COMMUNICATIONS

CRITERIA FOR EVALUATING SURROUND-SOUND SYSTEMS

MICHAEL A. GERZON

Mathematical Institute, University of Oxford, Oxford, England

Clear criteria are discussed for the evaluation and design of surround-sound recording and reproduction systems, free from any quadrifontal (four-source) assumptions. Some weaknesses in the quadrifontal approach are discussed, and difficulties in meaningfully testing systems are mentioned. It is observed that the ultimate aim of surround systems is to provide a good illusion of an intended encoded directional effect, and that this aim is not dependent on any particular choice of number of channels, number or position of loudspeakers, or on any particular method of originating surround program material. An appendix describes the similarities and differences between the resultant "kernel" system approach and the current matrix approach.

INTRODUCTION: In any field of engineering, progress is difficult if there is no clear statement of the aims to be achieved. The field of surround-sound recording, transmission, and reproduction has been particularly handicapped by a confusion of the desired end (the reproduction of an illusion of all directions around a listener) and particular so-called "quadraphonic" means of achieving that end. This paper is intended to present a clear statement of various possible aims in designing and evaluating surround-sound systems, and a description of known faults in that approach which treats just four independent sources [which approach is conveniently termed "quadrifontal" (four-source)]. It is not claimed that any particular observation in the following is new (for example, see [1]–[8]), but it is thought that a clear statement of aims will be helpful to those involved in taking decisions in this area. It is emphasized that the following comments are not merely based on a theoretical analysis, but have been extensively checked by practical experience in the development of surround-sound systems in connection with the ambisonic project of the U. K. National Research Development Corporation (N.R.D.C.).

While this paper is self-contained, it is also intended as a general introduction to a projected series of papers on surround-sound system design under the title, "The Rational Systematic Design of Surround-Sound Recording and Reproduction Systems."

KERNEL SYSTEMS

We start by assuming that the primary aim of any surround-sound system is to produce in the ears and brain of a listener the illusion of an intended pattern of directional sound. Such systems may be designed either to handle sounds just in the horizontal plane, or to handle a full sphere of directions ("periphony" [9]). In addition, such systems may be designed either to reproduce sounds at a single apparent distance from the listener, or to produce the illusion of sounds from any distance from the listener.

There are at least two steps required in satisfying such an aim, encoding and decoding.

Encoding is the process of assigning to every given direction of sound a way of including this sound among the n available recording or transmission channels used to convey the information. The method of inclusion is usually achieved by assigning the sound to each of the n channels with a different and stated gain (which may be real or complex) which is a function of the intended encoded direction. It is convenient to term systems in which the gains of the channels are expressed as a smooth and continuous function of direction, "kernel system," since such systems are mathematically described by "kernels" in the same way as so-called "matrix systems" are described by matrices (see Appendix). Examples of systems whose encodings have been specified in kernel form are the UMX systems [3], the RM system [10], and various periphonic sytems [9].

Decoding is the process of deriving from the encoded transmission channels signals suitable for feeding a stated loudspeaker layout so as to produce an approximation to the illusion of the intended encoded directional effect. Two points here are especially worthy of note. First, the decoding apparatus has to produce loudspeaker feed signals satisfying relevant psychoacoustic criteria at the ears of the listener, so that the design of decoders is primarily an engineering task dependent on a knowledge of what acoustic stimuli are best capable of producing the desired illusion. It by no means follows that an arbitrary choice of "ideal" loudspeaker feed signals will necessarily provide the best illusion. Second, the desired directional illusion does not depend upon where the loudspeakers happen to be placed, and it is thus necessary to vary the loudspeaker feed signals (and hence the decoder design) according to the size and shape of the loudspeaker layout chosen by the user. It is certainly not to be expected that the same loudspeaker feed signals will produce the same directional illusion for many different shapes of loudspeaker layout.

THE MAGIC NUMBER FOUR

It will be noted that the technical problems of surround sound as stated above have not mentioned the number four. The reason is that in general this has no obvious advantages in any application over other numbers. In typical laboratory work (and this is expected to be true also of much consumer equipment in the future) the original sound is recorded on three channels, transmitted through two or three channels, and decoded through five or six loudspeakers.

The number four does arise naturally in three ways in surround-sound systems.

1) The number of walls and corners of many domestic listening rooms is four, so that this number of loudspeakers may often be convenient for horizontal-only reproduction.

2) The minimum number of loudspeakers in a horizontal plane array that give acceptable (although not ideal) surround sound is four.

3) The number of channels required to capture the pressure and velocity components of a with-height periphonic sound field is four.

Case 3) is not relevant to horizontal-only sound. There is, of course, no reason why the number of channels should be the same as the number of loudspeakers. Compelling reasons to the contrary may be listed as follows.

