy the ear.

This maximal correlation, the engineering measure of azimuthal fidelity, provides the basis for the universe of matrix (UMX) systems, particular embodiments of which are the quadrasonic, two-channel matrix (BMX), three-channel matrix (TMX), and four-channel matrix (QMX). These are optimal in transmitting the maximum in azimuthal information for the number of channels used, and they all exhibit perfect localization patterns.

The design of these matrices is explained in detail. Experimental results showing that "discreteness" begins with the TMX are quoted. Further experiments, showing how the frequency response of some of the channels may be restricted, are explained, together with some of the system implications for cassette, FM, and disc media. In particular, it is shown that discrete-matrix multichannel stereo need not make excessively severe demands on current audio technology in these media.

The explanation of principles begins with a discussion of azimuthal mixing.

AZIMUTHAL MIXING

The keyword appears to be "azimuth." It is widely accepted that the loudspeakers shall be placed according to azimuthal designation, LF, LB, RF, and RB being the favorite designations, in a listening space for which the cardinal points in azimuth would be F, R, B, and L for front, right, back, and left (sometimes with the adjective "center," as in "center front" or "center right," etc., meaning "true" or "due" front, or right, etc.).

One way in which mixing (matrixing) may be done to derive a signal for the RF loudspeaker is shown in Fig. 1. A directional microphone with sensitivity curve M is aimed at the RF azimuth. Then, a source at LF will appear in the RF transmission channel with intensity T_{RF} which is the product MS_{LF}, where S_{LF} is the source strength and M is the value of the microphone sensitivity in the LF direction when pointed at RF, the height of the M curve at the LF location. Other sources at other locations will similarly excite this transmission channel in proportion to the plotted M values and the total T_{RF} signal will be the sum of these excitations.

The curve M actually plotted in Fig. 1, it is to be understood, is purely arbitrary. Practical microphones with a null response extending over 90° of azimuth are unknown. Also, the height of the plotted point S_{LF} is intended to indicate the source intensity, either as an instantaneous source-waveform value or as the value of a frequency component of the source, although such time or frequency structure cannot be indicated in this azimuthal plot. Generally, however, frequency components, characterized by magnitude and phase, will be so designated, although magnitude and phase cannot both be plotted so simply.

In addition to microphone mixing as in Fig. 1, mixing (matrix) circuits may be used according to the plan of Fig. 2. There a signal is thought to be derived directly from the source, "on-mic" pickup, for example, and the mixing is to be done in a circuit. Then M is the mixing curve, now centered on the source S_{LF}, and the height of the curve at RF specifies the proportion of S_{LF} that is to appear (added to corresponding amounts from other sources) in channel T_{RF}. Again, the curve M actually plotted is arbitrary and merely illustrative.

The extension of this circuit concept to the mixing of three sources is shown in Fig. 3. At each source location there is centered the mixing-specification curve (M of Fig. 2), shown dashed. The point-by-point sum of these curves is the solid-line curve $S(\theta)$. The height of this curve at any azimuth, at RF, for example, thus represents the sum of the source signals (in the proportions specified by the various translated M curves) that have been formed by mixing for transmission to the speaker (if

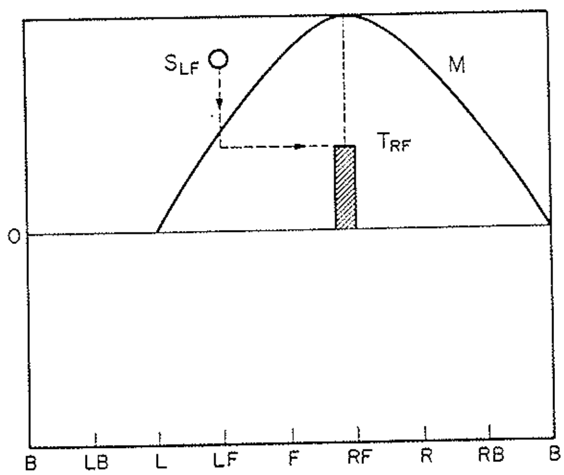

Fig. 1. Mixing (matrix) coefficient determined by microphone sensitivity curve M as if for a microphone directed for RF pickup, but detecting a source at LF.

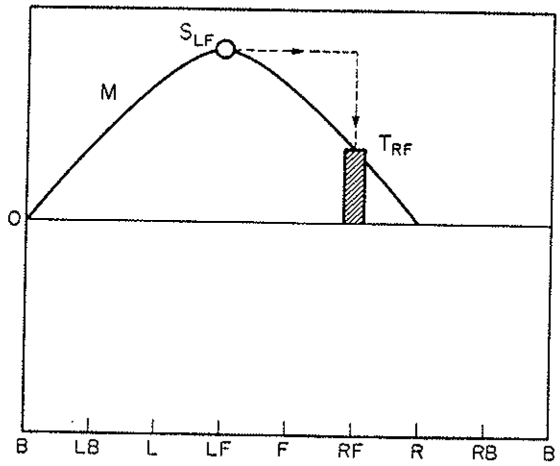

Fig. 2. Electrical mixing (matrix) coefficient specified by curve M whereby part of the source signal LF is mixed into the channel for RF.

any) for that azimuth. In this paper azimuth angles are to be measured, usually, counterclockwise from R.

In preparing Fig. 3, several important properties of the mixing-specification curve M have been illustrated. First, the curve is continuous, since it is necessary for a mixing coefficient to be specified for every possible source azimuth and loudspeaker, or pickoff, azimuth, and this coefficient must not to jump abruptly in value for a small change in azimuth. Second, the curve for each source is obtained simply by a translational motion of the curve, without changing its shape, to center it upon that source. This ensures that the treatment of each source, in its overall contribution, will be azimuth independent (azimuthal symmetry). Third, the curve M is periodic, since the azimuthal space shown is periodic; Fig. 3 may be regarded as formed from a cylinder that had been cut along the vertical line at B and unrolled to lie flat. Thus, in sliding M to be centered on the source between B and LB, this same curve must be allowed to extend beyond B without change, as shown on the right-hand side of Fig. 3.

A fourth and most important property of M relates to the fact that it has a principal maximum and a certain breadth about that maximum. Thus, if the pickoff azimuth happens to agree with a source azimuth, that source makes its maximum contribution to the transmitted signal. (It is possible, of course, for subsidiary maxima, not shown, to appear in M.) The breadth of the M curve is necessary for coverage, so that every source may contribute to at least one of the several transmission signals that are derived at differing azimuths. Hereafter, the M curve will be designated by the function $M(\theta)$ assumed to have a maximum at $\theta = 0$, for the untranslated version of M. When translated to have a maximum at θ_0, the curve is designated $M(\theta - \theta_0)$.

These properties of $M(\theta)$ determine the properties of the interpolated or mixed source pattern $S(\theta)$ in its relation to the actual source pattern. For example, $S(\theta)$ is guaranteed to be periodic. However, $S(\theta)$ is a limited-resolution representation of the source pattern in that it does not show a narrow bump at each source location, and its value at each source location does not exactly agree with the source value. Because of this limited resolution, $S(\theta)$ only approximates the actual source pattern. Of course, this limitation in resolution is a consequence of the need for coverage on the part of the $M(\theta)$ curve. Thus it is seen that coverage and resolution are mutually incompatible requirements.

The tradeoff between resolution and coverage may be managed by an optimization procedure whose outcome will depend upon the particular number of transmission channels that may be made available. The optimization criterion is to be the maximizing of the correlation between the actual azimuthal source pattern and the interpolated source pattern $S(\theta)$, since it is this correlation that is taken to be the engineering measure of azimuthal fidelity.

AZIMUTHAL OPTIMIZATION

There are actually two equivalent approaches to azimuthal optimization. In one, the difference between the source value S_i and the corresponding interpolated-source-curve value $S(\theta_i)$ is to be minimized in mean square value, where θ_i is the azimuth of the ith source. Thus it is the average value of

$$|S_i - S(\theta_i)|^2 \tag{1}$$

that is to be minimum, where the vertical bars denote magnitude without regard to phase, and the average is to be taken with respect to all possible azimuthal distributions. In expanding the squaring operation indicated in Eq. (1), however, there are computed two positive terms, $|S_i|^2$ and $|S(\theta_i)|^2$, together with a certain negative term. It is this negative term that, if maximized, achieves the minimum for Eq. (1) in the average. It is this negative term that is identified with the correlation

$$R = E\{S_i S^*(\theta_i)\} \tag{2}$$

in which E denotes the expectation or averaging operator, and the asterisk denotes the complex conjugate. (The conjugate of a complex quantity, or phasor, may be obtained by merely reversing the polarity of the phase.) It is the adjustment of $M(\theta)$, the mixing-specification curve, that is to be the basis for maximizing Eq. (2).

Of course, there is one trivial means of increasing the value of R in Eq. (2), namely, merely increasing the magnitude scale factor for $M(\theta)$ and thus for $S(\theta)$. To

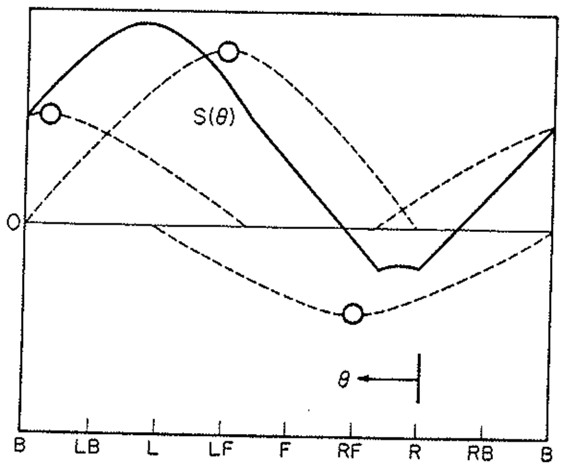

Fig. 3. Superposition of mixing curves resulting in interpolated source curve S, depending upon azimuth angle θ measured counterclockwise from RF.

guard against that, the maximizing is to be done for the total average energy

$$\mathcal{E} = \frac{1}{2\pi} E\{\int_0^{2\pi} |S(\theta)|^2 d\theta\} \tag{3}$$

being held constant.

