Directional Masking Coders for Multichannel Subband Audio Data Compression Systems

[2TM1.07]
Preprint 3261

GERZON, Michael A.;
Technical Consultant, Oxford, United Kingdom

Presented at the 92nd Convention 1992 March 24–27 Vienna

AN AUDIO ENGINEERING SOCIETY PREPRINT

Directional Masking Coders for Multichannel Audio Data Compression Systems

Michael A. Gerzon

Technical Consultant, 57 Juxon St.,Oxford OX2 6DJ, U.K.

Abstract

When monophonic subband audio data compression systems are used with multichannel sound reproduction systems, codec error artifacts are liable to severe "directional unmasking", especially if followed by a matrix processing stage. Methods are described of reducing this problem by adopting new subband coding strategies that align the direction of coding errors with that of the wanted signal, and two algorithms to implement such directional alignment are described. An additional analysis shows that residual directionally unmasked error artifacts can be reduced by up to 6 dB by a careful choice of an orthogonal or unitary post-codec matrix, rather than the highly non-orthogonal matrices used in some proposals for HDTV sound.

1. Introduction

If one has an n-channel audio signal, and codes it using n monophonic subband coding ("audio data compression") systems, such a codec (i.e. coding/decoding) operation C does not subjectively commute with audio signal matrixing operations M, i.e. subjecting the n-channel signal to

CM and MC

does not give the same degree of auditory masking of coding errors. In particular, in directional reproduction systems, coding errors will become more audible if they are reproduced from directions different to that of the wanted audio signal. This phenomenon, termed "directional unmasking", was discussed in detail by the author in ref. [1], where he showd that suitable encoding strategies can eliminate this effect by aligning the directions of the wanted signal and the coding error.

Directional unmasking becomes particularly serious in multispeaker reproduction systems if codec errors emerge from loudspeakers that have little wanted-sound output. The effect of such errors is two-fold:

(i) unwanted noise-like errors are heard from unactivated loudspeakers due to noiselike codec errors, and

(ii) amplitude modulation errors in the codec cause the apparent directionality of wanted signals to vary moment by moment, causing audible "pumping" and image instability effects.

It has been proposed to minimise the subjective effects of such codec errors in multichannel sound reproduction systems by using a "downmixing" transmission approach, whereby the loudspeaker feed signals for the most elaborate transmitted reproduction mode are separately encoded by monophonic subband encoders, and listeners receiving simpler reproduction modes using fewer loudspeakers implement a "downmixing" matrix in or after the decoders to derive the loudspeaker feed signals for their loudspeaker arrangement. See Meares [2] and Theile [3] for a discussion of "downmixing".

Downmixing has numerous disadvantages as compared to the alternative matrixing strategy known as "compatibility matrixing" (see [2], [3]) whereby matrixed transmission signals are used, with a suitable inverse matrix used by the end-user to derive speaker feeds, such that those listeners using fewer loudspeakers need only decode a correspondingly smaller number of transmitted audio signals. Compatibility matrixing permits the number of transmitted signals to be varied according to the number of transmission signals available in a flexible way, and does not require that the receiver be equipped with a different reproduction matrix for every possible transmission mode. In particular, this allows the introduction of new and more elaborate reproduction modes at a future time without rendering existing receivers incompatible, simply by adding additional audio transmission channels not used by existing receivers.

However, compatibility matrixing is found to give more audible directional unmasking when used with independent monophonic subband codecs. However, there exist directional sound reproduction systems, such as Ambisonics [4,5] in which independent speaker feed signals do not exist, and for which a downmixing approach cannot be used. For example, in B-format Ambisonics, three transmission signals W,X and Y are used (Conveying sounds with respective directional gains that are omnidirectional and have a forward and leftward-pointing figure-of-eight directional pattern) to convey feeds for four or more loudspeakers. In this example, sounds assigned to due front, due left, due right and due back will each give directionally unmasked errors from speakers in the opposite direction, but it is not possible to derive 3 signals corresponding to these four respective directions.