1) Based on low- and mid-high-frequency psychoacoustic criteria discussed in [4], one can prove a mathematical theorem that the various psychoacoustic criteria relevant to localization of sound will not be simultaneously optimized via a rectangle or square of four loudspeakers unless the number of nonredundant information channels feeding the decoder is not more than three. This rather surprising result is now well and widely confirmed by laboratory data and by theory (see, for example, [11, Figs. 17 and 20] and [4]). Loosely speaking, the effect of a fourth nonredundant channel of information fed to the loudspeakers is to create a detent effect whereby sounds in interspeaker directions are pulled toward the nearest loudspeaker, and to destabilize the localization of sounds between loudspeakers, especially at the sides of a listener. The result is to emphasize the four loudspeaker directions at the expense of all other directions. Similar considerations show that with-height reproduction from four channels is optimal only through six or more loudspeakers.

2) Domestic listening rooms vary widely in size and shape, and it is impossible to fit a single "standard" loudspeaker layout in even a small proportion of such rooms. As a result, the loudspeaker feed signals must vary in order to give the best results with different layout shapes, and it is thus not possible to transmit signals guaranteed to be optimal loudspeaker feed signals for most situations. Thus there can be no one-to-one correspondence between transmitted signals and loudspeakers.

3) For a given number of channels it is found that (provided the encoding specification is a suitable kernel specification) the more loudspeakers are used to reproduce the encoded signal, the better the results. (This is expected to fail with very large numbers of loudspeakers, but such numbers are in any case not practicable.) For example, it has been found that reproduction of three channels (via a suitable decoder) through five loudspeakers can give a significant improvement over reproduction via four loudspeakers. In particular, the "bispectral detent effect" whereby highly asymmetric sound waveforms (such as clapping, oboes) are pulled toward the loudspeakers is absent with five loudspeakers in a suitable layout. Thus it is important to leave the option available of designing decoders for loudspeaker layouts of various shapes and complexity, and this option should not be preempted by an overrestrictive choice of encoding specification.

MATRIX SYSTEMS AND THE QUADRIFONTAL APPROACH

There is, of course, a well-known alternative approach to surround sound, which we call the "quadrifontal" (four-source) approach. We avoid the word "quadraphonic" often used in connection with this approach (for example, see the title of a standard collection of references [12] on surround-sound systems), since different authors use the term in different meanings, (for example: four-speaker, four-channel, derived from four-track tapes, derived from four-output mixing desks, derived from four microphones, etc.).

In the quadrifontal approach the starting point is to assume that the desired information to be transmitted to the listener is four distinct source signals. The aim of all quadrifontal systems is to reproduce through four loudspeakers in the listener's room a simulation of the effect of the four distinct sources. This may be done in two ways. One is to provide four transmission channels to the listener (an approach termed "discrete") and the other (termed the "matrix" approach) is to matrix the four signals into two channels and to provide the listener with a device (termed "logic" or "variable matrix") that ensures that when only one source is transmitted, the loudspeakers corresponding to the other three sources are substantially unactivated. Such "logic" devices cannot, of course, reproduce all four sources simultaneously, each coming only from its assigned loudspeaker, when all sources are transmitted together.

In practice, sounds are often shared between the four source channels of quadrifontal systems in order to provide some degree of illusion of nonspeaker directionality. Two methods of doing this are widely used. One is pairwise mixing, and the other is the use of spaced microphone arrays (with spacing ranging from a few centimeters to several meters).

Pairwise mixing attempts to assign sounds to interspeaker directions by feeding sounds in phase to only the two loudspeaker source channels adjacent to the desired direction, the relative intensity depending on the desired direction. Such pairwise mixing has fared almost universally badly in experimental tests on localization behavior [13], [11], [14], [15]. If reproduced through a square loudspeaker layout, it is found that the front-stage images are unstable and suffer from the hole-in-the-

middle, as well as a pronounced elevation or in-the-head effect [13]–[16]. This is because of the overwide 90° angle subtended by the loudspeakers, which has long been known to give bad results in stereo. The side images are extremely unstable, tending to jump forward or backward to the corners, especially if the listener allows any small movement of his head. Noncentral listeners find that the apparent sound directions tend to be pulled toward the loudspeakers to which they are nearest, and also images tend to be drawn toward the loudspeaker the listeners are facing when they do not face forward.

Most "logic" or "variable-matrix" devices are designed to make the reproduction of material with only single sources activated behave like pairwise mixed quadrifontal material, since such a method of directional encoding of sound is often regarded as a reference standard.

It is difficult to make general comments about the use of spaced microphone quadrifontal material, since techniques vary widely. However, widely spaced (>1.5 m) microphones are well known in stereo to produce images where direct sounds are largely confined to the loudspeakers.

With smaller microphone spacing (5–50 cm), at low frequencies of sound, the wavelength of sound is large and the microphones thus are effectively coincident. Such coincident techniques define a kernel encoding in the sense discussed earlier, that is, each sound is assigned to the transmission channels with gains smoothly varying with direction. However, with the wide range of microphone techniques in use, such kernel encoding will vary in an arbitrary manner from recordist to recordist. At high frequencies (with wavelength small compared to microphone spacing), the microphones act as independent incoherent sources. Thus quadrifontal assumptions apply only at higher frequencies, with a largely random kernel encoding at lower frequencies. There is no reason to expect that the microphone characteristics considered "optimal" in the high-frequency quadrifontal region should be the same as those considered "optimal" in the kernel encoding region, and this is indeed almost never the case with undesigned microphone array systems.

