Psychoacoustic Decoders for Multispeaker Stereo and Surround Sound 3406 (J-8)

Michael A. Gerzon
Technical Consultant
Oxford, UK

Presented at
the 93rd Convention
1992 October 1–4
San Francisco

AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

Psychoacoustic Decoders for Multispeaker Stereo and Surround Sound

Michael A. Gerzon *

Technical Consultant, 57 Juxon St., Oxford OX2 6DJ, UK

Abstract

This paper overviews a new generation of multispeaker directional Sound Reproduction Technology, based on maximising the number of sound localisation cues heard by a listener that are consistent with one another. The paper describes methods for optimising the presentation of conventional and 3-channel stereo via three to five front-stage loudspeakers, and new "ambisonic" decoders giving enhanced frontal-stage image stability for 5- and 6-loudspeaker surround sound. Applications are discussed, including those for HDTV sound.

0. INTRODUCTION

Conventional 2-channel stereo, and many of the multispeaker systems proposed for HDTV surround sound, convey direct loudspeaker feed signals to create the illusion of a directional sound image. However, either when one wishes to reproduce the signals encoded for directional reproduction via one loudspeaker arrangement via another (e.g. 2-channel stereo reproduced via 3 loudspeakers), or where one wishes to optimise the subjective illusion of phantom image directions, it becomes appropriate to think of using psychoacoustic directional decoders, i.e. appropriately designed matrix algorithms (which may be frequency-dependent) to feed the incoming information to the loudspeaker arrangement used by a listener.

The idea of using psychoacoustic decoding is very far from new. Indeed, as long ago as 1931, Blumlein [1] proposed a psychoacoustic decoder (termed a "Blumlein Shuffler" [2]) for reproducing quasi-dummy-head signals via a stereo pair of loudspeakers.

However, the systematic design of such decoders for multispeaker stereo and surround sound was systematically developed by the author in the 1970's in connection with that surround-sound technology termed ambisonics. The full psychoacoustic design theory then used has only recently been given general circulation as ref. [3], but the resulting technology has been described in various places (see refs. [4] and [5], and references therein).

Recently, the author began the development of a second generation of psychoacoustic directional decoding technology (see refs. [6] and [7]), motivated by interest in stereo and surround-sound systems for TV and HDTV use using a front-center loudspeaker to stabilize sounds associated with on-screen visual action. Initially, he applied this to solving the problem of reproducing stereo intended for one number of loudspeakers via a larger number - a problem that dates back to the earliest days of stereo from work at Bell Telephone Laboratories [8]. The success at applying psychoacoustic methods to the frontal stage stereo case led to the extension of this work to a new generation of Ambisonic decoders [7] aimed at giving improved sound image stability in the due front direction so as to match the direction of on-screen images.

The design of these new decoders largely used the theory [3] developed in the 1970's for the older generation of Ambisonic decoders, (although detailed refinements in the use of this theory have occured over the years based on experience). However, the mathematical use of the theory to the new problems resulted in horrendously difficult mathematical problems, which required the solution of systems of simultaneous nonlinear equations - and these solutions in general occured near singularities in the space of solutions so that numerical methods had to proceed with great care, involving initial manual searches for the singularities.

Nevertheless, the resulting solutions are in many ways exceptionally well behaved. It is not the aim of this paper to describe the theory or solution methods used (these are detailed in refs. [6] and [7]), but to overview the properties of the new decoder solutions arrived at. The papers [6] and [7] are both rather dense and it is hoped that the present paper will provide a good introduction to the new decoder technology.

1. PSYCHOACOUSTIC CRITERIA

The general theory of directional psychoacoustics used in designing decoders has been given in detail in ref. [3], and summarised in a form applicable to specific design problems in refs. [6], [7] and [9], so that we do not intend to repeat that material here. However, it may be useful to give a more heuristic idea of what psychoacoustic criteria decoders meet, based on an early account of the theory published in ref. [10].

The aim of decoders is to create, in the ears and mind of a listener, the directional effect originally encoded into the source material. Since in general only a few loudspeakers are used in reproduction, this effect is obtained by providing cues used by the ear/brain system to localize sounds. In psychoacoustic decoders, the aim is generally to maximize the number of mutually consistent localization cues provided, so as to increase the reliabilty and robustness of localization, and to reduce listening fatigue caused by inconsistent cues.

The ears and brain localize sounds according to many different mechanisms. Among the most important cues used are low frequency interaural phase (applicable up to around 2 kHz, but dominant below 700 Hz) and localization by amplitude differences between the two ears, predominantly above about 1 kHz. While other cues are also important, we have found that satisfying both these cues, and making them mutually consistent for a central listener facing in any direction, leads to a particularly robust and reliable localization quality.

While it is generally quite easy to design reproduction that satisfies low frequency interaural phase cues, reproduction of the higher frequency amplitude or energy localisation cues cannot generally be made perfect using only a few loudspeakers. Moreover, there is generally a tradeoff between optimizing the interaural phase cues and the amplitude or energy cues at the ears. Because the dominant cue varies with frequency, this in practice means that psychoacoustic decoders are designed to vary the trade-off between the accuracy of the two cues as a function of frequency, by using frequency-dependent matrixing.