It is thus concluded that the downmixing approach not only involves a highly inflexible strategy making future enhancements difficult, and greatly complicating receivers using simple reproduction modes (due to the need to process a possibly large number of signals to derive a smaller number), but it also is inapplicable to some reproduction modes in any case. Additionally, downward mixing does not eliminate directional unmasking effects, but simply attempts to reduce some of its most extreme side effects, and it would be better to use a subband coding strategy that minimises directional unmasking in the first place by aligning the reproduced direction of coding errors with that of the wanted signal, as was shown to be possible in [1].

In the following, we indicate that such directional alignment of coding errors is possible even if n separate monophonic subband channels are used, as in systems such as those derived from MUSICAM or ASPEC or Dolby AC-2. This is done by modifying the quantisation strategy in each subband in the encoder, by choosing in each channel quantisation levels that take acount of the quantisation levels used in other channels in the same subband and the directional statistics in that subband of the wanted signal.

2. Joint Quantisation

Consider, by way of example, a 2-channel subband coding system using two monophonic coders, denoted by channel 1 and channel 2 respectively. In each subband, one will have a number (not necessarily the same) of quantisation levels in each channel, possibly with different quantiser step sizes in the two channels, giving possible pairs (q_1, q_2) representing possible quantisations of a wanted 2-channel signal (x, y). Such quantised signals are indicated by an "x" in the illustrated example of figure 1, which shows the signal (x,y) and possible quantisation levels. This illustrated example shows 6 quantisation levels for channel 1 and 9 for channel 2, with the channel 2 levels more closely spaced than those for channel 1.

The normal, so-called "independent" quantisation strategy for a 2-channel signal (x, y) quantises each of the channel signals x and y separately to the nearest available quantisation level in the respective channel. This minimises the overall quantisation noise energy, but gives a directional spread of quantisation noise that is uncorrelated in the two channels, having a typical form aligned along one of the channel axes, such as shown in figure 2 for the quantisation levels of figure 1. In general, the directional distribution of such quantisation noise and error is not aligned with the directional distribution of the wanted signal.

Were the directional distribution of quantisation noise to be accurately matched to that of the wanted signal, directional unmasking would not occur, and monophonic masking theory would be applicable.

It is possible to reduce the effect of directional unmasking by using an altered strategy to choose the quantised signal (q_1,q_2) to represent the wanted signal (x,y) such that the direction of the error signal $(q_1-x,\ q_2-y)$ tends to lie in the same direction as the wanted signal (x,y). For example, in the above illustrated 2-channel example, suppose that it is known that the signal has the direction of the arrow shown in fig. 3. Then we seek a quantisation point (q_1,q_2) that is close to the line of this arrow. For example, the quantisation point (level 1½,level 3) on figure 1 is the quantisation point that is closest in distance to the point (x,y), but the quantisation point (level 1½, level 2) is closer the the line of figure 3, and so is a better quantisation choice from the viewpoint of directional masking.

However, it is unwise to choose a quantisation point very nearly on the desired line if the resulting distance between the quantisation point and the wanted signal point is excessively large, since this would result in an exceedingly large coding error. One thus needs a strategy for choosing the quantisation point that helps to directionally align the coding error with the direction of the wanted signal without giving an excessively large coding error. There are many possible ways of doing this, and we shall give 2 examples of how this might be done.

3. Instantaneous alignment

One strategy helps align quantisation error sample by sample. For an n-channel signal $\mathbf{x} = (x_1,\dots,x_n)$ and a possible quantisation point $\mathbf{q} = (q_1,\dots,q_n)$, define a notion of "distance" between points. Define $\|(a_1,\dots,a_n)\| = \sqrt{(a_1{}^2+\dots+a_n{}^2)}$ or $\|(a_1,\dots,a_n)\| = |a_1| +\dots+ |a_n|$ (the former being a better choice than the latter for our purposes, but more difficult to compute). Conventional independent quantisation chooses that quantisation point $\mathbf{q}$ such that the distance $\|\mathbf{q}-\mathbf{x}\|$ is minimised. A technique of tending to align the direction of the quantisation error vector **q** - **x** with the instantaneous direction **x** of the wanted signal is given by an algorithm such as the following.