The quadrifontal approach is also expected by many [17] to feed several different shapes of loudspeaker layout. This is not unreasonable if what emerges from each loudspeaker is a distinct sound source having no component in common with the other three. Such "truly quadrifontal" material reproduces as four isolated sounds, and listeners may find it worthwhile to position these to their own taste. However, once related sounds appear in two or more loudspeakers, the directional results will vary with loudspeaker positioning, and the recording industry standard for reproducing such "shared quadrifontal" material (such as pairwise mixed material) has seemingly been chosen to be the square loudspeaker layout. Anyone using other layouts will not normally hear the effect heard by the record producer.

It is not evident that the best way to reproduce the encoded directional effect of pairwise mixed or of spaced microphone quadrifontal material is to feed the four source signals straight to four loudspeakers in a square. The "decoder" this implies is of crude design, and it is possible that a more elaborate decoder adapted to the properties of a particular method of recording (and possibly using a different loudspeaker layout) may give more accurate directional reproduction. In fact, this is the case. The difficulty is that to use a different decoder for each recording style is impractical for the consumer. Thus it is suggested that the circuitry required to produce signals suitable for feeding consumer decoders be, as far as possible, implemented as a part of the recording apparatus. Thus, for example, a spaced microphone technique would include the apparatus (a frequency-dependent matrix) required to make its low frequencies match a standard kernel encoding method, while treating its high frequencies as independent sources. Such an approach is no longer quadrifontal.

Thus it will be seen that hoping to treat a wide range of material derived in many different ways as four independent sources is unrealistic. As soon as interrelationships occur between the sources, one must ask the engineering design question of whether the relationships occurring have been designed to give the best possible results. If not, the design should be changed.

The most immediate conclusion of the above is that on no account should pairwise mixed material be used as a standard of reference, since its results are extremely poor. An analogy that may be helpful here is to note that in 1953 the choice of NTSC color TV standards was not based on getting the best results from 1953 Kodachrome film, but was based on getting the best results (within inevitable technical constraints) from a knowledge of the capabilities of human color vision. As the best modern color TV equipment shows, such a standard has permitted continual substantial improvement in color quality. In a similar way, standardization on pairwise mixed encoding as the basis of choice of any system is extremely unwise without optimizing designs of encoding (or "panning") according to the best available knowledge of human directional sound localization.

ENCODING METHODS

The problem of choosing encoding methods for sound field involves a number of constraints. One of these is that two of the encoded channel signals should be suitable for conventional L and R two-speaker stereo reproduction, and in addition the signal L + R should be suitable for monophonic reproduction. The author has considered elsewhere some of the possible compatibility criteria [18]. Clearly there can be no unique "best" compatibility choice, since different users will place different weight on different aspects of compatibility. The author believes that any system designed for universal use should not fail dramatically when fed with *any* current recording philosophy, and it is notable that most marketed systems obtain their advertised excellence with some recording philosophies only at the expense of failure or unusability with other philosophies. Among the philosophies currently in wide use are quadrifontal four-corner recording

+ front center, 360° pan-potted sound-stage recordings (especially in pop), front-stage recording + reverberant "splash" from the rear (especially for classical music), spaced microphone recording, and finally the transduction of the directional sound field at or near a point [19], [20].

The last of these philosophies clearly has a special place insofar as not only is it widely used for classical music recording and broadcast drama, but it is also the only philosophy amenable to objective testing of the soundness of the overall system, since it enables testing of the reproduced illusion versus an original sound field, that is, live sounds.

No system of transmitting and providing information can be designed rationally or be well-engineered unless there is a precise specification of how the information to be transmitted is represented in the transmission channels. A precise encoding specification for surround sound is, as we have observed, necessarily a kernel specification giving precise instructions as to how each spatial direction of sound is to be handled. Besides good mono and stereo compatibility and the capability of handling many recording philosophies, such a kernel specification of encoding should satisfy the following other constraints.

1) Suitability for transmission with high quality via most or all high-quality domestic entertainment media, including disc, FM radio (according to the different American, West European, and East European multiplex stereo standards), cassettes, and (possibly) videodisc. Otherwise chaos will result if, for example, discs cannot be broadcast or broadcasts cannot be recorded on cassette.

2) Capability of being decoded with accurate illusion of the encoded direction effect. This requirement is obviously crucial. It breaks down into several more detailed requirements.

a) Capability of being decoded with convincing and intended directional effect and lack of listener fatigue under prolonged listening without using signal-actuated variable matrixing. This requirement ensures that complex sound fields involving many different sources (including live reverberant sound fields) can always be reproduced without "pumping." Also, insofar as signal-actuated matrixing might provide an enhanced illusion of directionality for isolated sounds, any side effects of such variable matrixing are likely to be less audible if the basic fixed decoding is good in its own right. It is bad engineering to design a system to be bad, and to attempt to solve problems by a subsequent "fix," although such "fixes" may be acceptable for some applications to improve some aspects of a system that have already been optimized within available technical constraints.

b) Capability of decoding a good illusion of the encoded directional effect via loudspeaker layouts that fit a wide variety of shapes and sizes of domestic settings, and not just a narrow range of laboratory or studio monitoring arrangements.

c) Reasonable tolerance of normal inaccuracies in studio processing equipment, the recording or transmission chain, and in high-quality (but not state-of-the-art) consumer equipment. This is not to say that equipment not designed for surround sound (such as, many existing stereo tape recorders, cassettes, microphones, or loudspeakers) should always work well in surround sound (an impossible requirement), but only that manufacture, maintainance, and operation of equipment should not involve unreasonable care or expense.