The details are worked out in Appendix I, where it is shown that, for M, the consequence is the maximizing of the value of $M(0)$ subject to the constraint that the area under the curve $|M(\theta)|^2$ is to be constant. There is, however, a further constraint to be imposed because of the finite number of channels to be made available. This constraint imposes a limitation in azimuthal harmonic content, a subject to be examined next.

AZIMUTHAL HARMONIC CONTENT

Since the function $S(\theta)$ is guaranteed to be a periodic function in its dependence upon the azimuthal variable, it may always be represented by a Fourier series, either in trigonometric form,

$$\begin{aligned} S(\theta) = a_0 + a_1 \cos\theta + a_2 \cos 2\theta + \ldots \\ + b_1 \sin\theta + b_2 \sin 2\theta + \ldots \end{aligned} \tag{4}$$

or in complex exponential (phasor) form,

$$\begin{aligned} S(\theta) = a_0 + c_1 \exp(j\theta) + c_2 \exp(j2\theta) + \ldots \\ + c_{-1} \exp(-j\theta) + c_{-2} \exp(-j2\theta) + \ldots \end{aligned} \tag{5}$$

in which the c_n are also complex (i.e., phasors) having both a magnitude and phase angle. The c_n may be computed in terms of the a_n and b_n with the help of the well-known formula $\exp(jx) = \cos x + j \sin x$ and using the properties of the so-called imaginary unit j, obeying $j^2 = -1$, and having unit magnitude and 90° phase. Although Eqs. (4) and (5) are fully equivalent if an odd number of terms is used, it is Eq. (5) that must be used for the present purposes in which the number of terms may be even.

The complex exponentials, $\exp(j\theta)$, $\exp(j2\theta)$, $\exp(-j\theta)$, etc. are called azimuthal harmonic functions when θ represents the azimuth angle. Of course, $\exp(j\theta)$ is merely a phasor of unit magnitude and phase angle θ, and $A \exp(j\theta)$ is often written

$$A \exp(j\theta) = A e^{j\theta} = A \angle\theta \tag{6}$$

in older notation, while $\exp(-j\theta)$ is the conjugate of $\exp(j\theta)$, i.e.,

$$[\exp(j\theta)]^* = \exp(-j\theta). \tag{7}$$

The function $\exp(jn\theta)$ is an azimuthal harmonic function of harmonic number n, and, in Eq. (5), the corresponding Fourier-series coefficient c_n is the nth harmonic component, where n would be an integer.

Because of the Fourier-series theorem [10], the tabulation of a list of coefficient values c_n is fully equivalent to a specification of a periodic function $S(\theta)$, and this tabulation is a specification of the azimuthal harmonic content of that function. A finite list is said to specify a periodic function of limited harmonic content. Thus,

$$1 + \cos\theta = 1 + \tfrac{1}{2} \exp(j\theta) + \tfrac{1}{2} \exp(-j\theta) \tag{8}$$

is of limited harmonic content.

A need for limiting the harmonic content arises in azimuthal sampling.

AZIMUTHAL SAMPLING

The list of values of $S(\theta)$ taken at N equally spaced angles, namely, $S(\theta_0), S(\theta_1), S(\theta_2), \cdots, S(\theta_{N-1})$, where $S(\theta_N) = S(\theta_0)$, are said to be the periodic samples of $S(\theta)$. It is a remarkable fact (known as the azimuthal sampling theorem, demonstrated in Appendix II) that these N samples are sufficient to completely specify $S(\theta)$, provided that the harmonic content of $S(\theta)$ is limited to at most N azimuthal harmonic components. What this means is that an interpolating function $M(\theta)$ may be found such that

$$\begin{aligned} S(\theta) = S(\theta_0) M(\theta-\theta_0) + S(\theta_1) M(\theta-\theta_1) \\ + \cdots + S(\theta_{N-1}) M(\theta-\theta_{N-1}) \end{aligned} \tag{9}$$

provided that $S(\theta)$ obeys requisite harmonic limitation.

The implication of this azimuthal sampling theorem for the present application is that the interpolated source Pattern, derived for maximum correlation with the actual source pattern, may be fully represented by its N azimuthal-sample values, and faithfully transmitted via N transmission channels, without any further loss of azimuthal information. This preservation of azimuthal fidelity does, however, require $S(\theta)$ to be limited in its azimuthal harmonic content.

The requisite harmonic-content limitation for $S(\theta)$ also implies the same harmonic-content limitation for the interpolating or mixing function $M(\theta)$ whereby $S(\theta)$ is derived from the actual source pattern. Taking this limitation, together with the requirement that $M(0)$ be maximum while the area under $|M(\theta)|^2$ be constant, results, as shown in Appendix I, in the requirement that the azimuthal harmonic components μ_n of $M(\theta)$ be all equal to one another, a complete specification of $M_2(\theta)$, $M_3(\theta)$, $M_4(\theta)$, etc., the interpolators for 2, 3, 4, etc., transmission channels. These interpolators are to be discussed in further detail at a later point. At this point, however, the matrix specification is complete in view of Eq. (9), and it remains only to describe its structure in more detail.

For this description, two points of view may be taken, azimuthal sampling and azimuthal harmonic synthesis. For the N signals to be transmitted, there may be chosen, then, either the N equally spaced azimuthal samples or, what is fully equivalent in overall effect, the N azimuthal harmonic components. Azimuthal sampling is the one to be treated at a later point, and azimuthal harmonic synthesis is to be discussed next.

AZIMUTHAL HARMONIC SYNTHESIS—ENCODING

The coefficients in the Fourier series for the interpolated source function

$$\begin{aligned} S(\theta) = a_0 + c_1 \exp(j\theta) + c_2 \exp(j2\theta) + \ldots \\ + c_{-1} \exp(-j\theta) + c_{-2} \exp(-j2\theta) + \ldots \end{aligned} \tag{10}$$

are, after optimization and harmonic-content limiting, merely

$$a_0 = S_1 + S_2 + \cdots \tag{11a}$$
$$c_1 = S_1 \exp(-j\theta_1) + S_2 \exp(-j\theta_2) + \cdots \tag{11b}$$
$$c_{-1} = S_1 \exp(j\theta_1) + S_2 \exp(j\theta_2) + \cdots \tag{11c}$$
$$c_2 = S_1 \exp(-j2\theta_1) + S_2 \exp(-j2\theta_2) + \cdots \tag{11d}$$
$$c_{-2} = S_1 \exp(j2\theta_1) + S_2 \exp(j2\theta_2) + \cdots \tag{11e}$$

etc., where the number of coefficients is to be N, the number of transmission channels. In azimuthal harmonic synthesis, these coefficients are to be the actual transmission signals, and they are designated as follows:

$$T_\Sigma = a_0, \quad T_\Delta = c_1, \quad T_T = c_{-1}, \quad T_Q = c_2 \qquad (12)$$

etc., standing for sum-transmission, difference-transmission, tertiary-transmission, and quaternary-transmission channels, respectively. This channel list may, of course, be expanded, if further transmission channels should become of interest.

The sum-transmission channel T_Σ as shown in Eq. (11a) is merely the sum of all source signals without regard for azimuth. Thus, T_Σ is an omnidirectional channel and corresponds to the ideal of a mono channel. In making a recording, ideal mono compatibility would be assured by guaranteeing that T_Σ would be selected in playing the recording in the mono mode.

The difference-transmission channel T_Δ as shown in Eq. (11b) is again the sum of all source signals, but with each signal being shifted in phase to lag their counter parts in T_Σ by a phase angle that is equal to the azimuth angle for each source. It is called the difference signal because, in making a two-channel stereo recording, stereo compatibility is obtained by setting

$$T_\Delta = T_R - T_L \qquad (13a)$$

with

$$T_\Sigma = T_R + T_L \qquad (13b)$$

in which T_R and T_L are the usual right-transmission and left-transmission channels. With these definitions, a source at the cardinal point R appears entirely in T_R and not at all in T_L, while a source at the cardinal point L appears entirely in T_L. Stereo compatibility is to be discussed in more detail at a later point.

The tertiary-transmission channel T_T is a conjugate-phase (except for phases intrinsic to the source signals) version of T_Δ. Thus it may be combined with T_Δ in two different ways to make the azimuthal dependence appear in both sine and cosine form, so that 3-channel matrices need not be phasor matrices in overall effect. Optimal 4-channel matrices, however, make use of T_Q, involving azimuthal harmonics of the second harmonic, and these are phasor matrices.

Means for deriving these transmission signals are illustrated in Fig. 4 in the case of the signal T_T. The derivation of T_Σ and T_Δ will be obvious, and the same will be true of T_Q, allowing for its double-angle dependence. The sin θ and cos θ dependence upon the source angle may be introduced either electrically or acoustically. In the electrical formation, elementary voltage dividers may be used to assign an "on-mic" signal (S_{θ_1} in Fig. 4) to a specified direction θ_1, and this may be done in panpot fashion by using a sine–cosine potentiometer [11]. In the acoustical formation, the directional properties of a pair of coincident dipole or ribbon microphones may be used to introduce the sine and cosine dependence, as for the source S_{θ_2}.