Step 1 Compute the normalised unit-length vector

$$\hat{\mathbf{x}} = (\hat{x}_1, \ldots, \hat{x}_n) \tag{1}$$

where

$$\hat{x}_i = x_i / \sqrt{(x_1{}^2 + \ldots + x_n{}^2)} \tag{2}$$

for i = 1 to n.

Step 2 Compute for each possible quantisation point **q** the component of the vector error in the direction of **x**, which is the vector

$$((\mathbf{q}-\mathbf{x}).\hat{\mathbf{x}})\hat{\mathbf{x}} = ((q_1-x_1)\hat{x}_1+\ldots+(q_n-x_n)\hat{x}_n)(\hat{x}_1, \ldots, \hat{x}_n)$$

$$= (b_1, \ldots, b_n) \tag{3}$$

Step 3 For a prechosen constant k between 0 qnd 1, choose the quantisation point that minimises the distance

$$\|(\mathbf{q}-\mathbf{x}) - k((\mathbf{q}-\mathbf{x}).\hat{\mathbf{x}})\hat{\mathbf{x}}\|. \tag{4}$$

By removing a proportion k of the component of the error signal vector **q-x** along the direction of x in computing the "magnitude" of the quantisation error, the error is biased towards lying in the same direction as the wanted signal. The best subjective choice of the constant k needs to be determined empirically, but a choice around k = 0.7 or 0.8 may be appropriate. In practice, one would not wish to compute the "distance" of equ. (4) for every possible quantisation point **q** for reasons of the amount of computation involved, but would confine the minimisation to a selection of quantisation points **q** in the immediate neighbourhood of the wanted signal **x**, possibly at the 2^n corners of the n-cube of quantisation points around the point **x**, or at the 3^n or 4^n quantisation points formed by examining the nearest 3 or 4 quantisation levels in each of the n channels.

The above computation in steps 1 to 3 above can be simplified, avoiding a square root operation, by computing equ. (3) as

$$((\mathbf{q}-\mathbf{x}).\mathbf{x})\ \mathbf{x}\ /\ (x_1{}^2+...+x_n{}^2) \qquad (5)$$

and minimising the <u>square</u> of the distance of equ. (4). The division operation can also be avoided to minimise computaional effort by instead minimising the distance of equ. (4) multiplied by $(x_1{}^2+...+x_n{}^2)$, i.e. by minimising the quantity (or its square):

$$\| (\mathbf{q}-\mathbf{x})(x_1{}^2+...+x_n{}^2) - k((\mathbf{q}-\mathbf{x}).\mathbf{x})\mathbf{x} \| \ . \qquad (6)$$

This reduces the optimisation of the choice of quantisation point to steps involving adds, multiplies and selecting minimums.

4. Statistical Alignment

The above quantisation strategy was based on aligning the quantisation error direction according to the instantaneous direction of each n-channel sample in that subband. A more sophisticated strategy aligns the directional statistics of the error signal **q-x** to the average directional statistics, determined over a number of samples, of the subband signal **x**.

The directional distribution of a subband signal **x** is conveniently measured by its $n \times n$ <u>correlation</u> <u>matrix</u> **C**, which for a signal $\mathbf{x} = (x_1, \ldots, x_n)$ is the matrix with entries

$$c_{ij} = \text{average}(x_i x_j) \;, \tag{7}$$

where "average" is a suitably weighted time average over a chosen number of samples with a chosen weighting. Since the correlation matrix **C** has $\frac{1}{2}n(n+1)$ independent entries, ideally the averaging should be over at least n^2 samples, although the following will still work with shorter averages.

The idea of the statistical choice of quantisation point is to multiply the quantisation error **q-x** vector by the inverse $\mathbf{C}^{-1}$ of the correlation matrix, and to minimise its length. While this strategy is conceptually simple, and aligns the quantisation error to have the same directional statistics as the wanted signal, it has two problems:

(i) the matrix **C** may be singular, or nearly so, meaning that $\mathbf{C}^{-1}$ may not exist or may be extremely ill-conditioned, and

(ii) this strategy can result in an extremely high error energy when the matrix **C** has some eigenvalues much smaller than others.

Also, one has the problem that computing a matrix inverse is computationally quite complicated, and involves division operations which are expensive in real-time signal processing.