3) The possibility of converting existing quadrifontal material into some reasonable approximation of the kernel encoding specification. As already explained, it cannot be expected that all quadrifontal material will convert to the correct encoding exactly, or even that a single conversion apparatus will give an optimal approximate conversion for all quadrifontal material. Nevertheless, reasonably convenient encoding means for getting moderately acceptable results from quadrifontal material must be possible. (There is the possibility that the conversion means may in some cases actually result in better results than obtained from simple quadrifontal four-speaker reproduction.)

4) The encoding system should be such as to provide the option of acceptable consumer decoders covering a wide range of cost and quality. This requirement, while not essential, undoubtedly will be helpful to recording companies, since a cheap-end consumer market would justify commercially the extra efforts involved in surround programming, while the existence of state-of-the-art consumers will help to maintain the impetus toward improving the quality of recorded programs.

5) The encoding specification must be realizable at reasonable cost for a wide range of recording philosophies, and with good results in as many of these philosophies as possible.

6) The encoding specification must leave room for both minor and major developments of the art without incompatibility or obsolescence. In particular, it must not unduly hinder use for conveying a wide range of artistic ends, ranging from concert hall live recording and live drama to purely electronic music performed or composed with spatial effect as a fundamental part of the musical language. Technically, the system should be capable of extension to include recording of the effect of sound-source distance and eventually to full-sphere with-height with-distance three-dimensional space recording. In fact, such recordings have been made and are already routinely possible in the laboratory [19], and a wide range of decoding apparatus suitable for with-height with-distance reproduction suitable for domestic settings is now known.

PRODUCING THE ENCODED MATERIAL

Requirement 5) needs some elaboration, since it is not at first clear how sound suitable for a kernel encoding specification can be produced. Just as the NTSC color TV specification would be useless if it were not possible to design color TV cameras and flying-spot scanners producing suitable R, G, B signals (that is, red, green, blue) from the original scene, so we must have means of sound signal production from our raw sounds capable of being processed to meet the encoding specification. In the color TV case, R, G, B signals are themselves not the final NTSC color signal, but they are a natural first stage in the production of the NTSC signal (and equally of the signals

of other color systems such as PAL and SECAM). In a similar way, it turns out that operationally, the simplest signal production apparatus for surround sound need not necessarily be in the consumer's kernel encoding specification (whose production will usually involve elaborate phase shifting circuitry and often band-limiting circuitry as well). Also, in a situation in which it is not clear yet whether a universal encoding system will be adopted, it is wise to produce studio signals capable of being matrixed into all well-behaved kernel systems. We have seen that pairwise mixing is unsuitable for this role as the "studio kernel specification" (which we shall term "studio format").

Whichever kernel encoding specification is adopted for either studio or consumer use, it is evident that in principle a suitably complex joystick arrangement connected to a sufficiently complex arrangement of potentiometers in a complex circuit can provide a means of positioning a single monophonic input into any desired encoded position (at least for horizontal encoding). For reasons of economy and reliability, it is desirable that such pan pots should not be too complex. The author has described elsewhere [8b] a simple pan-pot design for horizontal sound capable of producing a studio-format kernel-encoded signal that is actually simpler than a pairwise pan pot; many alternative designs are possible.

The studio format should avoid use of phase-shift circuitry (because at studio quality this is expensive), and should be amenable to a wide range of processing with the simplest possible circuitry. In addition, it should be robust under tape recording, that is, it should not be subjectively too badly affected by tape azimuth errors, dropouts, or mild noise reduction system mistracking. It has been found that any format representing direct loudspeaker feed signals does not meet this robustness condition adequately unless a quite low standard of directional reproduction is tolerated, and that the studio format discussed in [8b] is much better in this respect. Other studio formats may also be considered.

Providing an effective surround-reverberation effect is not easy. Treating distinct reverberation unit outputs as monophonic sources does not provide the desired directional continuity [23] of the reverberation characteristic of good surround reproduction. A kernel specification should ideally permit the design of reverberation devices providing a uniform continuous distribution of reverberant sounds meeting that specification. In the studio format it is also desirable to have means of rotating a whole encoded sound field and not just monophonic sounds.

Finally, it must be possible to design microphone systems capable of accurately encoding live sound fields to the kernel specification. One means of doing this for a range of kernel encoding specifications has been described elsewhere [19]. Note that even for horizontal sound reproduction, such microphones are inevitably exposed to reverberant sound from nonhorizontal directions, and should be designed so that the effect on horizontal reproduction of these nonhorizontal sounds is subjectively uncolored and as natural as the lack of height and system design constraints will allow.