Such frequency-dependent matrixing is characteristic of psychoacoustic decoders for directional sound reproduction via loudspeakers (although in some cases such as the 4×3, 5×4 and 5×3 decoders for frontal-stage stereo systems reported in [6], such frequency-dependence proves not to be necessary.) However, this frequency-dependence is primarily in the way the sounds are distributed among the reproduction loudspeakers, and in general it is a design aim to ensure that the overall frequency response, i.e. the total energy gain from all loudspeakers, is substantially flat for all sounds.

One of the aims of psychoacoustic decoding is to ensure what has become (perhaps misleadingly) known as "crosstalk cancellation" at low frequencies for a central listener, i.e. to ensure that the interaural phase localization cues are correct below around 700 Hz at the listener's head. The calculation of this via models for the head is somewhat tedious and complicated, and in ref. [10], the author noted that a much simpler way existed of calculating whether the interaural phase cues were correct at low frequencies.

If one considers, as in fig. 1, the signals arriving at the two ears of a listener, then the information available to the listener is equivalent to:

(1) The sum of the signals arriving at the two ears, and

(2) the difference of the signals arriving at the two ears.

At low frequencies, where the head gives little acoustical obstruction to arriving sounds, the sum signal (1) has an omnidirectional gain polar pattern to sounds arriving from different directions (see figure 1), and the difference signal (2) has a sideways-facing figure of eight (cosine) gain polar pattern to arriving sounds.

Thus consideration of the information available to the ears at low frequencies shows that we may consider the information as being that picked up by an omnidirectional and a sideways figure-of-eight microphone positioned at the listener. A relatively simple mathematical analysis which we shall not give here shows that interaural phase cues are given by the information in components of sounds in the figure-of-eight signal that are in 0° or 180° phase relationship to the corresponding sounds in the omnidirectional signal, and that signal components in quadrature (90°) phase relationship between the figure-of-eight and omnidirectional signals have no effect on interaural phase. This observation is the basis of the practical use of the localization metatheory given in ref. [3].

When we allow the listener to rotate his/her head, the "sideways" figure of eight microphone also rotates, having now both a forward and a sideways figure-of-eight component. Thus, when head rotation is included, the information available to the ears and brain of a listener at low frequencies is given by the outputs of three microphones positioned at the listener, namely as shown in figure 2: an omnidirectional signal W, a forward-facing figure of eight signal X and a sideways-facing figure-of-eight signal Y. (Strictly speaking, this is only true for horizontal head rotations. Rotations in other planes also use the information of a fourth vertical Z figure-of-eight signal). As before, components of the figures of eight in quadrature phase relationship to the corresponding sound components in the omnidirectional signal do not affect interaural phase at low frequencies, and thus do not contribute to low frequency localization.

The three directional pick-up patterns at a point that characterise low frequency directional localization, shown in figure 2, constitute the directional encoding method termed B-format, which in the hor izontal plane consists of 3 signals: a signal W containing all sounds with equal (say unity) gain, a signal X containing all sounds with gain $\sqrt{2}\cos\theta$ and a signal Y containing all sounds with gain $\sqrt{2}\sin\theta$, where θ is the azimuth angle of sound arrival direction measured anticlockwise from due front. (The factor $\sqrt{2}$ is a convention that helps to ensure that all channels of B-format encoded sound typically have similar average energies.)

While low-frequency interaural phase localization is thus equivalent to providing pressure and orthogonal velocity information for the ears, localization at higher frequencies is instead determined by the directional flow of <u>energy</u> in the sound field near the listener's head. For these higher frequencies (typically between 700 Hz and 4 kHz), the omnidirectional signal used in the low frequency localization theory may be replaced by the total energy of each sound at the listener, whereas the figure-of-eight signals may be relaced by the directional components of the <u>sound intensity</u> of the sound field, which measures the vector directional flow of energy.

The details of the resulting <u>energy vector</u> theory of sound localization, apt at higher frequencies, are given in refs. [3,6,7,9,10]. However, it can be shown that if the sound arrivals at the listener from different loudspeakers are phase-incoherent, then the low and the high frequency theory then give the same result! Thus, in particular for very noncentral listeners at different distances from all the loudspeakers, the "high frequency" energy vector theory is also applicable at lower frequencies.

In practice, there is in any case a broad range of frequencies around 700 Hz at which both theories are partially applicable. For this and other reasons, one of the primary design aims of all psychoacoustic decoders designed by the author is to ensure that the same localisation is given for sounds by both of the above theories. This consistency of localization condition helps ensure lower listening fatigue and a greater reliability of localization than if only one theory is correct. Except for the center and the two loudspeaker directions, this consistency condition fails to hold for conventional two-loudspeaker stereo.

Thus, although the <u>quality</u> of localization (i.e. image sharpness and stability) may be better according to one theory than the other at any given frequency, with the trade-off varying with frequency, a psychoacoustic decoder will at least ensure that the basic image <u>direction</u> is broadly consistent between the two theories.

There is no completely general method of designing psychoacoustic decoders that guarantees this consistency, which is why some of the more recent designs reviewed in this paper took so long to arrive at after the basic psychoacoustic localization theory used in their design was known.