The division operation can be avoided by multiplying an inverse matrix by the determinant of the original matrix, which gives a matrix all of whose $n \times n$ entries are determinants of $(n-1) \times (n-1)$ submatrices of the original $n \times n$ matrix, and so invlves no divisions in its computation.

For a matrix $\mathbf{M}$, define $\hat{\mathbf{M}}$ to be the matrix

$$\hat{\mathbf{M}} = (\det\mathbf{M})(\mathbf{M}^{-1}) , \tag{8}$$

computed by an algorithm involving no divisions, such as the Gaussian reduction algorithm without divisions. Then a possible statistical algorithms for aligning the direction of quantisation errors with that of the signal is as follows:

Step 1 Compute the correlation matrix $\mathbf{C}$ of the signal in each subband via equ. (7).

Step 2 For a predetermined constant $\varepsilon > 0$ chosen to avoid singularities, compute the $n \times n$ matrix

$$\mathbf{C}' = \mathbf{C} + \varepsilon(\text{average}(x_1{}^2 + \ldots + x_n{}^2))I_n , \tag{9}$$

where the concept of "average" in equ. (9) is the same as in used in equ. (7), and I_n is the $n \times n$ identity matrix. ε may be chosen to have a value such as $1/(2n)$, or else chosen adaptively, for example by choosing the coefficient of the I_n matrix in equ. (9) to equal the largest eigenvalue of $\mathbf{C}$ multiplied by 0.2 or 0.3 .

Step 3 Compute the matrix $\hat{\mathbf{C}}'$ from equs. (8) and (9), and the distance (or its square)

$$\|\hat{\mathbf{C}}'(\mathbf{q}-\mathbf{x})\| \tag{10}$$

and choose that quantisation point **q** that minimises the distance of equ. (10).

The problem of the maximum error energy becoming too large is controlled in two ways by the above algorithm:

(i) the choice of ε places an upper bound on the increase in error energy by controlling the degree of singularity of **C'**.

(ii) by limiting the quantisation points **q** over which the minimum distance (10) is determined, for example by only examining say the 2^n, 3^n or 4^n points determined by using the 2, 3 or 4 quantisation levels in each of the n channels closest to the signal level in that channel.

Note that $\hat{\mathbf{C}}'$ need only be computed once, so that only step 3 need be repeated for different quantisation points **q** .

5. General points

The above algorithms are by no means the only ones that tend to align the directional distribution of n-channel quantisation noise with that of the wanted signal. In particular, there is a class of algorithms that quantise one channel at a time, subtracting a multiple of the quantisation error of that channel from the remaining channels before quantising them. The coefficients of that multiplication can be determined from the correlation matrix **C**. Such successive quantisation strategies with error subtraction can give a useful degree of reduction of directionally unmasked errors, typically 3 dB for 2-channel systems, and has the advantage of not needing an explicit minimisation over a large choice of quantisation points.

Such algorithms generalise "noise shaping" methods, but it would complicate this report too much to give full details here.

The performance of any joint quantisation strategy can be optimised by choosing the bit allocation among the n channels carefully. In general, the best directional alignment of error energy (which would be when the correlation matrix of the error signal is proportional to that of the wanted signal in each subband) is not obtained if the bit allocation is such as to minimise total error energy, but may be better (e.g. see ref. [1]) if a more nearly equal number of bits is used for all n channels. In practice, the bit allocation should probably be such that more quantisation levels should be allocated to the most energetic channels, but such that there is not too great an inequality between the number of levels allocated to the quieter and the louder channels.

To some extent, the bit allocation strategy needs to be determined empirically, but perhaps will be such that, for example, 70% of the bits are allocated equally among the channels and the remaining 30% as though the aim were to minimise quantisation error energy (the Shannon-optimal bit allocation strategy), see ref. [6].

6. The effect of matrixing

Were the correlation matrix of the error signal accurately proportional to that of the wanted signal, no amount of post-codec matrixing would cause any directional unmasking of codec errors, since all linear combinations of signals (including those involving large cancellations of common components) would still mask errors which would remain below signal masking thresholds.