Clearly, as many sources as desired are automatically mixed, each according to its natural azimuth, by the acoustical method. However, each "on-mic" channel must be afforded its own voltage divider, or panpot, if each is to be assigned to a different azimuth, and additional inputs are needed for the summing amplifiers, denoted by Σ in Fig. 4.

The 90° amplifier provides the 90° phase lead (multiplication by j) for the sin θ-proportioned signals with respect to the cos θ-proportioned ones. This phase shift need be provided only with reference to the output of the reference amplifier, since the input–output phase relation is of no interest. It is necessary, however, that this phase shift (so referenced) be maintained essentially constant over the audio band. Circuits for providing such shifts are known [12]. The reference phase for all transmission channels must also be the same.

If only two channels are to be encoded, they may as well be encoded directly as T_R and T_L, as defined by Eqs. (13), and as shown in Fig. 5. The channel T_T may also be derived from the same hardware, choosing the correct polarity from the 90° amplifier. The channel T_Q would, however, require additional facilities, a quadrupole microphone, for example, or a double-angle sin–cos pot.

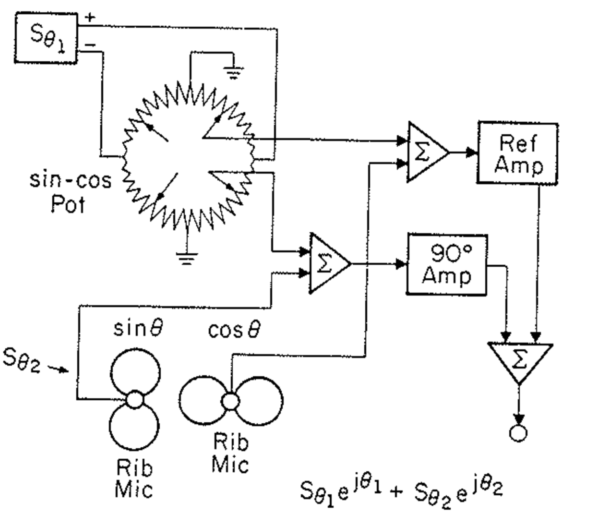

Fig. 4. Azimuthal harmonic synthesis showing both electrical and acoustical mixing.

AZIMUTHAL HARMONIC SYNTHESIS—DECODING

Decoding is actually the synthesis of $S(\theta)$ according to Eq. (10). However, this synthesis need be provided only for the listener-space azimuths at which loudspeakers are to be placed. To avoid confusion with the source-space azimuths, these listener-space, or presentation, azimuths are named ϕ_1, ϕ_2, etc., and the kth one is called ϕ_k. Also, the kth presentation signal is called $P(\phi_k)$, although it is identical with $S(\theta)$ evaluated at that particular azimuth. With these changes in notation, Eq. (10) is rewritten as

$$P(\phi_k) = T_\Sigma + T_\Delta \exp(j\phi_k) + T_T \exp(-j\phi_k) + T_Q \exp(j2\phi_k) + \cdots \qquad (14)$$

as the signal supplied to the kth loudspeaker located at azimuth ϕ_k.

This decoding or synthesis may be performed by the same means as in encoding, namely, processing the transmitted signals through phase shifters and thence through voltage-divider networks. A schematic diagram for a two-channel decoder is shown in Fig. 6. Outputs for a

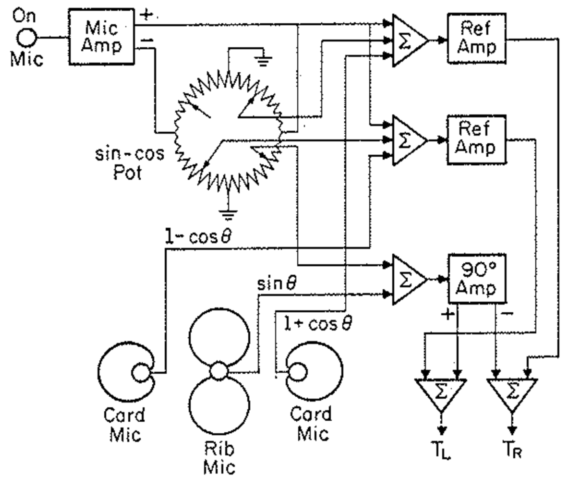

Fig. 5. Azimuthal harmonic synthesis of channels T_R and T_L.

great variety of presentation angles may be provided, either as the multiple outputs shown, or via switching. The user may thus place his loudspeakers to suit his listening-room configuration.

The overall encode–decode treatment of the source signals may be deduced by inserting the equations for T_Σ, T_Δ, T_T, and T_Q into Eq. (14). The result is

$$P(\phi_k) = S_1[1 + \exp(j\phi_k - j\theta_1) + \exp(-j\phi_k + j\theta_1) + \exp(j2\phi_k - j2\theta_1)] + S_2[1 + \exp(j\phi_k - j\theta_2) + \exp(-j\phi_k + j\theta_2) + \exp(j2\phi_k - j2\theta_2)] + \cdots. \quad (15)$$

That is to say, each source appears in the speaker signals multiplied by an overall-matrix coefficient (square brackets) that depends solely upon the difference between the speaker azimuth and the source azimuth. These square-bracket functions are overall interpolation functions to be discussed shortly.

AZIMUTHAL-SAMPLE INTERPOLATION

Fig. 7 shows in its upper part the overall encode–decode system, in which the intermediate or transmission channels are the azimuthal harmonic components. For brevity, the encoder is called a Fourier encoder and the decoder is called a Fourier decoder. The overall function is that of interpolation, however. An interpolation is made among a number of azimuthally deployed sources to provide signals for some number of speakers deployed according to some other, possibly more regular, azimuthal pattern.

One may ask whether these speaker signals might not, themselves, serve as transmission signals. Once the original source-pattern interpolation is made to derive an azimuthal harmonic content limited source function that correlates maximally with that source pattern, the sampling theorem guarantees no further information loss, provided the number of channels is sufficient to handle that limited content. Thus the answer to the question is "yes."

In the instance shown in Fig. 7 the requisite number is four, and channels corresponding to four equally spaced points in azimuth, Φ_1, Φ_2, Φ_3, and Φ_4, suffice. These azimuths need not agree with the final speaker azimuths. The encoder and decoder blocks each encompass the functions of the whole system shown in the upper block. In principle, this encoding may be cascaded endlessly without further information loss, provided that the number of channels is sufficient to represent the azimuthal harmonic content.

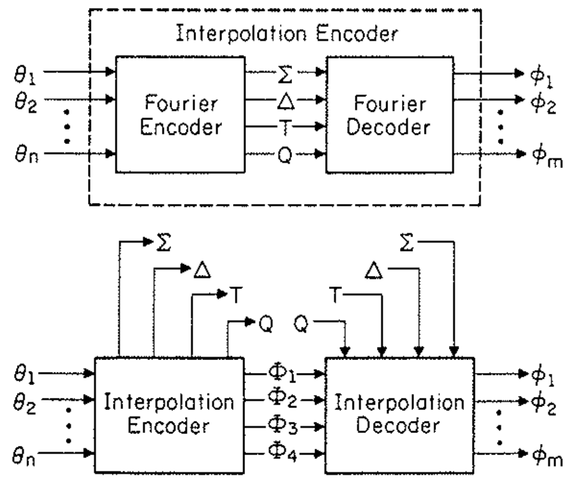

Fig. 7. Comparison of Fourier encoding and decoding (upper part) with interpolation encoding and decoding (lower part).

The interpolation encode–decode system of the lower part of Fig. 7 bears a close resemblance to the so-called discrete proposals for multichannel stereo, in that the transmission channels, themselves being azimuthally designated channels, would serve perfectly well as loudspeaker signal channels. The mixing prior to that, here called interpolation encoding, corresponds to the usual process of quadrasonic mixdown from the original multichannel material, except that it is optimized here for the maximal transmission of azimuthal information (see Appendix III).

The alternate azimuthal harmonic synthesis channels are, in principle, also available as shown, designated Σ, Δ, T, Q. These actually may be preferable transmission channels, because addition or deletion from the scheme of T_Q, for example, automatically changes the azimuthal

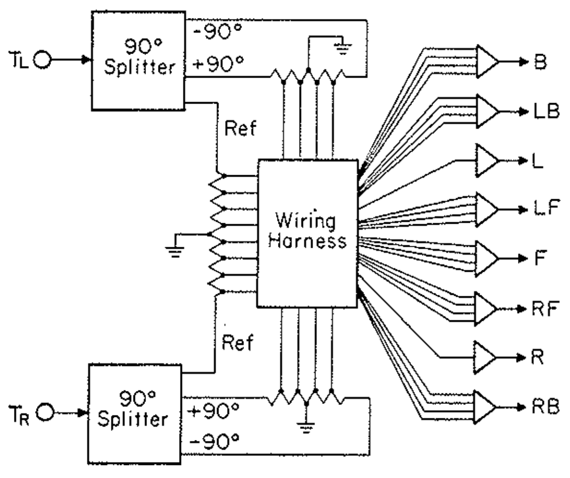

Fig. 6. Decoding of channels T_L and T_R to produce a number of loudspeaker signals.