Above 4 or 5 kHz, other factors affect sound localization, as noted in [10], notably the effect of pinna colourations at the ears, which play an important role in determining the localization of sounds at these highest frequencies. The design of decoders to optimize these highest frequencies is somewhat more empirical than at lower frequencies, although in ref. [10] we suggested that these colourations can be partially incorporated into loudspeaker feeds to improve localization.

It is not always possible to make all localization cues consistent, and in such cases, one should at least attempt to get some localization cues correct. This is generally easiest at low frequencies, where it proves possible to provide correct phantom image cues even to the sides of a listener. Thus, although low frequency localization cues are not generally the most important ones, they are the ones we can always get right, and that is one reason why these cues prove to be important in the design of psychoacoustic decoders. However, in ref. [9], an example was given of the dangers of designing equipment solely on the basis of low frequency cues, where it was shown that this can lead to very suboptimal localization at higher frequencies.

Although the transition between the low and high frequency theories is nominally centered at 700 Hz, listeners are likely to be seated away from a central listening position, which creates a likelihood that the "high" frequency energy vector theory is in fact apt at lower frequencies. For this reason it is in practice found that a lower transition frequency than 700 Hz is best used for the frequency ranges in which psychoacoustic decoders are optimized for low and high frequency localization. For domestic environments, a transition frequency of 400 Hz has been found apt, with lower transition frequencies appropriate for larger listening environments.

2. OLDER AMBISONIC DECODERS

In ref. [4], we overviewed and described a number of Ambisonic decoders for surround sound reproduction based on the psychoacoustic localization ideas just outlined. In this older generation of Ambisonic decoders (see refs. [4] and [11]), the idea was to create in the decoder by means of an initial decoding matrix signals W', X' and Y' representative of reproduced pressure and forward and sideways components of acoustic velocity at the listener, and then to subject these velocity signals and pressure signals to shelf filters, typically centered around 400 Hz, so as to alter the relative gains of reproduced pressure and velocity at low and at high frequencies.

The general block diagram of such an older Ambisonic decoder is shown in figure 3. An initial matrix converts the encoding scheme in use (which may be B-format, conventional amplitude stereo or a dedicated 2-channel surround-sound encoding system such as UHJ [4]) into a pressure signal W' and orthogonal velocity signals X' and Y' such that the signal components in X' and Y' not in quadrature phase relationship to the signal components in W' have a directional pattern such as that shown in figure 2. The subsequent shelf filtering generally increases the gain of the W' signal above 400 Hz and reduces that of the X' and Y' signals above 400 Hz, in such a way that:

1) the overall energy frequency response of the decoder is roughly flat for all input directions, and

2) the quality of localization for the energy vector theory is optimized above 400 Hz.

These modified pressure and velocity signals are then fed to an output amplitude matrix to provide loudspeaker feed signals. The precise form of this matrix depends on the shape of the loudspeaker layout use and the number of loudspeakers, having an adjustment (termed a "layout control") to optimize it for various loudspeaker layouts so as to give the intended reproduced pressure and velocity (and vector sound intensity direction) for a central listener. By way of example, figure 4 shows a typical arrangement for use with a variety of different shapes of rectangular layouts of loudspeakers.

The form of the first decoding matrix and the choice of shelf filters is determined entirely by the directional encoding system used for the input signals, and does not depend on the loudspeaker layout, whereas the output matrix depends on the reproducing loudspeaker layout but not on the input directional encoding system.

The kind of decoder used in figure 3, however, only works for a quite restricted class of loudspeaker layout shapes - including rectangles, regular polygons with 5 or more sides, and for loudspeaker layouts where the loudspeakers are arranged in pairs diametrically opposed to each other with respect to a central listener.

This restriction arises because the architecture of figure 3 can be shown to give identical localization directions for both the low and high frequency localization theories only for such layouts - thanks to various mathematical theorems proved in detail in ref. [3].

This restriction is one of the problems with the older generation of Ambisonic decoders. Another problem is that they are limited in the degree of image stability they can give at high frequencies. This was not a particular problem for audio-only applications, because the relative directions of different sounds moved broadly together as the listener moved around, giving a still convincing surround-sound field with only a moderate directional distortion. However, when used with visual media such as film or TV, the mismatch of on-screen visual action and associated sounds becomes a problem.

The reduction of image instability especially for the frontal stage can only be done to a limited degree using the older generation of decoders. A three-channel surround-sound encoding format such as B-format gives a better image stability than 2-channel surround-sound encoding systems such as UHJ [4], and modern digital transmission media allow more channels to be conveyed. However, the older generation of decoders strictly limit the attainable improvement.

Even within the older generation of psychoacoustic decoders, however, a more elaborate architecture than that of figure 3 could give better trade-offs among conflicting factors affecting localization quality. Figure 5 shows the block diagram of an older-generation horizontal decoder with such extra features. The input decoding matrix now also produces a fourth signal B', which is a 90° phase shifted version of the pressure signal W'. Because it is in quadrature phase relationship to W', B' can be combined with the velocity signals without affecting localization direction, but only localization quality. By carefully designing the amount of B' combined with Y', and the shelf-filtering of B', it proves to be possible simultaneously:

1) to improve localization stability of the frontal stage at the expense of the rear stage,

2) to reduce an unpleasant quality, termed "phasiness", caused by phase differences between reproduction loudspeakers for frontal stage sounds, and

3) to ensure a much flatter frequency response for sounds from all directions.