DECODERS

We have already talked about decoders extensively, but it is necessary to state some aspects of the engineering design and evaluation problems that they involve, as well as clearing up some myths.

As we have commented, kernel-encoding systems should be designed to give the best possible decoded illusion through the best possible decoder for that system. The optimization problem is thus double: for each possible encoding system we have to optimize the decoder, and then we have to find out of all possible encoding systems the one that gives the best decoder. Stated this way we have an impossible task, since we do not know what "best" means, but it does emphasize that the design of systems starts not with the choice of a kernel encoding specification, but with a consideration of the ideal stimuli at the ears of a listener. Papers have been published [4], [24], [25] dealing with mathematically tractable criteria for sound localization, and in a future paper we hope to present general mathematical methods for handling directional psychoacoustics incorporating most existing theories; much of this work is implicit in the nonmathematical account of [4].

However, just as a good encoding should not fail dismally with any coherent philosophy of recording, so it is desirable that a good decoding method should not be too critically dependent on any highly specific assumptions about the ears, the position of the listener, or the position or number of his loudspeakers. As suggested in [4], the ears appear to use a large number of methods of sound localization and to take a "majority decision" as to the apparent sound position. Under these conditions, any particular cue for localization can be removed without affecting localization provided a sufficient number of other cues remain correct. Thus a primary rational design aim in decoders should be to satisfy as many as possible of the cues used by the ears. Such a design will clearly be more "robust" under abuses (such as, chairs in front of a loudspeaker, component tolerances, or distortion) than designs based on just one assumption. Experience bears out this assertion, and theoretical designs maximizing the number of correct cues have invariably "worked first time."

In this work (which will be reported in detail in future papers, but see [8b]) it has been found that many of the assertions made about sound localization are myths based on inadequate interpretation of evidence. For example, it is often asserted that "bass frequencies are not important for sound localization." This is based on experiments [26] in stereo that show that if the bass of the stereo L - R signal is removed, the position of sounds remains unaltered. However, we have already noted that cues used by the ears can be omitted yet correct localization will be heard *if* other cues are still present. In experimental work it has been found that sounds nominally positioned at the side of a listener in a square loudspeaker layout are *not* heard as at the side unless the bass frequencies are modified to satisfy criteria discussed in [4], [24]; a modification affecting frequencies below 500 Hz only (see [8b]) is sufficient to

produce a very strong side localization. In this case, the bass provides cues that otherwise are absent [14]. In another experiment, tympani and double bass sounds have been noticeably displaced from their "correct" positions by modifications affecting only frequencies below 200 Hz significantly. In a similar way, it is believed than any assertion that any particular range of frequencies "is not needed" is suspect, as is any assertion that satisfying one theory of sound localization is either "necessary" or "sufficient" for reliable sound positioning.

For domestic use it is desirable that decoders should give a reasonably accurate directional illusion not only for a forward-facing central listener, but for a listener facing in other directions and for a listener sitting away from the center. Clearly, one would not expect a system using only two transmission channels to do this over such a large listening area as a system using three channels, but the results should nevertheless be domestically usable. As important as the correct illusion of directionality are other qualities of the sound. Is the localization sharp or diffuse? Is the image single or double? Is it in-the-head or elevated? Is the bass quality clean or lumpy [13]? Is the treble quality clear or harsh? When two sounds in different directions occur together, are they both well located in their respective directions? Is there any sensation of "pumping"? Is the ambience uniform around the listener, or is there a "tunnel" in one particular direction? Is there any front/back ambiguity? Most important of all, when listening to music, does listener fatigue set in, or does the sound have an unobtrusive quality that makes one forget the technical means of reproduction? This last question, of course, is ultimately what the system is used for, and it is surprising how many systems tested fail it.

Insofar as perfect reproduction cannot be obtained, it is desirable that any inevitable compromises should be biased toward improving the effect heard by a central forward-facing listener, but this step should only be taken once one has done as well as one can for the "general" listener at arbitrary positions. It seems obvious to the author that surround sound cannot be viable unless and until it is better than stereo for sounds in the front quadrant of positions, and it would seem that failure to meet this minimal requirement is punished by the reaction of the public in the market place.

As emphasized earlier in this paper, decoders should be designed for a variety of loudspeaker layouts suitable for use in the home. Designs have been published suitable for regular polygon loudspeaker layouts [3], regular polyhedron layouts [9], and perhaps more realistically, rectangle layouts with arbitrary front/side ratio [8b]. Other irregular layout designs are perfectly possible, such as for trapezia.

We remark that an optimal decoder can be a relatively sophisticated apparatus, being essentially a layout-adjustable frequency-dependent matrix [8b].