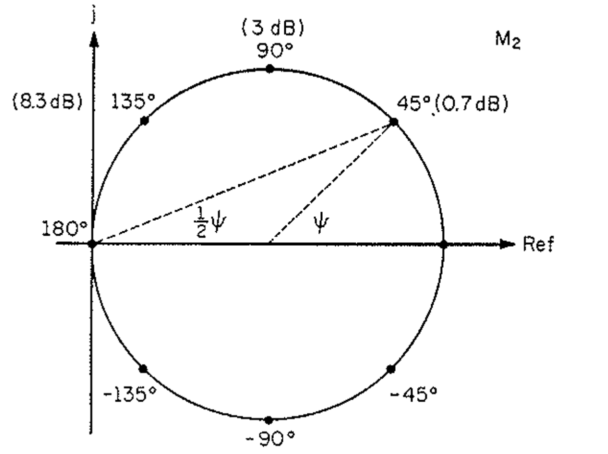

Fig. 8. Interpolation function M_2 as phasor plot. A line from origin to a point on the circle indicates by its length and angle the magnitude and phase, respectively, of the overall mixing coefficient, for differences in azimuth angle as marked on circle. These magnitudes are also quoted in dB.

harmonic content to be appropriate to the number of channels used. On the contrary, the loss of one of the azimuthally designated channels, say T_{Φ_i}, leaves the harmonic content inadequately represented, to cause actual errors in the interpolation, so that an optimal three-channel system would not be the result. It is the Σ, Δ, T, Q system, then, that provides for simple adjustment, as required by circumstances, of the channel capacity of the system.

The quality of the system, for varying adjustments of the number of channels, may be assessed by examining the overall interpolation functions whose plots are discussed in the next section.

INTERPOLATION PLOTS

The functions shown in square brackets in Eq. (15) are the overall interpolation functions for the optimal matrix system. The number of terms included depends upon the number of transmission channels used. These functions may be written as

$$M_2(\psi) = 1 + \exp(j\psi) = \exp(j\psi/2)\cos(\tfrac{1}{2}\psi) \quad (16a)$$
$$M_3(\psi) = 1 + 2k\cos\psi \quad (16b)$$
$$M_4(\psi) = \exp(j\psi/2)[\cos(\tfrac{1}{2}\psi) + q\cos(\tfrac{1}{2}3\psi)] \quad (16c)$$

in which ψ is the difference $\phi - \theta$ between loudspeaker azimuth and source azimuth, and the subscript designates the number of channels. The parameters k and q may be adjusted by permitting a certain amount of interchannel blending. The informational-optimal values are unity in each case. In writing Eqs. (16), overall proportionality constants have been omitted.

A plot of M_2 is shown in Fig. 8. This is a phasor plot using reference-phase and j-phase axes. The term $\exp(j\psi)$ is then a unit phasor extending (as would a vector) at an angle ψ from the reference axis and extending from the unit point on that axis. As ψ changes, the tip of that phasor traces a unit circle whose center is displaced one unit from the origin. The diameter along the reference axis is the overall transmission for a loudspeaker whose azimuth agrees with the source azimuth ($\psi = 0$), two units in the plot, to be taken as 0 dB.

Proceeding counterclockwise, the next point marked is for a loudspeaker 45° counterclockwise from the source location. The length of the line from the origin to that point is 0.7 dB below the two-unit reference, and the angle made with the reference line is 22.5°. Proceeding in this way, the magnitude and phase for all the off-source loudspeaker signals may be deduced. The null for the loudspeaker that is 180° away from the source is evident. What is seen here is an azimuthal "beat," alternating constructive and destructive interference between T_Σ and T_Δ, determining the loudspeaker feeds. If the magnitude of the feeds alone were plotted, in a polar plot, a broad cardioid pattern would be shown with a single null at 180°.

A cardioid pattern is shown in polar plot in Fig. 9 for M_3, the overall interpolating or mixing curve for three channels. The dashed curve is the optimal one. It is down 9.5 dB for loudspeakers 90° and 180° away from the source azimuth and it has nulls at ± 120°. The nonoptimum solid-line curve shows a possible tradeoff between backlobe suppression and 90° separation, bringing the 90° separation figure to 6 dB, a tradeoff whose advantages may be doubted. In either case, a marked improvement over the two-channel case is exhibited.

The cardioid plot for M_4, Fig. 10, shows a further sharpening of directivity. In the optimal dashed-line plot, nulls at ± 90° and 180° are shown (infinite separation), the usual mark of a four-channel "discrete" system. The solid curve shows a price that may be paid for backlobe suppression, a 90° separation of 9 dB. In either case one may feel assured that the ear would be well satisfied with such a high order of directivity.

The sequence of plots of Figs. 8, 9, and 10 shows members of a universe of optimal multichannel matrices, which may be called UMX. It starts with the nearly trivial, but ideal, omnidirectional mono system MMX, which with the addition of one more channel becomes the basic, or bimodal, two-channel matrix BMX. Then, successive augmentations in the number of channels in the optimal manner results in the trimodal TMX, and then the quadramodal QMX, etc. This ability to move upward or downward in the UMX hierarchy has important consequences in regard to compatibility and

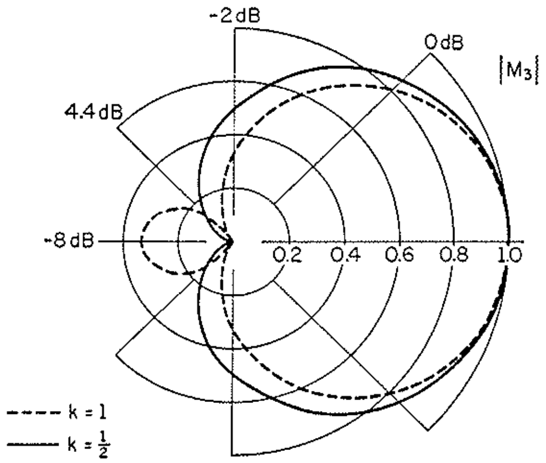

Fig. 9. Interpolation function M_3 as polar plot of magnitude, shown in dashed line. The solid line is a nonoptimal variant that may be obtained for backlobe suppression.

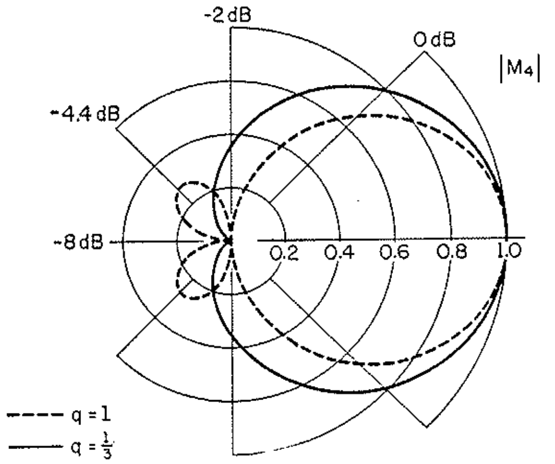

Fig. 10. Interpolation function M_4 (dashed line) in plot similar to that of Fig. 9.

growth potential. The examination of such questions starts with the BMX, or basic, matrix.

BASIC TWO-CHANNEL MATRIX

The basic two-channel matrix (BMX) uses the two channels T_Σ and T_Δ which correspond, in record-groove modulation, to lateral and vertical modulation, or in FM transmission, to baseband and subcarrier transmission. Alternatively, these may be represented as

$$T_R = T_\Sigma + T_\Delta \tag{17a}$$

$$T_L = T_\Sigma - T_\Delta \tag{17b}$$

corresponding to azimuthal cardinal points R and L, respectively. With this observation, the signals T_R and T_L may be deduced for various source azimuths from Fig. 8, or more conveniently from Fig. 11, showing the signals for T_R as the phasor arrows drawn from the origin. The phasor signals for T_L are diametrically opposite those for T_R (dashed line).

In compatible ordinary-stereo playback, T_R feeds the right speaker (usually placed at RF), T_L feeds the left speaker (usually placed at LF), and the ratios of these signals become the ordinary-stereo separation figures. These figures may be read from Fig. 12 for various source azimuths. It is seen that T_L leads T_R by 90° in phase for sources in the front half of the azimuthal space but lags T_R for the back half, while the magnitudes of the separations are the same front and back. Taking magnitude and phase into account, apparent ordinary-stereo localizations may be computed, as shown plotted along the diameter of the circle. It is seen that an excellent range of localizations is obtained as an indication of full ordinary-stereo compatibility.

These localizations, confirmed in psychoacoustic trials, were calculated from the theory of Makita [13], from which the curves shown in Fig. 13 were computed. Makita's theory is a physical wavefront-localization theory. The computation is of the propagation direction of the phase front for the composite acoustic wave at the listener position. As such, it corresponds to the psychoacoustic localization obtained by the listener who seeks to orient his head to face the apparent source.

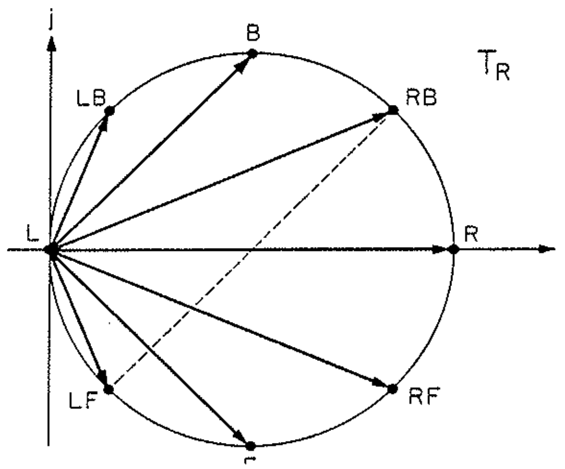

Fig. 11. Phasor values, shown by arrows, of T_R signal for various source locations. The dashed line shows how corresponding T_L values may be read.