The improvements of figure 5 over figure 3 are particularly relevant to decoding 2-channel directional encoding systems such as UHJ or 2-channel stereo into four or more loudspeakers, and are not so relevant to B-format, which can already be decoded with better results via a basic decoder of the form of that of figure 3.

In the decoders of figures 3 or 5, it has been found that it is important that the shelf filters be "phase compensated", i.e. should have accurately matched phase shifts, and should differ only in amplitude response. Such phase compensated shelf filters may be implemented using first order shelf filters of non-minimum-phase type, with all shelves designed to give a 90° phase shift at the same frequency around 400 Hz.

Another feature shown in fig. 5 is the use of high-pass filters in the velocity signal paths. These filters compensate for the effect of curvature of the sound field at the listener due to the finite distance d of the loudspeakers from the listener. These filters are first order high-pass filters with a -3 dB frequency of 53/d Hz (where d is in metres), and it is generally the effect these filters have on the phase response that is most important, not their effect on amplitude response. The effect of these distance compensation filters is very subtle on most sounds, but is occasionally audible as an improvement on drums, double basses and the like. These filters may most easily be understood as the inverse filters to the familiar bass boost "proximity effect" encountered when close to velocity microphones. Regarding the ear system as incorporating velocity microphones as in figs. 1 and 2 means that one should equalise for the proximity of the loudspeakers.

An important aspect of the older generation of Ambisonic decoders is that the whole technology was also applicable to full-sphere (4π steradian) with-height surround sound, termed "Periphony" ([12,13]) by the author.

Full sphere surround sound is encoded in 4 B-format channels W, X, Y, Z whose directional pick-up gain patterns are shown in fig. 6, now including a vertical figure of eight pickup Z. Such signals can be provided either by full-sphere panpots or by means of a Soundfield Microphone such as the AMS Sound Field Microphone.

The typical form of an Ambisonic decoder of the older generation for periphony is shown in fig. 7, where now all three velocity signals are shelf filtered, and where a more complex layout control is used. Figure 8 shows several possible loudspeaker layouts that can be used with such a decoder, and more details of such decoders can be found in refs. [11] and [13].

Because of the elaborate loudspeaker layouts required, such periphonic decoders may only have limited domestic use, but in the author's opinion, full-sphere surround sound is subjectively a marked improvement on even the best horizontal surround sound, being both more natural and much easier to listen to when listening for subtleties of sound. However, as in the horizontal case, there is a problem of matching the direction of visual and sound images for noncentral listeners when using this older generation of decoders.

3. MULTISPEAKER STEREO

The older Ambisonic decoders could only be designed for full surround-sound loudspeaker layouts. With the advent of Television with stereophonic sound, however, the author became interested in applying psychoacoustic decoding ideas to frontal stage stereo systems using three or more loudspeakers distributed across a frontal sector of directions, as reported in detail in ref. [6]. The idea here was to see if it was possible to decode signals originally intended for a smaller number (say two) of stereo loudspeakers via a larger number (say three or four) so as to obtain an enhanced and more consistent stereo directional illusion.

The restricted range of directions covered by the stereo loudspeakers, as compared to the 360° covered by previous Ambisonic decoders meant that the optimization of psychoacoustic decoders for this case involve quite different trade-offs, even though the underlying psychoacoustic theory [3] used is identical. These different trade-offs mean that the structure of psychoacoustic decoders in this case is quite different from that of Ambisonic decoders.

For example, the lack of sound images to the sides of the listener in the stereo case means that it is not so important to completely optimise low frequency localization psychoacoustics, and that small deviations from ideal low-frequency interaural phase localization are permissable, provided that the low-frequency localization is broadly consistent with the energy localization theory across the frontal sector of directions. It was found, as a result of detailed studies reported in [6], that it was not necessary in this case to make the decoding matrix different below and above 400 or 700 Hz, since it was found that the resulting improvements in low frequency localization quality were negligible.

However, in this application, it was found that optimizing the behaviour in the highest frequency region above 5 kHz proved to be important, since otherwise it was not possible to get adequate stereo width in three-loudspeaker reproduction of two-channel stereo. Thus in the stereo application, it was found that the decoder frequency-dependence was centered around 5 kHz, and not the 400 Hz of the Ambisonic decoder case. This is the transition frequency for a different set of auditory localization mechanisms, notably the Haas effect and pinna colouration.

In figure 9, we show the loudspeaker layouts considered for frontal stage stereo using from one (regarding mono as the trivial case of "one-loudspeaker stereo"!) to five loudspeakers, showing both the angles and the notations used for each loudspeaker feed signal. For simplicity, we here only consider the special case where all loudspeakers are at the same distance from a "central" listener position, and with all the loudspeakers pointing at the listener.