THE TESTING OF SYSTEMS

It is very important to realize that any particular design of decoder is not a part of an overall encoding system specification. Since the job of a decoder is to provide an optimum illusion of the encoded directional effect, it is possible that improved knowledge of psychoacoustics or improved engineering may permit improvements in the future that should not be constrained. However, as in the case of color TV receivers, it is not possible to optimize the decoding equipment unless the precise nature of the encoding specification followed in transmission is known to the designer. Some experimental tests of quadrifontal and surround systems have tended to ignore this and to treat a particular decoder design as an essential feature of the transmitted (encoding) system.

It is clear that not all encoding systems will permit the design of decoders giving equally good results, and as far as is possible, the choice of encoding system should be based on the best possible results obtainable with the "best" possible decoder. The snag here is that different users have different criteria of what is "best," and there is a perfectly serious argument for providing different users with different decoders adapted to their particular criteria of goodness. A practical problem here is, of course, to ensure that the user is indeed provided with his own "best" decoder, but these considerations do show how fraught with difficulty is any comparative investigation between surround systems. Additional difficulties arise in such tests if equipment designed to meet the correct kernel encoding specification of each system is not used [27], [28], especially when material of known badness (notably pairwise mixed material) is used in the evaluation of all systems. If a given surround system becomes widely used, it is clear that the associated studio equipment will be adapted to get the best out of that system, and one can only evaluate the relative merits of each system by comparing the best they can do with each of a range of recording philosophies (with the respective equipment for that philosophy suited to each individual system). No such test has yet been performed anywhere, to the author's knowledge.

CONCLUSIONS

Surround-sound systems have to be considered as complete systems from the initial means of sound production and handling in the studio right through to the ears and brain of a listener casually seated in his own oddly shaped listening room. The number of domestic, technical, psychoacoustic, commercial, and artistic constraints on a system is considerable, and only a balanced compromise involving all these constraints can hope to be fully viable. The philosophy has been expounded in this paper that a system should be designed to give reasonable results with a wide range of legitimate needs, rather than to be optimized according to the arbitrary choice of any particular need at the expense of others. The precise balance of choices within such a "wide-range compromise" is open to legitimate argument, but seems to exclude a narrow "quadrifontal" approach.

It has been observed that the "quadrifontal" (four-source) approach is irrelevant to the engineering design problems and evaluation of surround-sound systems. Indeed, it is positively misleading since it leads to an

incorrect statement of aims, which can be illustrated by asking "have you ever been in a 4-channel concert-hall?"

In practice it has been found that a systematic design of the whole chain—from microphones, through studio equipment, encoding methods, to decoders—based on the psychoacoustic, technical, commerical, domestic, and artistic aims discussed in this paper leads to results that are predictable, consistent, and far better than any achieved with systems in which parts of the chain are left to chance. There is, of course, no unique technological answer to any of the individual design problems involved, and it is hoped that the statement of aims in this paper will be helpful to other designers who may have been confused by the logical contradictions implicit in the quadrifonfal approach as practiced. Some details of work based on aspects of the philosophy of this paper is to be found in [3], [4], [8b], [18], and it is hoped to publish further detailed papers on the design of decoders and studio equipment in the future.

APPENDIX

MATRICES AND KERNELS

This Appendix is intended to help clarify mathematically the concept of a kernel encoding system and its relation to matrix systems.

First we define the relevant mathematical concepts. Recall that a "matrix" a_{ij} consisting of $m \times n$ numbers is a method of relating n quantities b_j ($j=1,2, \cdots, n$) to m quantities c_i ($i=1,2, \cdots, m$) via the equation

$$c_i = \sum_{j=1}^{n} a_{ij} b_j. \tag{1}$$

The mathematical notion of "kernel" is the analogous idea when an infinite number of quantities occur. Suppose that we have a function $b(y)$ on a space of variables y on which a notion of integration is defined (such as the circle of direction azimuths around a microphone, with the integration

$$\int_{-\pi}^{\pi} f(y)\,dy,$$

or the sphere of directions around the listener, with the integration being with respect to the surface area). Let $c(x)$ be a function on a second space (for example, a line, circle, sphere, or even just four discrete points in which last case x takes the values 1,2,3,4). Then a kernel $k(x,y)$ defines a relationship between the functions $b(y)$ and $c(x)$ via the equation

$$c(x) = \int k(x,y)b(y)\,dy \tag{2}$$

where $\int \cdots dy$ is the relevant notion of integration.

It will be seen by comparing Eqs. (1) and (2) that kernels are analogous to matrices except that summation is replaced by integration. Kernels are relevant whenever an infinite number of quantities ($b(y)$, one for each of an infinite number of y's) occur. In the real world of sound recording, the number of channels associated with a live sound field is indeed infinite (although it is true that in theory a reasonable practical approximation to such a field can be achieved using only around a million channels—which is still infinite for present-day practical purposes).