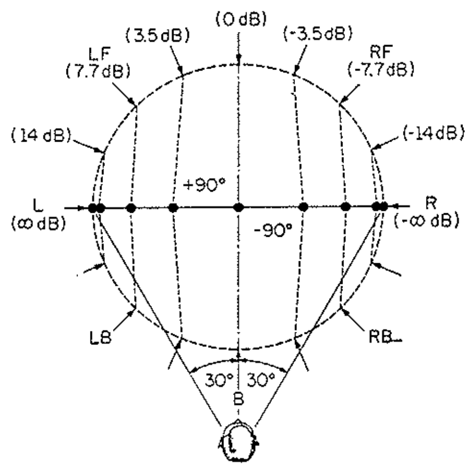

Fig. 12. Stereo compatibility of BMX. For various source azimuths the ratios of T_L to T_R are quoted in dB. Front-half insertions cause T_L to lead T_R by 90°, while the opposite is true for the back half. Stereo localizations for these data are shown as computed from the theory of Makita.

The question of the advisability of the 90° phase shift has been examined with regard to anomalous source spread. It has been found that mono signals divided between the two channels experience no stereo source spread when the 90° phase shift is imposed, but merely produce the localizations shown on Makita's 90° curve (Fig. 13). With several such signals divided in different proportions in the two channels, the spacing in localization increases when the 90° shift is used, in excellent agreement with the comparisons that may be made between the 0° and the 90° curves of Makita.

On the other hand, if a single source, e.g., a piano, had been picked up with two spaced microphones, there may already exist frequency-dependent phase differences between the two channels that can impart a localization spread to the stereo image. Then, with the further provision of 90°, these phase variations could, in overall effect, exceed 120°, and invoke occasional anomalous localizations. Thus, while it may be inadvisable to subject some of the older more wildly phased stereo recordings

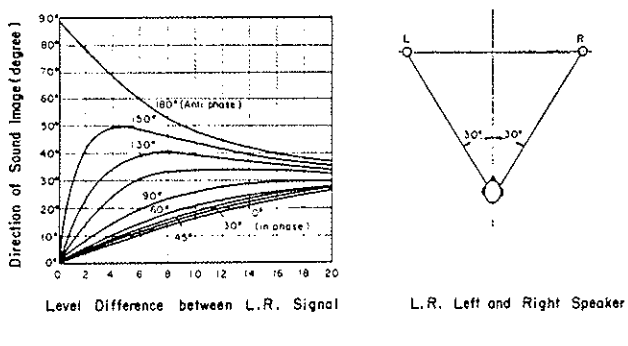

Fig. 13. Plot of stereo-localization theory of Makita, showing results for wide range of phases and separation ratios.

to a further interchannel phase shift, no such strange effects need occur in new well-planned quadrasonic mixes. Thus, in contrast to some matrix proposals that impose phase shifts approaching 180°, this conservatively phased matrix should offer no problems with stereo compatibility.

In fact, BMX matrix records have been carefully auditioned in ordinary-stereo playback, and they have proven to be splendid stereo records indeed. These BMX records also, of course, are ideal in mono playback, because of the omnidirectional quality of the sum channel. There are no center-channel buildups nor back-channel dropouts. The mono-feed, equal signal to all four channels, is a perfectly acceptable four-channel input to the matrix.

Localizations in four-channel playback may also be examined with the help of Makita's theory, as explained in the next section.

MATRIX LOCALIZATIONS

The quadrasonic extension of the Makita localization theory predicts the composite wavefront to propagate from a direction given by

$$\tan\Theta = \frac{\mathrm{LF\cdot LF - RB\cdot RB + RF\cdot LF - RF\cdot RB + LF\cdot LB - LB\cdot RB}}{\mathrm{RF\cdot RF - LB\cdot LB + RF\cdot LF + RF\cdot RB - LF\cdot LB - LB\cdot RB}} \qquad (18)$$

in which Θ is measured counterclockwise from the RF azimuth, and the notation of the dot products is interpreted as

$$\mathrm{RF\cdot LF} = |P_{\mathrm{RF}}|\,|P_{\mathrm{LF}}|\cos(\varphi_{\mathrm{RF}} - \varphi_{\mathrm{LF}}) \qquad (19)$$

for example. Here, $|P_{\mathrm{RF}}|$ is the magnitude of the signal for the RF speaker and φ_{RF} is the phase of that signal. The use of RF as the reference azimuth for the measurement of Θ departs from the usual custom observed in this paper, but is made for some mathematical simplifications. In interpreting Eq. (18), the relative polarity of numerator and denominator serves to identify the quadrant for Θ in the usual manner, as needed for the correct computation of an arctangent function.

The wavefront source localizations predicted from Eq. (18) pertain to an observation made at a point equidistant from the four speakers arrayed at equal intervals in azimuth. Phase-front propagation directions may be computed for other parts of the listening space, using more complicated formulas. Because of the differing time delays involved, however, the interpretation of such calculations requires the input of psychoacoustic parameters relating to time resolution, complicating the appealing physical simplicity of the Makita theory. Since this more elaborate theory has not yet been developed, the risks involved in such elaboration need not be considered further, and only the present simple theory need be used.

It is gratifying to be able to report that all of the UMX matrices (excluding the mono matrix) provide loudspeaker-feed signals that, when inserted into Eq. (18), result in wavefront localizations in exact agreement with the actual source locations. This fact is illustrated by the UMX pattern shown in Fig. 14, in which the points shown on the outer circle are source locations, and the points on the inner circle are apparent source locations, shown paired with the input sources via connecting lines.

The patterns marked "other" in Fig. 14 are for certain other (nameless herein) widely publicized two-channel matrices. Both are phasor matrices, and one had been designed for extreme left–right separation at the expense of front–back separation, while one uses the same magnitude-separation pattern as in the BMX, but a rather different phase pattern. Needless to say, the localizations for the UMX have been confirmed psychoacoustically, while it is possible to report only an impressionistic confirmation for the "others."

The impressionistic experience suggests some interesting facets of localization theory. There are actually two kinds of such a theory. One pertains to the localizations obtained when the listener is expected to turn his head to face the apparent source, so that his judgement is based on equal (as nearly as may be) conditions at his two ears. The other is a fixed-head theory and it attempts to predict a judgement based on unequal conditions. An example of the latter kind is that of Mertens [14]. While these theories ought to agree for actual sources, there is no guarantee that they will agree for a "phantom" source with localization to be construed from a composite wave.

It is believed that listeners dislike making fixed-head localization judgements, presumably because the style is incompatible with maintaining a critical frame of mind, although this dislike may be conquered with a little coaching. However, the listener is probably correct in his natural preference for turning his head. He knows that he can be deceived in a complicated acoustical situation and that if he "really needs to know," he must turn his head. Thus, matrix designers relying upon fixed-head localization phenomena are probably relying upon a weak reed, one which will sway with changing listening-room acoustics and one which may collapse if the listener turns his head. This observation applies with particular force to limited-separation matrices.

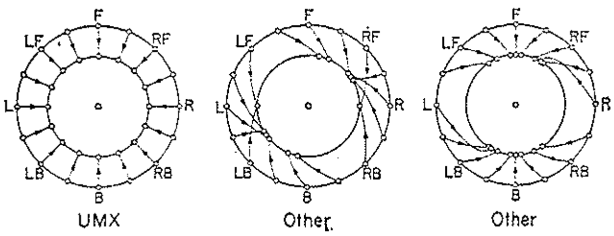

Fig. 14. Localization results from quadrasonic extension of Makita's theory. It is seen that a perfect localization pattern is obtained for UMX, while patterns for two "other" matrices are not so good.

The authors' impression of the "other" matrices of Fig. 14 is that they sound "not bad" in fixed-head listening, but that the sources become fugitive upon a turning of the head and do finally reside more or less at the plotted positions. On the other hand, the BMX, the two-channel member of the UMX, is also a limited-separation matrix, but one whose source localizations become much less fugitive, if at all, in changing from a fixed-head listening style to a turning-head listening style, so that the UMX pattern of Fig. 14 is invariably confirmed. Mixing engineers report their delight in making direct tape–matrix comparisons that support Fig. 14, even for the BMX.

The excellence of the BMX localizations is readily understood on an intuitive level. For a source at LF, for example, the RF and LB flanking speakers are at exactly matching levels, so as to aid either a LF or RB localization. However, the LF speakers are supplied with a 3-dB stronger signal, and the RB speaker is actually silent. Moreover, the LF-speaker signal is supplied with a phase which is exactly centered between the phases of the RF and LB flanking speakers, $-45°$ and $+45°$, respectively. It is no wonder that the LF localization is reliably compelling. The psychoacoustic principle is nothing more than the familiar fill-speaker or interpolation-speaker phenomenon, a most stable one, with proper management of phasing. This is the only matrix for which side-effect signals are actually localization aiding.

Because the Makita theory is not a psychoacoustic theory, however, it may be used to examine matrix performance only in regard to satisfying one necessary requirement, the correct orientation of the composite phase fronts, the theory being silent with regard to further requirements. Thus all UMX systems are equally good in the Makita theory, even the limited-separation BMX. For the BMX presentation, tests show that upon moving away from the center, the excellent UMX localization pattern does change, exhibiting localizations biased toward the listener position.