Typically for frontal stage stereo use, the total subtended loudspeaker layout angle ($2\theta_2$, $2\theta_3$, $2\theta_4$ and $2\theta_6$ for the respective cases of 2, 3, 4 and 5 loudspeakers) is around 60°, although it may be larger for the three, four or five loudspeaker cases. Also typically, the angles between adjacent pairs of loudspeakers are all the same, being respectively $\frac{1}{2}$, $\frac{1}{3}$ and $\frac{1}{4}$ of the total subtended angle of the layout in the respective cases of 3, 4 and 5 loudspeakers. However, the following discussion is not confined to this preferred equal-angle case.

THe simplest case we consider is the decoding of two-channel stereo, originated for two-loudspeaker reproduction, via the three-loudspeaker stereo layout, using a 3×2 reproduction decoder matrix as shown schematically in figure 10.

Two obvious requirements on such a decoder matrix is that
1) it should be left/right symmetrical - i.e. the results should be unchanged if left and right signals are swapped at both inputs and outputs, and
2) it should preserve the total reproduced energy of input stereo signals.

As discussed in ref. [6], the energy-preservation property is desirable for several reasons: to preserve the input signal balance, to prevent "comb-filter" colorations with spaced microphone techniques, to maximize reproduced image width, and to preserve distance effect encoded into a recording by early reflection cues, as discussed in detail in ref. [14].

The general form of a 3×2 matrix decoder satisfying these properties 1) and 2) can be shown (see [6]) to be given by figure 11 with width gain $w = 1$. (The width gain w provides an optional adjustment of reproduced width). In figure 11, the MS matrices are sum and difference matrices with gain such that the energy of signals passing through is preserved (which means that the sum and differences of inputs are given gains 0.7071), and the sine/cosine gain adjustment splits the energy of the input sum signal between the center and the outer loudspeakers of the three-loudspeaker layout.

It is found that no value of the angle parameter ϕ gives ideal localisation quality all across the stereo stage at all frequencies, although $\phi = 45^{\circ}$ in fig. 11 is a moderately good frequency-independent compromise. For this reason, it is found that making ϕ frequency-dependent, with a value around 35° below 5 kHz and about 55° above 5 kHz gives optimum results, with the highest degree of consistency between different localization theories below 5 kHz, and improved stereo width and distribution of sounds among the loudspeakers above 5 kHz. (See ref. [6] for the details).

Such a frequency-dependent psychoacoustic 3×2 decoder can be implemented as shown in fig. 12, by using a cross-over network before the sine/cosine energy splitter, with the high-frequency value ϕ_H of ϕ set at 55^o and the low-frequency value ϕ_L of ϕ set at around 35^o. If the crossover network sums to an all-pass response, a matching all-pass network is required to phase compensate the difference signal, as indicated schematically in fig. 12 before the width gain w adjustment.

There are alternative implementations of the psychoacoustic 3×2 stereo decoder other than that of fig. 12. One such equivalent implementation, derived in ref. [15], is shown in figure 13. Here, the transition between low and high frequency behaviour around 5 kHz is achieved by passing the sum signal M_2 through a first order all-pass network with gain -1 below 5 kHz and gain +1 above, and giving it a gain of 0.172, and then passing the resulting three channel signals M, S and T into a 3×3 "transmission decoding matrix" D_{33} of the form described in ref. [15]. The only filtering in this implementation is provided by the all-pass network. This implementation can be simpler than that of fig. 12, and moreover naturally generalizes to the 3-loudspeaker stereo decoder for B-format signals discussed later.

The advantage of a psychoacoustic 3×2 decoder of the kinds just described is severalfold: they markedly improve the stability of central images as listener position varies as compared to two-loudspeaker stereo, localization cues across most of the reproduced stereo stage are made much more consistent for a listener at the ideal "stereo seat", and the listener away from the stereo seat hears a reproduced stereo stage that is much less "distorted" than with two-loudspeaker stereo, giving a stereo stage that is still distributed fairly evenly between left and right loudspeakers. The result of these improvements is stereo that can be heard across a larger listening area, and that gives lower listening fatigue than two-loudspeaker stereo, as well as better alignment of central sounds with associated on-screen visual images for TV.

Moreover, it is found that these improvements appear with a wide variety of stereo recording techniques, even those that do not use amplitude panning of stereo position, such as for example, widely spaced omnidirectional microphones or Blumlein shuffled dummy head recordings.

As shown in ref. [6], similar decoders can be found for reproducing signals intended for multispeaker stereo reproduction via any one number n_1 of stereo loudspeakers via any larger number n_2 of stereo loudspeakers, as shown schematically in fig. 14.

Such $n_2 \times n_1$ matrix reproduction decoders can be designed either substantially to <u>preserve</u> the originally intended directional effect achieved via n_1-loudspeaker stereo via the larger number n_2 of loudspeakers (in which case they are termed <u>preservation</u> decoders), or else to improve the reproduced directional effect, when they are termed improvement decoders.