In the real world, however, exemplified by one of the above quantisation strategies, such perfect tracking of the error and the signal correlation matrices is not achieved, so that a reduced residual amount of directional unmasking will still occur. Therefore, the choice of post-codec matrixing still has some relevance.

One crude measure of the effect of a pre-and-post codec matrixing is what effect it has on the total energy of codec errors in reproduction. While such an approach takes no account of the directional unmasking effect, it at least quantifies how much error is laible to such unmasking.

The matrix strategies that are least liable to cause an excessive decrease in signal to noise ratio when signal noise are not directionally aligned involve the use of orthogonal matrices, which preserve all energies, both of wanted signals and of error noises. Such orthogonal matrices have all eigenvalues having a magnitude equal to 1, whereas if a matrix is used the smallest of whose eigenvalues has a small magnitude relative to the largest of whose eigenvalues, then if the wanted signal is an eigenvector having the smallest eigenvalue, and if noise is aligned along the direction of the eigenvector having the largest eigenvalue, then the signal-to-noise ratio will be degraded proportional to the ratio of the two eigenvalues.

A simple example in connection with two proposed 3-channel stereo coding systems illustrate this point. Use L_3, C_3, R_3 to denote signals intended to feed a respective left, centre and right speaker of a 3-speaker layout. In refs. [7-9], the author has proposed encoding these signals into 3 transmission signals M, S, T given by

$$\begin{pmatrix} M \\ S \\ T \end{pmatrix} = \begin{pmatrix} 0.500 & 0.707 & 0.500 \\ 0.707 & 0.000 & -0.707 \\ 0.500 & -0.707 & 0.500 \end{pmatrix} \begin{pmatrix} L_3 \\ C_3 \\ R_3 \end{pmatrix} \qquad (11a)$$

which are recovered by the inverse matrix equation

$$\begin{pmatrix} L_3 \\ C_3 \\ R_3 \end{pmatrix} = \begin{pmatrix} 0.500 & 0.707 & 0.500 \\ 0.707 & 0.000 & -0.707 \\ 0.500 & -0.707 & 0.500 \end{pmatrix} \begin{pmatrix} M \\ S \\ T \end{pmatrix}, \qquad (11b)$$

which are both orthogonal (energy preserving) matrix equations.

In refs. [2] and [3], Meares and Theile report on a different transmission and reception matrixing given by:

$$\begin{pmatrix} L \\ R \\ T \end{pmatrix} = \begin{pmatrix} 1.000 & 0.707 & 0.000 \\ 0.000 & 0.707 & 1.000 \\ 0.000 & 0.707 & 0.000 \end{pmatrix} \begin{pmatrix} L_3 \\ C_3 \\ R_3 \end{pmatrix} \qquad (12a)$$

which is a nonorthogonal transmission encoding matrix with the inverse transmission decoding matrix

$$\begin{pmatrix} L_3 \\ C_3 \\ R_3 \end{pmatrix} = \begin{pmatrix} 1.000 & 0.000 & -1.000 \\ 0.000 & 0.000 & 1.414 \\ 0.000 & 1.000 & -1.000 \end{pmatrix} \begin{pmatrix} L \\ R \\ T \end{pmatrix} . \qquad (12b)$$

Consider the two transmission systems of equs. (11) and (12) when they are both handling a signal fed only to the C_3 speaker, and suppose that the coding error in each coded transmission channel has an energy that is a fixed multiple k of the coded signal energy (a reasonable approximate description of subband codecs), with the errors in the 3 transmission channels being uncorrelated.

Then for the system of equ. (11), the respective error energies in the respective M, S and T channels are $\frac{1}{2}k$, 0, $\frac{1}{2}k$, so that the final reproduced error energies in the L_3, C_3 and R_3 reproduction channels are respectively

$$\tfrac{1}{4}k,\ \tfrac{1}{2}k,\ \tfrac{1}{4}k\ . \qquad (13)$$

For the system of equ. (12), the respective uncorrelated error energies in the L, R and T transmission channels are $\frac{1}{2}k$, $\frac{1}{2}k$ and $\frac{1}{2}k$, so that the final reproduced error energies in the respective L_3, C_3 and R_3 reproduction channels are

$$k, k \text{ and } k \ , \qquad (14)$$

which gives a total error energy 3 times as large as for equ. (13). Even worse, the directionally unmasked L_3 and R_3 speaker feeds are actually 4 times as large in this case for the system of equs. (12) as for the system of equs. (11).