There are two approaches in handling kernel systems with actually or potentially an infinite number of input channels. The first, adopted by Cooper and Shiga [3], is to imagine initially an idealized continuous circle (or sphere [9]) of loudspeakers for reproduction, and to attempt to relate the whole process of encoding to decoding as a relationship between two functions on the circle (or on the sphere). Thus, for example, Cooper and Shiga consider the variable y to be the angle θ of arrival of an azimuthal sound, and the variable x to be the angle ϕ of a loudspeaker through which reproduction occurs. In that approach the input sound field occurs as a function S_θ of the arrival angle, the sound P_ϕ is fed to the loudspeaker at azimuth ϕ, and the kernel equation is of the form

$$P_\phi = \int_{-\pi}^{\pi} k(\phi,\theta) S_\theta\,d\theta \tag{3}$$

where the kernels $k(\phi,\theta)$ have the form $k(\phi,\theta) = 1+e^{-j(\phi-\theta)}$ for BMX, $1+2\cos(\phi-\theta)$ for TMX, etc. Cooper and Shiga lay great emphasis on the rotation symmetry properties of $k(\phi,\theta)$ (namely, that $k(\phi,\theta)$ may be written in the form $k(\phi,\theta) = a(\phi-\theta)$ where $a(\psi)$ is a function such that $a(-\psi) = a(\psi)^*$, * being complex conjugation). There is no obvious reason why such rotational symmetry (also considered in [9] for the sphere) should lead to the best system.

The second approach recognizes that an infinite number of loudspeakers may not always be a suitable idealization for decoders, and takes the variable x to vary over a number of values $i=1,2, \cdots, m$ associated with m loudspeakers, in which case, the equation takes the form

$$c_i = \int k_i(y)b(y)\,dy \tag{4}$$

which lies about halfway between the matrix equation (1) and the full kernel equation (2). It is a matter of psychoacoustic theory to design suitable kernels $k_i(y)$ (which may vary with frequency) for a particular layout of m loudspeakers.

In practice, it is convenient to break up a kernel system (4) into at least two stages. The first stage is to encode the sound onto n channels, resulting in a kernel encoding equation ($j=1, \cdots, n$),

$$d_j = \int K_j(y)b(y)\,dy \tag{5a}$$

describing how directional sounds $b(y)$ are converted to the n channel signals d_j. Usually the kernel $K_j(y)$ is frequency independent, since frequency-dependent encoding can lead to poor mono and stereo compatibility, and may complicate the design of some decoders. The second stage (decoding) is to derive m loudspeaker feed signals c_i ($i=1, \cdots, m$) via a matrix equation:

$$c_i = \sum_{j=1}^{n} a_{ij} d_j \tag{5b}$$

where the matrix a_{ij} depends on the encoding system, the loudspeaker layout, and may be frequency dependent, or even (in the case of signal-actuated matrices) be dependent on frequency *and* the form of the n-channel signal.

The overall encode/decode kernel $k_I(y)$ of Eq. (4) is given from Eqs. (5a) and (5b) by

$$k_I(y) = \sum_{j=1}^{n} a_{ij} K_j(y). \tag{6}$$

It will be seen that kernels need only occur in the encoding process; such kernel specifications may be implemented in practice by suitably designed pan pots [8b] (which vary the n gains $K_j(y)$ as the pot setting y varies), or by suitably designed sound-field microphones [19] which transduce directly n channels from the infinite number of acoustical channels of the original sound field. The rest of the system (studio processing, reencoding to a consumer coding specification, and decoding) is described by matrix equations such as Eq. (5b), and so may be implemented by fairly conventional matrix circuits. Thus in terms of hardware and software, kernel systems in practice largely resemble matrix systems; the main difference lies in the possibility of rationally designing these matrices in terms of what they do to a continuous (infinite-channel) original sound field, and thus of getting optimal results for all sound directions.

There are cases when the idealization used in [3] and [9] and Eq. (3) of an infinite number of reproducing loudspeakers (that is, kernel *de*coding) is useful. It turns out that provided the real-world loudspeakers used lie on a sufficiently regular configuration (such as square, regular polygon, regular polyhedron), then the loudspeaker feed signals associated with the directions of each loudspeaker according to the kernel equations (2) or (3) do indeed give a good approximation to results of the ideal circle or sphere of loudspeakers. However, once more irregular loudspeaker layouts are used (such as rectangles [8b]), the correct loudspeaker feed signals are not given by the kernel decoding equations, and more elaborate methods of decoder design must be used, which we shall describe in detail in a later paper or papers.

ACKNOWLEGMENT

The author would like to thank Professor Peter Fellgett for helping develop these ideas, members of the BBC research department for discussions that helped clarify the concepts described herein, and the British NRDC for their support during this work.

REFERENCES

[1] M. A. Gerzon "What's Wrong with Quadraphonics?," *Studio Sound*, (5) pp. 50, 51, 56 (May 1974).

[2] B. J. Shelly "Quadraphonic Quandry," *Wireless World*, vol. 80, pp. 235-236 (July 1974).

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

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

[5] P. B. Fellgett, "The Ambisonic Surround-Sound System, Pts. 1 and 2," *The Gramophone*, vol. 53, pp. 1266, 1269 and 1397-1398 (Jan. and Feb. 1976).

[6] J. R. Ashley, "On the Psycho-Acoustic Basis for Two- and Four-Channel Home Music Systems," presented March 2, 1976, at the 53rd Convention of the Audio Engineering Society, Zurich, Switzerland, preprint B-5.