It was not obvious, in advance of the detailed work reported in ref. [6], that such a thing as a preservation decoder could actually be designed. After all, there are many possible ways of panning sounds via n_1-loudspeaker stereo, with many different panpot laws (see for example ref. [9]). However, by adjusting the matrix parameters of an $n_2 \times n_1$ matrix such that 1) the matrix is left/right symmetrical, 2) the matrix is energy-preserving for all input signals, and 3) for a set of input "test" stereo signals which excite either only one or (equally in phase) two of the input n_1 loudspeakers, ensuring that the localization direction after decoding is consistent at low and high frequencies according to the two localization theories - one finds that:
a) there is an essentially unique such matrix, and that
b) such a matrix is indeed a preservation decoder (to a very good degree of approximation) for panned sounds according to a wide variety of panning laws.

However, the preservation of the input directional psychacoustics is not (and clearly cannot be) perfect - the image stability of some directions originally near input loudspeakers is degraded somewhat after decoding, whereas the image stability of other directions can be improved. Also, there is an inevitable reduction of the total reproduced angular width of the total stereo image after decoding due to crosstalk in the matrixing - but the proportionate loss of angular width turns out to be small (around the 90% mark) for n_1 equals three or more, in the case that the angles between adjacent pairs of loudspeakers in fig. 9 are equal.

A remarkable aspect of this work is that we have found that within wide limits, the actual form of the preservation matrix is largely insensitive to the precise angular disposition of the loudspeakers in the n_1- and n_2-loudspeaker layouts, so that in practice, a single matrix can be used for all reasonable values of the angles θ_p in figure 9. Again, this was not obviously the case in advance of the detailed work in ref. [6].

The actual details of the design of these matrices, whose equations are given in the appendix of ref. [6], are highly mathematical, and involve systems of simultaneous nonlinear equations in parameters describing the decoding matrices. The solution of these equations is by somewhat tedious numerical methods, which involve an initial hand search for rough solutions before computer numerical methods are used, since the solutions are close to mathematical "singularities" in the nonlinear problem involved.

The derivation of solutions is greatly simplified by an observation enshrined in the schematic of figure 15. This observation is that if we cascade an $n_2 \times n_1$ preservation decoder with an $n_3 \times n_2$ preservation decoder, with $n_1 < n_2 < n_3$, then the result must also be a preservation decoder, since the localization effect is preserved at each stage. Thus we can solve and design first for $(n+1) \times n$ preservation decoders for n = 2, 3, 4 etc, and then design other cases simply by cascading such decoders.

The schematic of fig. 15 shows all possible decoders for all numbers of input and output loudspeakers up to five, where one can "go in" in stereo for n_1 loudspeakers and "come out" via any larger number n_2 of loudspeakers. Fig. 15 also shows how stereo signals originated for different numbers of loudspeakers may be mixed together.

Of course, fig. 15 is not intended to imply that, say, a 5×3 decoder should actually be implemented as a cascade of a 4×3 and a 5×4 decoder - only that it can be designed as that matrix obtained by computing the cascade of the two "component" matrices. Its actual implementation will be as a single matrix.

Improvement decoders can similarly be designed as equivalent to cascades of component improvement decoders via the schematic of fig. 15. In general, unlike preservation decoders, improvement decoders will be frequency depedent (as shown for the 3×2 case in figs. 12 or 13) in order to optimise localization psychoacoustics above 5 kHz.

The remarkable thing about these stereo psychoacoustic decoders is that their actual implementation is technologically simple - and they would actually have been economically viable even with mid 1960's technology. This is a case where the use of a technology was not inhibited by technical difficulties in implementation, but by a lack of appropriate <u>ideas</u> of what to implement. Were the optimization to have been done twenty years ago, psychoacoustic 3×2 decoders could have been in general use then - but instead much energy was put into attemping to optimize what proved to be a quadraphonic dead end [5].

However, when the loudspeakers are no longer equally distant from a central listener, these decoders will only work as well as possible if the loudspeakers closer to the central listener are fed via time delays to ensure that all sounds arrive at the same time. High quality time delays have only become cheap in recent years with the advent of low-cost digital technology.

The psychoacoustic localization theory can also be applied to n-loudspeaker stereo systems in another way - to devise panpot laws for panning directional sounds among the n loudspeakers in a manner that ensures that low and high frequency localizations are consistent - something that is not true for conventional 2-loudspeaker stereo. The reader is referred to ref. [9] for the details of this work, especially with reference to its figure 4.

4. NEW AMBISONIC DECODERS

The improved front-stage image stability obtained using three or more stereo loudspeakers, especially for use with TV/video reproduction, motivated the development of improved Ambisonic decoders achieving a similar improvement of frontal stage image stability. It was evident from what was known that
1) such an improvement could not be achieved with Ambisonic decoders of the older type, and
2) that improved frontal image stability was only possible via loudspeaker layouts having extra loudspeakers at or near the due front direction, as shown for example in figure 16 for the 5-loudspeaker case or figure 17 for six loudspeakers.

In these loudspeaker layouts, one or two near-front loudspeakers are added to four loudspeakers in a rectangular or trapezium shaped layout.