Although this retreat from wide-ranging localization is graceful, and not altogether displeasing, it is rather fast for the BMX, in comparison to the more firmly locked localizations provided by its bigger brothers, the TMX and the QMX. Even the TMX gives a discrete-matrix effect.

DISCRETE-MATRIX EXPERIMENTS

The structure of the QMX decoder is such that it always contains the basic or BMX matrix, if the T_T and T_Q channels are silent, or that it always contains the TMX as an augmentation of the basic matrix if only the T_Q channel is silent. Thus it is a simple matter to change channel capacity without changing the loudspeaker configuration, or without changing the four-channel source material, and thus explore the psychoacoustic relevance of the various levels of directional acutance indicated in the plots of M_2, M_3, and M_4.

At first the well-centered listener is not impressed when, starting from the BMX, an augmentation is made by admitting the T_T channel. He reports a slight loudness increase, but claims the localizations to be essentially unchanged. Rather, it is the off-center listeners who say, "Ah," and who describe the change as comparable to switching over to the original four-channel tape.

Starting from the TMX, however, and deleting the T_T channel does sometimes elicit a response from the centered listener, illustrating a general principle that deprivation is more keenly felt than augmentation, at least for moderate changes. That the change is only moderate for the centered listener is evident in the difficulty he usually experiences in describing the difference in words.

By contrast, the off-center listeners never exhibit any difficulty in describing the BMX–TMX difference. The reaction is emphatic, whichever direction is taken in making the change, and the word most often used is "discreteness." If he is sufficiently off center, he explains that the TMX restores the localizations that the BMX had allowed to fall too near to his own position.

The TMX–QMX comparisons are marked by the onset of diminishing returns. Detection of the difference is extremely difficult for the centered listener, and only the most attentive and knowledgeable listeners achieve any reliable ability at distinguishing the two, even at far, off-center locations.

In general terms, the augmentation of the BMX upon adding the acutance-augmenting channels T_T and T_Q results in an effect called discreteness, which may be principally identified as increasing the area of the usable listening space. In particular, this discreteness is only marginally increased in moving to the QMX, the TMX being essentially fully discrete.

BAND-LIMITED ACUTANCE

The ability to switch between matrices of the same general character, differing only in directional acutance, made it possible to test the conjecture that human hearing makes rather little use of the high-frequency content in judging localization. For example, it might be supposed that for wavelengths appreciably shorter than the distance around the head from ear to ear, the relatively complicated phase and amplitude relations might make for ambiguities that would be preferably ignored, especially if the lower frequency more reliable information were available for judging the localization. What is preferred, as a localization strategy, would seem to be the question, since it is recognized that high-frequency localization is possible. As a practical matter, the question is of great interest, since the economics of providing additional channels could well depend upon the question of bandwidth.

Accordingly, trials were made with phase-equalized low-pass filters being inserted by means of switches into the T_T and T_Q channels. In one test the filters had a 12-dB point at 3 kHz and an effective noise bandwidth of about 1.3 kHz. The switching was also arranged to make an equalization correction to compensate for the augmented energy transmission in the lower frequencies, in the filtered cases. The filtered cases, then, consisted of the TMX or QMX matrix trials for frequencies below the cutoff region of the filter, but only the BMX for frequencies above.

The tests showed that, as a practical matter, the conjecture is valid. Listeners found it exceedingly difficult to detect either the loss or the gain in high-frequency acutance, whether the lower frequency acutance was that of the TMX or the QMX. Also, the ability to distinguish the TMX from the QMX was maintained at about the same low level of reliability, regardless of filtering.

It should be noted that all of these tests were of the A/B type, affording listeners the opportunity to make close comparisons based on sudden changes. With any appreciable time lapse between comparisons, the ability to make distinctions rapidly fades. It is believed, for example, that the TMX system with band-limited acutance would remain distinguishable from the BMX even on the basis of a time-lapse comparison, but that any distinction more marginal than that one would become impossible. Under such circumstances, the TMX with band-limited acutance becomes the practical equivalent of a fully discrete four-channel system.

BAND-LIMITED ACUTANCE SYSTEMS

The point of view developed in the previous section, that the BMX may be regarded as a basic two-channel matrix which may be augmented through the use of acutance-augmenting channels T_T and T_Q, appears to be a most powerful one in considering compatibility questions.

For example, in Hanson's proposal for expanding the Musicassette format from four to eight tracks to provide a quadrasonic capability [15], he would split the L track in two to carry LF and LB information (proposed tracks 1 and 2, respectively), and he would split the R track in two to carry the RB and RF information (proposed tracks 3 and 4, respectively). The existing two-track stereo players would then sense front–back blended signals for the L and R stereo signals with a slightly worsened signal-to-noise ratio (SNR) because of the unrecorded guard space between the new tracks. A disadvantage is that the front–back blend matrix offers no interesting dematrixing possibilities, unless a pair of front–back difference signals could be made available.

The UMX approach would be to record the BMX signals T_L and T_R across the full width of the tracks originally provided for L and R signals and to record T_T and $-T_T$ in tracks 1 and 2 of Hanson's format, and T_Q and $-T_Q$ in tracks 3 and 4, as superpositions upon these full-width signals (Fig. 15). In this proposal no formal provision for additional unrecorded guardspace need be made. The ordinary players, suitable for decoding into a BMX quadrasonic presentation, would then sense the mono–stereo compatible matrix signals of the BMX while new players equipped with track-splitting heads would also sense the acutance-augmenting channels T_T and T_Q for discrete-matrix quadrasonic presentation.

Any means of subdividing the cassette tracks to provide additional channels necessarily tends to worsen the SNR, as Hanson notes. In the cassette medium this is a critical matter. The UMX offers, however, the possibility of a restricted bandwidth for the acutance channels, and a factor 6.25 reduction in bandwidth, reducing the rms noise by a factor 2.5, would allow a peak signal reduction by that same factor 2.5 without loss of SNR. Now, without this factor, each subtrack would devote ⅓ of the peak capability to each of T_Σ, T_Δ, and T_T (or T_Q), but with this factor, the ratios become 1.25/3, 1.25/3, and $1/(3\times1.25)$, so that the sum of T_Σ and T_Δ may account for 2.5/3 of the peak modulation, a loss of 1.6 dB in SNR. On the other hand, each of T_Σ, T_Δ, T_T, and T_Q show a ratio 1.25/3 relative to the present stereo, or −7.6 dB, which signals, on fourfold coherent combination, show an improvement by $(12-6)$dB to restore the net loss of 1.6 dB.

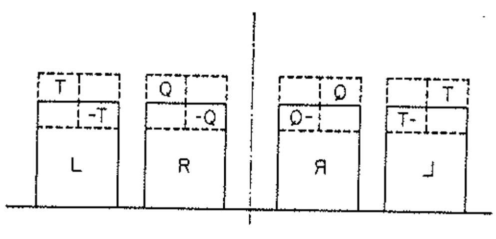

Fig. 15. Possible track allocation for cassette.

In Hanson's proposal, by contrast, the loss of recording space, because of the need for a guardspace, causes the SNR loss to be 1.6 dB for ordinary stereo, while the SNR loss for each speaker channel is 3.8 dB according to the proposed trackwidths. The SNRs for the UMX band-limited-acutance approach, then, are the same as Hanson's for the compatible stereo, and better than Hanson's by 2.2 dB for discrete-matrix playback. The general expectation, then, is that the SNR loss that might be marginally noted in the Hanson format should go fully unnoticed in the UMX format.

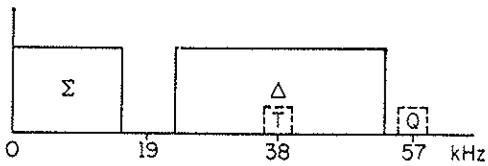

Fig. 16. Possible frequency allocation for FM multiplex.

The UMX format, moreover, offers four compatible modes of playback:

1) mono,
2) stereo,
3) two-channel matrix (BMX) quadrasonics,
4) discrete-matrix quadrasonics,

with the fourth mode available to players with track-splitting heads. These same track-splitting UMX band-limited-acutance methods are also possible in the cartridge format without incurring the playing-time loss of the Q-8 provision, and similar possibilities are available to the open-reel tape format.

In quadrasonic FM multiplex (QFMX) the UMX approach also promises to offer the same four compatible modes of presentation. Using band-limited acutance signals, there is little problem in making compatible spectrum allocations. For example, the availability of the quadrature-phased 38-kHz subcarrier for modulation has been previously noted [16]. The possibility that some receivers may treat the two 38-kHz sidebands unequally and thus introduce crosstalk between the quadrature-phased carrier signals becomes improbable if the additional signal, say T_T, is of sharply reduced bandwidth. There is also the possibility of the placement of a narrow-band modulation upon a synchronized 57-kHz carrier (3rd harmonic of the 19-kHz pilot) for T_Q (Fig. 16). Thus no interference with SCA (subsidiary communication authorization) need be contemplated in these and other possible spectrum allocations.

Of greatest importance among these media is the ordinary disc record. The introduction of carrier channels is a current proposal [17]. This proposal appears to involve:

1) frequency modulation at 30 kHz covering the frequency range up to 45 kHz,
2) tracing-error compensation,
3) adaptive carrier and deviation levels,
4) special pickup construction,
5) special stylus shapes,
6) special hard plastics,
7) reduced baseband signal levels,
8) reduced-speed recording,
9) sum and difference matrixing

as key points of feasibility and compatibility. Because the list is so long, many observers seem to feel that this carrier-channel disc is too near the very frontiers of disc-recording technology to be regarded as typifying the conservatively engineered product acceptable in current studio practice and capable of gaining widespread consumer acceptance.