This example illustrates that a bad choice of matrixing can substantially increase both the total energy and the degree of directional unmasking of directionally uncorrelated components of codec error. The system of equs. (12), however, has a no worse error energy performance for L_3 or R_3 isolated signals than the system of equs. (11), and has better directional masking for such signals. However, since it is often considered that central signals are the most important, and many 3-channel stereo mixes have most energy concentrated near the middle, the poor performance of the system of equs. (12) with central signal via subband codecs without directional alignment of errors is of considerable practical significance.

Thus, where possible, transmission encoding and decoding matrices should approximate being energy preserving or orthogonal, with their smallest and largest eigenvalues having similar magnitudes, in order to minimise the effects of directional unmasking. The main problem in this regard, besides that arising from matrices using "cancellation" of frontal-stage stereo centre-speaker channels, arises from transmission systems using cancellation of back channel information from frontal channels [2,3], where a frontal channel contains, for reasons of compatibility of mono and stereo reproduction, a proportion of rear stage sounds. When such rear stage information is cancelled from the frontal speakers, the codec errors from the rear channel signals remain in the frontal speaker outputs, increasing the risk of directional unmasking.

This problem is reduced if the proportion of rear signal contained in the frontal signal is kept low, and the proposal of Meares [2] to mix rear sounds in with front sounds 6 dB down will perform markedly better than if the mixing is at full level, as has been proposed in connection with the Japanese 3:1 HDTV system [10].

Thus all transmission encoding and decoding matrixing used with subband coding systems should be evaluated not only for directional unmasking, but also for a liability to decrease reproduced signal to noise ratio. Matrices nearly proportional to orthogonal matrices will perform much better than those whose largest and smallest eigenvalues have very different magnitudes, and are strongly preferred. The frontal stereo coding systems proposed by the author in refs [8,9] have this desirable orthogonality property, unlike systems reported by Meares and Theile [2,3]

7. Principal Vector Coding

Besides directional alignment of codec errors with wanted signals, there is a second quite distinct strategy for minimising directional unmasking effects, which works on the principle of "gating" out both the wanted signal and the error noise signal away from the dominant wanted signal direction.

This principle has long been familar in so-called "logic" or adaptive matrix surround-sound decoders. The gating of subsiduary direction signals has the advantage of preventing spurious sounds from the wrong directions, and the disadvantage of affecting the sound quality of the wanted signal in a possibly undesirable manner.

An example of the "gating" approach is the proposal in MUSICAM [11] to transmit the higher frequency subbands of a stereo signal monophonically, but with separate amplitude gain coefficients for each stereo channel so as to "pan" the sound in that subband to the dominant stereo position. The effect of this proposal is to suppress stereo signal information in that subband in the stereo-channel direction orthogonal to the direction of the predominant signal energy, i.e. to gate it out.

This proposal can be generalised to n-channel systems by transmitting in some subbands only a smaller number $m \leqslant n$

of channel signals corresponding to the m components of the n-channel signal resolved in the directions of the m eigenvectors of the signal correlation matrix corresponding to the m largest eigenvalues. This has the effect of gating out the remaining n-m channels corresponding to the directions of the n-m eigenvectors with the smallest eigenvalues.

More generally, only $m \leqslant n$ channels will be transmitted lying in a vector subspace approximating to that spanning the m principal eigenvector of the correlation matrix, possibly with an additional gain adjustment so that the total reproduced energy remains similar in that subband to the original n-channel signal.

This kind of principal value coding of some subbands not only has the desirable effect of gating unwanted error noise in non-principal directions, but also gives a reduction of the transmitted data rate by reducing the number of channels that need to be transmitted in an adaptive manner. By this means, significant economies in transmitted data rate can be achieved for systems using a large number of channels, by a factor 2 or more over separate monophonic subband coding.