From what was learned in the stereo case, it became obvious that with such loudspeaker layouts, one could no longer rely on a simple shelf filtering of pressure and velocity signals with associated matrixing to implement ambisonic decoders for the new loudspeaker layouts. Rather, one had to go back to square one and to design decoders separately in the low (below 400 Hz) and in the high (above 400 Hz) frequency regions, optimizing each for image quality and stability for the theory of sound localization apt in its frequency range, and to use cross-over networks to combine these two designs into a single decoder. Even worse, this design procedure has to be repeated for every possible shape of loudspeaker layout (determined by the angles ϕ, ϕ_F, ϕ_B and ϕ_C in figures 16 and 17).

Unlike in the older ambisonic decoders, every aspect of the decoding interacts, with the decoder matrixing being dependent on encoding system, loudspeaker layout shape and on frequency.

Also, as in the stereo case, the system of nonlinear equations ensuring psychoacoustic consistency of direction of different localization mechanisms could only be solved by tedious numerical methods, without the elegance of the mathematical results of ref. [3] in the older case.

Although, as reported in ref. [7] which outlines the design procedure, there is an analytic solution method (albeit one which is very lengthy), the problem is that it has a number of free parameters that have to be carefully optimized for best results, and the process of optimizing these parameters is a lengthy and time-consuming process. (Initially, a decoder design for each and every loudspeaker layout took about 5 hours work even with computer aids).

A part of the reason why this process is time consuming is that the analytic solution is multivalued and close to singular behaviour in the free parameters, so that a close search is required before the best values (very close to the singularities) are found. Another difficulty is that the "optimization" process is itself not well defined, being a trade-off among conflicting psychoacoustic defects, and it took some time to identify a reasonable set of "objective" design criteria that gave a reasonable trade-off for a wide variety of loudspeaker layouts.

The reason why such an "objective" criterion was desirable was that if one did a purely subjective "optimization" of the low and high frequency decoders for each loudspeaker layout, then there was no guarantee that for layouts with an intermediate shape that an "interpolated" design between ones that had been arrived at would be OK. With "objective" criteria for the trade off, it is possible to compute designs for a range of layouts - and to use interpolation methods for intermediate layouts without having to do a new and lengthy design.

Even then, it turns out that the way that the decoder design varies with layout is quite singular, so that interpolation does not give as accurate results as might be hoped.

In the 1970's, the author postponed consideration of designing ambisonic decoders for the layouts of figs. 16 and 17 because of an intuition that it was not going to be easy - an intuition that proved to be wholely correct!

The basic architecture of the new ambisonic decoders is, however, conceptually straightforward, if more complicated to implement than older designs. As shown in figure 18 for B-format decoding, and ignoring for the present the input "forward dominance" block, the B-format signal is fed to a phase-compensated crossover network centered around 400 Hz, and the low frequency and high-frequency B-format components are fed to separately designed decoding matrices, whose outputs are then recombined. For reasons of simplicity, these matrices do not directly derive the loudspeaker feed signals in their frequency range, but instead derive the sum and the difference of the signals fed to left/right symmetric loudspeaker pairs (eg. L_F and R_F, or L_B and R_B, or C_L and C_R), which requires the use of the sum and difference matrices shown at the output of fig. 18 to derive the actual loud-speaker feed signals.

For different loudspeaker layouts, it is thus necessary to alter all 9 matrix coefficients that define each of the low and the high frequency decoding matrices shown in figure 18 a total of 18 adjustments as the layout changes. Once the values of these coefficients have been computed by the above tedious design theory, values can be downloaded from memory or a look-up table (possibly via an interpolation procedure) for the layout shape used by the listener.

Unlike in older Ambisonic decoders, it turns out that the directional gain polar pattern of the reproduced pressure signal is frequency-dependent for these new decoders, being omnidirectional at low frequencies, but having a backwards-facing subcardioid characteristic above 400 Hz, the precise shape of which is layout dependent.

In doing the design of decoders for a given loudspeaker layout, it is necessary to ensure not only that the localizations for all directions given by both localization theories are identical within each frequency band and the same for both frequency bands, but that the localization does not change in the crossover frequency region around 400 Hz. Fortunately, provided that the crossover network is phase compensated (i.e. that the low- and high-pass filters have identical phase responses), it is found that deviations of reproduced direction in the crossover region are at most only a fraction of a degree.

It is found that typically, a 5-loudspeaker ambisonic decoder designed by these methods typically has about twice the frontal stage image stability of a corresponding older type of 4-loudspeaker Ambisonic decoder, without a marked reduction of image stability across the rest of the 360° azimuthal sound stage. The 6-loudspeaker decoders typically have about 2½ times the 4-loudspeaker frontal stage image stability.

There is, however, a price paid for this improved B-format frontal stage image stability. This is that, if sounds are reproduced from their originally encoded directions, it is found that sounds from the back stage (which are the most unstable in localization quality) are reproduced louder than sounds from the front stage. This is generally not acceptable, since preservation of the originally intended level-balance is important, and is also necessary for proper redproduction of any recorded distance effect [14].