[7] J. Eargle, "Beyond Quad," presented May 17, 1973, at the 45th Convention of the Audio Engineering Society, Los Angeles, preprint 921.

[8](a) P. B. Fellgett "Ambisonics; Pt. I, General System Description," *Studio Sound*, (8), pp. 20-22, 40 (Aug. 1975);

(b) M. A. Gerzon, "Ambisonics; Pt. II, Studio Techniques," *Studio Sound*, (8) pp. 24-26, 28, 30, (Aug. 1975). Correction: *Studio Sound*, (10), p. 60 (Oct. 1975).

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

[10] "The Sansui QS Regular Matrix System and a New Technique to Improve Interchannel Separation Characteristic," QS Regular Matrix System Tech. Anal. D1 (Distributed by Sansui Electric Co. Ltd.).

[11] O. Kohsaka, E. Satoh, and T. Nakayama, "Sound-Image Localization in Multichannel Matrix Reproduction," *J. Audio Eng. Soc.*, vol. 20, pp. 542-548 (Sept. 1972).

[12] "Quadraphony, an Anthology of Articles in the JAES," Audio Eng. Soc., New York, 1976.

[13] P. A. Ratliff, "Properties of Hearing Related to Quadraphonic Reproduction," BBC Research Dept., Rep. BBC RD 1974/38 (Nov. 1974).

[14] P. Damaske and Y. Ando, "Interaural Crosscorrelation for Multichannel Loudspeaker Reproduction," *Acustica*, vol. 27, pp. 232-238 (1972).

[15] G. Theile and G. Plenge "Localization of Lateral Phantom Sources," presented March 2, 1976 at the 53rd Convention of the Audio Engineering Society, Zurich, Switzerland, preprint B-5.

[16] K. de Boer, "A Remarkable Phenomenon with Stereophonic Sound Reproduction," *Philips Tech. Rev.*, vol. 9, pp. 8-13 (1947).

[17] B. B. Bauer, "From Stereo to SQ," *Elektor 17*, vol. 2, pp. 9-34 to 9-38 (Sept. 1976).

[18] M. A. Gerzon, "Compatible 2-Channel Encoding of Surround Sound," *Electron. Lett.*, vol. 11, pp. 615-617 (Dec. 11, 1975).

[19] M. A. Gerzon, "The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound," Collected papers, presented March 6, 1975, at the 50th Convention of the Audio Engineering Society, London, England.

[20] T. Yamamoto, "Quadraphonic One Point Pickup Microphone," *J. Audio Eng. Soc.*, vol. 21, pp. 256-261 (May 1973).

[21] M. A. Gerzon, "A Geometric Model for Two-Channel Four-Speaker Matrix Stereo Systems," *J. Audio Eng. Soc.*, vol. 23, pp. 98-106 (Mar. 1975).

[22] B. B. Bauer, "Directional Ambiguity of Quadruphonic Matrices," *J. Audio Eng. Soc. (Letters to the Editor)*, vol. 19, pp. 315-316 (Apr. 1971).

[23] M. A. Gerzon, "Recording Techniques for Mul-

tichannel Stereo," *Brit. Kinematography, Sound & Telev.*, vol. 53, pp. 274-279 (July 1971).

[24] B. Bernfeld, "Simple Equations for Multichannel Stereophonic Sound Localization," *J. Audio Eng. Soc.*, vol. 23, pp. 553-557 (Sept. 1975).

[25] M. Nishimaki and K. Hirano, "Localization of Sound Sources in 4-Channel Stereo," QS Regular Matrix System Tech. Anal., D3 (Distributed by Sansui Electric Co. Ltd).

[26] D. S. McCoy, "Distortion of Auditory Perspective Produced by Interchannel Mixing at Low and High Frequencies," *J. Audio Eng. Soc.*, vol. 9, pp. 13-18 (1961).

[27] T. W. J. Crompton (based on work by), "The Subjective Performance of Various Quadraphonic Matrix Systems," BBC Research Dept., Rep. BBC RD 1974/29 (Nov. 1974).

[28] J. G. Woodward, "NRQC Measurement of Subjective Aspects of Quadraphonic Sound Reproduction, Pts. I and II," *J. Audio Eng. Soc.*, vol. 23, pp. 2-13 and 128-130 (Jan./Feb. and Mar. 1975).

About the Author:

Michael A. Gerzon received an M.A. degree in Mathematics at Oxford University in 1967. Since completing postgraduate research in axiomatic quantum theory at Oxford University, he has been involved in research into the properties of linear and nonlinear multichannel systems, including analytic system theory.

One aspect of this work has been a study of surround sound recording and reproduction systems. This includes the abstract mathematical theory of these systems, the study of mathematical models for non-directional and directional psychoacoustics, and the design of microphone arrays, studio processing equipment and decoders based on these studies. Since the start of 1974, much of this work has been carried out in connection with the British N.R.D.C. (National Research Development Corporation) and its 'ambisonic' systems of surround sound.

Mr. Gerzon has published over 20 papers and articles on the theory and practice of surround sound systems, as well as papers on a number of other audio topics. He is a member of the Audio Engineering Society.