Because of such doubts, it is interesting to see the differences that might be made in the carrier-channel disc if the UMX with band-limited acutance channels were used. The changes that become possible are conveniently keyed to the above list:

1) reduction of carrier frequency and increase of FM modulation index keeping all frequency components below 30 kHz,
2) tracing-error compensation probably not needed,
3) adaptive carrier and deviation level probably not needed,
4) present high-quality pickups OK,
5) present stylus shapes OK,
6) special plastic not needed,
7) negligible reduction in baseband level,
8) speed reduction not so severe,
9) optimal matrixing with four compatible playback modes.

The improvement in regard to items 3) and 7) stems from the SNR improvement traceable to the bandwidth reduction directly, and also indirectly, because of the lowering of the carrier frequency, making possible the increased modulation index (Fig. 17). At the same time, the lowered carrier frequency is easier to track and trace (items 2-5), while the carrier signal is less fragile mechanically (item 6)). The speed reduction in recording is more like 2:1 than 3:1, and may be obviated at a sooner date via cutter improvements. The improvement in regard to item 9) appears in the other UMX applications. These changes tend toward bringing the carrier-channel disc more nearly within the range of well-practiced and conservatively engineered disc technology, and should sharply enhance consumer acceptance.

In regard to item 9) the question may be asked as to whether the front–back sum-and-difference matrix may not also be used with a bandwidth reduction for the difference. It should be clear, however, that front–back localizations would then be treated on a different basis than left–right localizations, and not merely in acutance, but in the localization itself. It is the azimuthal symmetry of the UMX that places the further channels in a mere acutance-enhancing role for localizations that are already correctly made. It is this aspect that helps conceal the effect of a reduction in bandwidth for those channels.

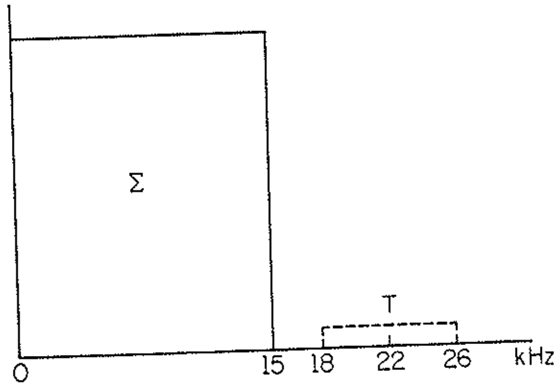

Fig. 17. Possible frequency allocation for disc.

The hardware realizations of these UMX discrete-matrix band-limited-acutance systems are currently under development, with highest priority being given to the disc. It is expected that highly reliable conservatively-engineered consumer-acceptable recordings may soon be offered by the industry following these approaches.

CONCLUSIONS

In this paper it has been shown how the basic periodic structure of azimuthal space, together with an optimization principle, the maximizing of azimuthal fidelity, can lead to a unique solution to the problem of mixing to obtain an optimal azimuthal signal presentation, taken to be the goal of multichannel matrix stereo. The details have involved considerations of optimal azimuthal interpolation of the source pattern, within constraints imposed by a limitation of the number of channels which, because of the azimuthal sampling theorem, is equivalent to a limitation in azimuthal harmonic content.

Azimuthal harmonic synthesis of encode and decode matrices is the means of using the azimuthal harmonic components as transmission channels, whereby the overall encode–decode interpolation curves are seen to be cardioid functions exhibiting a directional acutance that increases with increasing number of channels. These cardioids, M_2, M_3, M_4, etc., specify the performance of a universe of optimal multichannel matrices, UMX, containing a two-channel member, BMX, a three-channel member, TMX, and a four-channel member, QMX, all suited to quadrasonic presentation, but with increasing directional acutance.

Both psychoacoustic experiments and localization theory confirm the azimuthal fidelity of these matrices. These experiments further show that the increase in directional acutance, as the number of transmission channels increase, is directly related to a quality called discreteness, whose practical meaning relates to the usable fraction of listening space. The gain in discreteness is essentially complete in the TMX.

Thus the BMX may be regarded as a basic matrix that may be augmented by adding the further channels of the TMX or QMX, so that these further channels play the role of acutance augmentation. These experiments show that the bandwidth of the acutance-augmenting channels may be sharply limited to frequencies below some 2 kHz.

With band-limited acutance, cassette, FM, and disc become media for which a four-channel discrete-effect matrix system appears to be fully feasible. This feasibility depends on technology that is within the state of the art, avoiding marginal engineering, and making for the possibility of reliable consumer-acceptable practice.

This consumer-acceptable practice embraces four compatible modes of consumer use: mono, ordinary stereo, matrix-stereo quadrasonics, and discrete-matrix quadrasonics. The engineering realizations of these systems are currently under development.

It is believed that the optimal structure of the UMX system of matrices, together with the wide range of compatible applications, and the engineering economies they appear to offer, make them uniquely suited for international standardization.

APPENDIX I
OPTIMAL AZIMUTHAL CORRELATION

The azimuthal smoothing of a source pattern consisting of random source signals S_1, S_2, $\cdots$, S_i, $\cdots$ at azimuth angles θ_1, θ_2, $\cdots$, θ_i, $\cdots$ generates the interpolated source-pattern function

$$S(\theta) = \Sigma_n S_n M(\theta-\theta_n) \tag{20}$$

in which $M(\theta)$ is the interpolation function. The correlation between the discrete S_i pattern and the pattern $S(\theta)$ is

$$R = E\{S_i^* \Sigma_n S_n M(\theta_i-\theta_n)\} \tag{21}$$

in which the asterisk denotes the complex conjugate and E denotes the statistical averaging operator. If, as assumed, S_i is independent of θ_i, and S_i is independent of S_n and is zero mean, then Eq. (2) may be evaluated to be

$$R = \Sigma_n E\{|S_n|^2\} M(0). \tag{22}$$

Since R may be arbitrarily increased by merely increasing the scale of M, its maximization is to be subject to a constant-energy constraint. This energy is

$$\mathcal{E} = E\left\{\frac{1}{2\pi}\int_0^{2\pi} |S(\theta)|^2 d\theta\right\} \tag{23}$$

which is found to be

$$\mathcal{E} = \Sigma_n E\{|S_n|^2\} \frac{1}{2\pi}\int_0^{2\pi} |M(\theta)|^2 d\theta. \tag{24}$$

Thus the maximization is to be the maximizing of $M(0)$, holding the area under the curve $|M(\theta)|^2$ constant.

Since M must be of limited azimuthal harmonic content, it may be represented as the N-term Fourier series

$$M(\theta) = \Sigma_k \mu_k \exp(jk\theta) \tag{25}$$

from which it is evident that

$$M(0) = \Sigma_k \mu_k \tag{26}$$

and that

$$\mathcal{E}_M = \frac{1}{2\pi}\int_0^{2\pi} |M(\theta)|^2 d\theta = \Sigma_k |\mu_k|^2 \tag{27}$$

the latter deriving from Parseval's theorem. The maximizing of $M(0)$, holding $\mathcal{E}_M$ constant, produces the result

$$\mu_k = \sqrt{\mathcal{E}_M/N} \tag{28}$$

for all k, so that the correlation R is proportional to

$$M(0) = \sqrt{\mathcal{E}_M N} \tag{29}$$

a quantity which increases as the permissible harmonic content increases.

This same harmonic content limitation obtains for $S(\theta)$, so that it also may be represented by an N-term harmonic series

$$S(\theta) = \Sigma_k c_k \exp(jk\theta). \tag{30}$$

Inserting Eq. (25) into Eq. (20) and making use of Eq. (28) is the means of showing that each c_k must be proportional to

$$c_k = \Sigma_i S_i \exp(-jk\theta_i). \tag{31}$$

The necessity for limited azimuthal harmonic content is shown in Appendix II.

APPENDIX II
AZIMUTHAL SAMPLING THEOREM

Let the N periodic samples $S(\theta_n)$ be obtained from the source function

$$S(\theta) = \Sigma_m c_m \exp(jm\theta) \tag{32}$$

and let thereby constructed the presentation function

$$P(\theta) = \Sigma_n S(\theta_n) M(\theta-\theta_n) \tag{33}$$

by means of the overall matrix interpolation function

$$M(\theta) = \Sigma_k \mu_k \exp(jk\theta). \tag{34}$$

Then the presentation function is

$$P(\theta) = \Sigma_k \Sigma_m \lambda_{km} \mu_k c_m \exp(jk\theta) \tag{35}$$

in which

$$\lambda_{km} = \Sigma_n \exp[j(m-k)\theta_n]. \tag{36}$$

The goal of the sampling theorem is the determination of conditions such that λ_{km} must vanish for $k \neq m$, so that there will be obtained $P(\theta) = S(\theta)$. Clearly, for $0 < |m-k| < N$, the sum of Eq. (36) is over a set of phasors equally spaced over 360° so that the sum is zero, but in the $m = k$ case, the sum is N. Thus if

$$\mu_k = 1/N \tag{37}$$

then $P(\theta) = S(\theta)$.