The disadvantage is that the gating of the lesser principal vector signal components in each subband may have audible side-effects on signal quality, especially for applications where high sound quality is critical. It is easy to provide rationalisations about the ears' ability to "mask" such gating errors in the presence of the dominant directional components, as has been done in the past in connection with "logic" decoders, but the evaluation of such proposals requires great care.

In particular, it is not enough that there be no very obvious change in sound quality or directional effect. It is also necessary to ensure that the resulting sound can give low listening fatigue on extended listening, something that "logic" decoders have proved to be poor at. There is currently not an adequate psychoacoustic understanding to design such systems without considerable trial and error, but transmission of the principal vector components in general would be expected to be better than transmission of only a very restricted range of possible vector components.

8. Conclusions

In this report, we have shown that it is possible to modify the quantisation strategy in coding n monophonic transmission channels used to convey n-channel directional sound such that any subband codec errors are largely directionally aligned with the wanted signal. Such alterations of coding/ quantisation strategy still allow independent monophonic decoders to be used for individual transmission channels, and do not require any change of coding system specifications, being purely an encoder process.

Two such modified "joint quantisation" methods have been detailed, and others outlined. The reduction of directional unmasking effects allows the use of compatibility matrixing of multichannel sound, rather than the inherently inflexible method of "downward mixing", which in any case cannot avoid directional unmasking effects for some types of sound reproduction system.

Residual directional unmasking effects caused by the imperfect alignment of the directional distribution of errors with wanted signals can be minimised by using compatibility matrixing that uses approximately orthogonal matrices, or those whose smallest and largest eigenvalues have similar magnitudes. It was shown that some existing 3-channel proposals strongly violate this desirable property.

REFERENCES

[1] M.A. Gerzon, "Problems of Error-Masking in Ausio Data Compression Systems", Preprint 3013, 90th Audio Engineering Society Convention, Paris (1991 Feb.)

[2] D. Meares, "High Quality Sound for High Definition Television", Proc. 10th AES International Conference, "Images of Audio", London (1991 Sept.) pp. 163-177

[3] G. Theile, "HDTV Sound Systems: How Many Channels?" Proc. 10th AES International Conference, "Images of Audio", London (1991 Sept.) pp. 147-162

[4] M.A. Gerzon, "Ambisonics in Multichannel Broadcasting and Video", J. Audio Eng. Soc., vol. 33 no. 11, pp. 859-871 (1985 Nov.)

[5] M.A. Gerzon, "Criteria for Evaluating Surround Sound Systems", J. Audio Eng. Soc., vol. 25 no. 6, pp. 400--408 (1977 June)

[6] L.D. Davisson, "Rate Distortion Theory and Application", Proc. IEEE, vol. 60, pp. 800-808 (1972 July)

[7] M.A. Gerzon, "Optimal Reproduction Matrices for Multispeaker Stereo", Preprint 3180, 91st Audio Engineering Society Convention, New York (1991 Oct.)

[8] M.A. Gerzon, "Hierarchical Transmission System for Multispeaker Stereo", Preprint 3199, 91st Audio Engineering Society Convention, New York (1991 Oct.)

[9] M.A. Gerzon, "Production Approaches to Multispeaker Stereo", Studio Sound, vol. 34 (1992) to appear.

[10] K. Nakabayashi, K. Korozumi, E. Miyasaka, S. Komiyama & A. Morita, "Three-One Quadraphonic Sound System for High Definition Television", Proc. AES 10th International Conf. "Images of Audio", London (1991 Sept.) pp. 141-6

[11] Y-F. Dehery, "MUSICAM Source Coding", Proc. AES 10th International Conf. "Images of Audio", London (1991 Sept.) pp. 71-79