The question then becomes one of determining conditions under which $|m-k| < N$. It is assumed that the set of integers for k includes the set of integers for m. The number of integers in each set may be written $N\{k\}$ and $N\{m\}$. Then, $|m-k| < N$ for

$$N\{k\} + N\{m\} \leqq 2N \tag{38}$$

provided that, because the k set includes the m set,

$$N\{m\} \leqq N \tag{39}$$

where the inclusion of the m set within the k set automatically implies

$$N\{m\} \leqq N\{k\}. \tag{40}$$

The violation of Eqs. (38) or (39) leads to an interpolation error, involving a spurious azimuthal harmonic content for $P(\theta)$ called alias-harmonic content. The violation of Eq. (40) produces an interpolation error involving a deficiency of azimuthal harmonic content, a Fourier-series truncation error. Choosing $N\{m\} < N$ is harmless, but it makes for an inefficient use of azimuthal samples.

An example of alias-harmonic content may be indicated. Let $k = N + m$ for some certain pair of values k,m. Then $\lambda_{km} = N$ for that pair, and there will appear in $P(\theta)$ the term $c_k \exp[j(k-N)\theta]$, a component originally at the kth harmonic newly appearing at the $k-N$ alias harmonic. It is the avoidance of this error that demands $|n-m| < N$.

APPENDIX III
NONOPTIMAL MATRICES

Every matrix that is not a member of the UMX hierarchy is nonoptimal in not maximizing the transfer of azimuthal information for the number of channels used. Among two-channel matrices, bad localization patterns can be the result, as has been shown. Among three-channel matrices, one non-UMX example has been described by Eargle [2], and it shows a single 180° phased loudspeaker feed at a point not symmetrically located relative to the source, so that a suboptimal localization pattern is to be expected. If, however, Eargle's matrix be regarded merely as an algebraic object, then its azimuthal significance is subject to reinterpretation by way of noncyclic permutations in channel designations, and one such interpretation agrees in speaker-feed pattern, at least for the four source azimuths, with the corresponding TMX pattern.

With three channels and four channels it is possible to devise a nonoptimum matrix that shows a good localization pattern, but merely fails to exploit the full information capacity of the number of channels used. Examples may be seen in the nonoptimum adjustment of the k and q parameters in the TMX and QMX. Another example is seen in the pairwise mixing (PWM) formula for stereo mixing applied to four channels.

PWM uses a mixing curve $M(\theta)$ that equals $\cos\theta$ provided that $\cos\theta > 0$; otherwise the function $M(\theta)$ is zero. This mixing curve is clearly not of limited azimuthal harmonic content.

For four sources at angles θ_1 counterclockwise from RF, θ_2 counterclockwise from LF, θ_3 counterclockwise from LB, and θ_4 counterclockwise from RB, the four presentation channels are derived from the matrix equation

$$\begin{bmatrix} P_{RF} \\ P_{LF} \\ P_{LB} \\ P_{RB} \end{bmatrix} = \begin{bmatrix} \cos\theta_1 & 0 & 0 & \sin\theta_4 \\ \sin\theta_1 & \cos\theta_2 & 0 & 0 \\ 0 & \sin\theta_2 & \cos\theta_3 & 0 \\ 0 & 0 & \sin\theta_3 & \cos\theta_4 \end{bmatrix} \cdot \begin{bmatrix} S_1 \\ S_2 \\ S_3 \\ S_4 \end{bmatrix}$$

in PWM. Since the determinant of this matrix vanishes for $\tan\theta_1 \tan\theta_2 \tan\theta_3 \tan\theta_4 = 1$, it is seen that one row of the matrix may be formed as a linear combination of the other three for this combination of angles, so that one of the four channels is redundant. The channel capacity of this matrix varies between three and four channels, depending upon source placement.

Although the PWM provides a perfect localization pattern in the Makita theory, it does not contain an omnidirectional mono channel as its four-channel sum, since this four-channel sum exhibits a 3-dB between-the-speaker buildup. As a source mix for the UMX system, PWM thus produces suboptimal results. The azimuthal uniformity of energy in the UMX is thus subjected to a 0.8-dB fluctuation in the BMX, a 0.6-dB fluctuation in the TMX, and a 0.5-dB fluctuation in the QMX. Localization errors, however, never exceed 2° in the BMX and are even smaller for the higher order UMX members. That the UMX is degraded so little through the use of a nonoptimal source mix is attributed to the parametric insensitivity that characterizes optimal systems.

It is fortunate that PWM produces a source mix showing such a high degree of compatibility with the UMX systems, since this mixing formula is very popular. Once its advantages are more widely appreciated, however, QMX mixing may be used for the source mix.

Another popular four-channel source mix is obtained by using, as channel signals, the signals from four coincident cardioid microphones directed at RF, LF, LB, and RB. However, one of these channels contains redundant information, since the microphone sensitivity curve is the same as $M_3(\psi)$ (Fig. 9), the overall mixing specification curve for the three-channel matrix TMX. Thus only three coincident microphones, oriented 120° apart in azimuth, would have provided the same overall information. While suboptimal for QMX, such a four-channel source mix is optimal for the matrices BMX and TMX.

REFERENCES

[1] P. Scheiber, "Four Channels and Compatibility," *J. Audio Eng. Soc.*, vol. 19, pp. 267–279 (April 1971).

[2] J. M. Eargle, "Multichannel Stereo Matrix Systems: An Overview," *J. Audio Eng. Soc.*, vol. 19, pp. 552–559 (July/Aug. 1971).

[3] P. Scheiber, "Suggested Performance Requirements for Compatible Four-Channel Recording," *J. Audio Eng. Soc.*, vol. 19, pp. 647–650 (Sept. 1971).

[4] B. B. Bauer, D. W. Gravereaux, and A. J. Gust, "A Compatible Stereo–Quadraphonic (SQ) Record System," *J. Audio Eng. Soc.*, vol. 19, pp. 638–646 (Sept. 1971).

[5] L. Feldman, "On Matrix Quadraphonic Systems," *Audio Mag.*, vol. 55, p. 20 (Oct. 1971).

[6] C. Stark, "One Approach to Four-Channel Sound," *Audio Mag.*, vol. 55, p. 30 (Oct. 1971).

[7] R. Itoh, "Proposed Universal Encoding Standards for Compatible Four-Channel Matrixing," presented at the 41st Convention of the Audio Engineering Society (Oct. 1971).

[8] P. Scheiber, "Analyzing Phase-Amplitude Matrices," *J. Audio Eng. Soc.*, vol. 19, pp. 835–839 (Nov. 1971).

[9] T. Nakayama, T. Miura, O. Kosaka, M. Okamoto, and T. Shiga, "Subjective Assessment of Multichannel Reproduction," *J. Audio Eng. Soc.*, vol. 19, pp. 744–751 (Oct. 1971).

[10] A. Papoulis, *The Fourier Integral and Its Applications* (McGraw-Hill, New York, 1962), p. 42.

[11] G. A. Korn and T. M. Korn, *Electronic Analog and Hybrid Computers* (McGraw-Hill, New York 1964), p. 330.

[12] W. J. Albersheim and F. R. Shirley, "Computation Methods for Broadband 90° Phase Difference Networks," *IEEE Trans. Circuit Theory*, vol. CT-16, p. 189 (1969).

[13] Y. Makita, "On the Directional Localization of Sound in the Stereophonic Sound Field," *EBU Rev.*, pt. A, no. 73, p. 102 (June 1962).

[14] H. Mertens, "Directional Hearing in Stereophony —Theory and Experimental Verification," *EBU Rev.*, pt. A, no. 92, p. 146 (Aug. 1965).

[15] E. R. Hanson, "Musicassette Quadrasonic: Tape Record Compatibility," *J. Audio Eng. Soc.*, vol. 19, p. 24 (Jan. 1971).

[16] D. H. Cooper, "Time-Division Multiplex for Four-Channel FM," *J. Audio Eng. Soc.* (Letter), vol. 18, p. 198 (Apr. 1970).

[17] T. Inoue, N. Takahashi, and I. Owaki, "A Discrete Four-Channel Disc and its Reproducing System (CD-4 System)," *J. Audio Eng. Soc.*, vol. 19, pp. 576–583 (July/Aug. 1971).

THE AUTHORS

DUANE H. COOPER

TAKEO SHIGA

Duane H. Cooper was born in Gibson City, Ill., on August 21, 1923. He received the B.S. and Ph.D degrees in physics from the California Institute of Technology, Pasadena, in 1950 and 1955, respectively.

Since 1954 he has been with the University of Illinois, Urbana, where he is presently Associate Professor of Electrical Engineering and Physics. His research interests involve statistical and information theoretic problems in instrumentation and information transmission systems. In his consulting work, he has published papers in the areas of multichannel sound reproduction and distortion in disc-recording systems. He teaches courses in circuit analysis, dynamic systems, communications systems, random processes, and acoustics.

Dr. Cooper serves as referee for a number of professional journals. He is a member of the Acoustical Society of America, the American Physical Society, a senior member of the Institute of Electrical and Electronics Engineers, a Fellow of the Audio Engineering Society, and a recipient of that Society's Emile Berliner Award.

•

Takeo Shiga was born in Tokyo, Japan, in 1924. He received an engineering physics degree in 1946 and a doctorate in applied physics in 1962, both from Tokyo University. In 1946, Dr. Shiga joined Nippon Columbia Company Ltd., where he has been responsible for the development of loudspeakers, phonograph pickups and stereo phonographs. Lately, his studies have been devoted to stereophonic sound. At present he is the Director of the Acoustical Products Division. Dr. Shiga is a member of the Audio Engineering Society and the Acoustical Society of Japan.