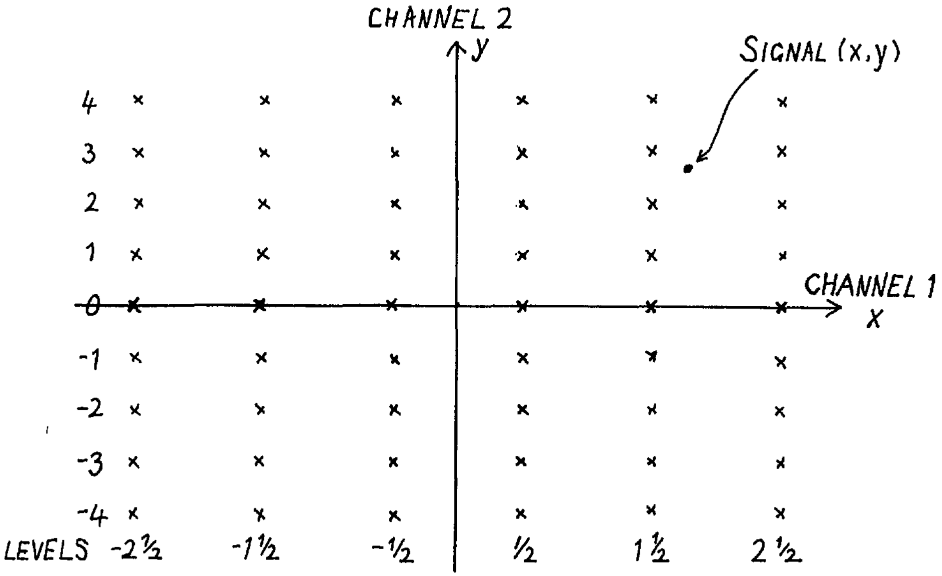

Figure 1. Showing possible quantisation points (marked "x") for a 2-channel signal with components x and y.

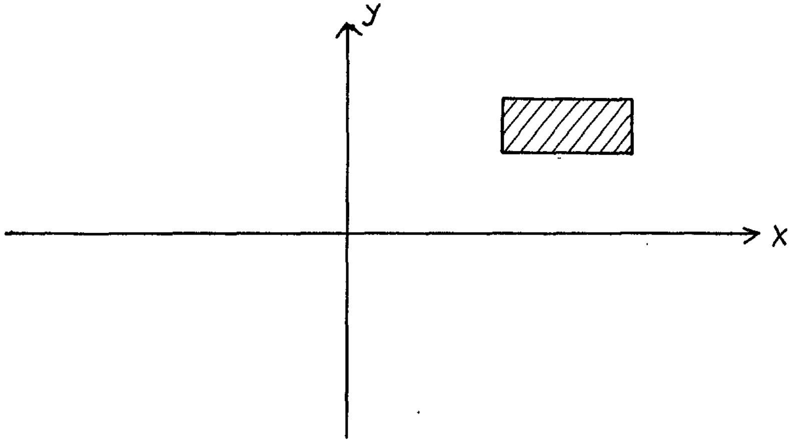

Figure 2. Directional distribution of quantisation noise for independent quantisation as in figure 1.

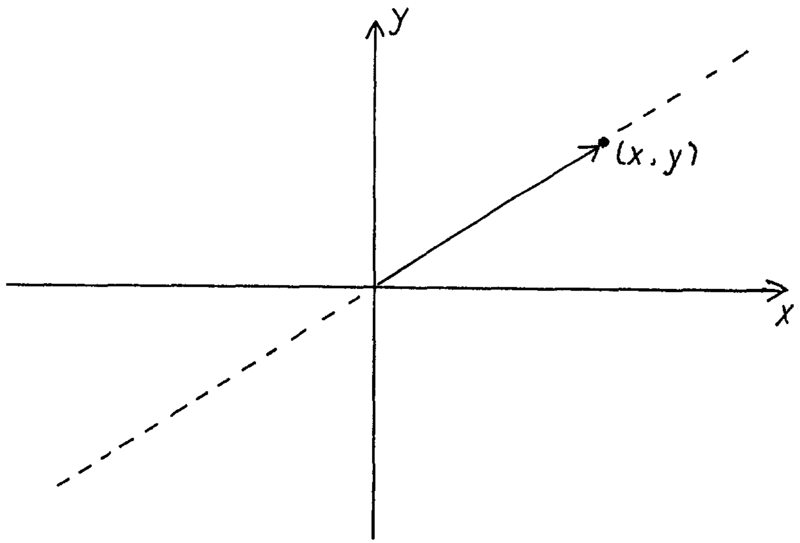

Figure 3. Showing the instantaneous direction line of a 2-channel signal (x,y)