The initial "forward dominance" block shown in fig. 18 is intended to solve this problem. Forward dominance [7] is a transformation of B-format encoded signals that has the effect of altering the positions and gains of sounds encoded within the B-format 360^o azimuthal stage, while still maintaing the relationships among channel gains characteristic of B-format. By way of example, a suitable forward dominance matrix transformation increasing the level of front-stage sounds by 3 dB and reducing the level of back stage sounds also by 3 dB has the effect of moving all encoded azimuths towards the front as illustrated in fig. 19, reducing the width of the frontal stage somewaht and correspondingly increasing the width of the back stage.

Although the use of forward dominance transformation of B-format alters image width and gains, the fact that the resulting signal is still a B-format directionally-encoded signal means that the result of decoding it via an ambisonic decoder still satisfies all the requirements of consistency between different localization cues and of image localisation quality and stability. Thus the effect of preceding the new type of ambisonic decoder by a forward dominance transformation is to compensate for the front/back level imbalance of the decoder, but also to slightly narrow the angular widths of sound at the front and to widen the sound stage at the back a little.

In practice, the forward dominance gain compensation need not be implemented separately as shown in fig. 18, but can be incorporated into the decoding matrix coefficients.

For TV applications, the slight narrowing of the frontal stage may actually be desirable, since otherwise the reproduced surround-sound front stage width may be too wide when the ordinary stereo fold-down [16,17] is matched to the size of the visual image.

It is also possible to design ambisonic decoders for systems other than horizontal B-format using additional near-front loudspeakers. By way of example, 2-channel surround sound systems such as UHJ may be decoded via the layouts of figs. 16 or 17 using the decoder architecture of fig. 20. This is similar to that of fig. 18, except that the encoded 2-channel signal is converted to four signals W_2, X_2, Y_2 and B_2 similar to those used in the older decoder of fig. 5 using a phase--amplitude matrix, and these signals are then fed to a phase compensated cross-over network and to two decoding matrices more or less as before (although the decoding matrices now each have 12 adjustable real coefficients instead of 9 due to the extra B_2 signal in quadrature phase relationship with W_2.)

Such new decoders for 2-channel encoded sound may prove useful in improving ambisonic decoder results from 2-channel source material, including the back-catalog of material encoded into the 2-channel UHJ format [4] available on compact disc, which besides classical and world music repetoire, includes recordings of music by Paul MacCartney, Tina Turner, Alan Parsons, Steve Hackett and others.

However, although 2-channel ambisonic decoders can also be used for decoding 2-channel surround material originated for the cinema, it is expected that the greatest benefits of the new decoding technology will be obtained for material originated in B-format or in further enhanced modes using four or five channels described in refs. [7] and [16].

In ref. [7], we also show how the basic B-format 5- or 6-loudspeaker decoder of figure 18 can be used as the basis for decoding 4 or 5 channels allowing further improved image stability beyond that attainable from 3-channel B-format. The possibility of such further improvement is dependent on the use of a decoding technology already capable of improved front-stage image stability from B-format. The extra channels supplementing the three B-format channels can be used to provide completely stable images at due front, which can be locked to the important center-of-screen visual position for cinema and HDTV productions.

Design work is still proceding, based on the same methods already described for the 5- and 6-loudspeaker horizontal B-format case, for more elaborate loudspeaker layouts for auditorium and possible home use, and experience is still being acquired concerning the optimum trade-offs between different localization qualities.

5. 3-LOUDSPEAKER DECODING OF B-FORMAT

B-format is a surround-sound 360^{o} azimuthal directional encoding format, but the above psychoacoustic methods have also been applied to optimize the decoding of B-format via frontal 3-loudspeaker stereo, as described in ref. [16]. The reason for wanting to do this is that many listeners in the future may only have a three-loudspeaker stereo layout, and it turns out that a suitable decoding of B-format into such a layout can give better image quality and stability than simply decoding a 2-channel stereo folddown into the 3 loudspeakers via the psychoacoustic 3×2 decoders of figs. 12 or 13. Also, it may be desired in TV production applications to provide a three-channel stereo feed from a sound field microphone.

The 3-loudspeaker stereo decoder for a B-format input is designed under the constraint that back sounds should be picked up no louder than the front, and that frontal stage sounds picked up across a 120^{o} sector of directions should more-or-less fill the stereo stage. The solution we have arrived at is a further development of the 3×2 stereo decoder shown in figure 13.

Figure 21 shows a schematic of the 3-loudspeaker stereo decoder for B-format. The B-format signal is first subjected to an optional forward-dominance transformation, allowing adjustment of reproduced stage width and front/back level balance without affecting psychoacoustic localization quality of the decoder. Ignoring this, the B-format signal is then passed through a first matrix to produce signals M_3, T_3 and S_3 whose respective polar diagrams (gains as a function of encoded B-format direction) are forward and backward-facing hypercardioids and leftward-facing figure-of-eight. The hypercardioids are arranged to have nulls 135° off their respective axes, i.e. a front-to-back ratio of 15.31 dB. The two hypercardioid signals are then passed into a frequency-dependent rotation matrix, which rotates the two signals in the M and T channels by an angle ϕ - 45°, where ϕ is the same angle parameter used in the 3 X 2 stereo decoder schematic of fig. 11. The form of this rotation matrix is shown in fig. 22, where it is seen to simply be a natural elaboration of the all-pass plus gain used in fig. 13 to optimize frequency dependent decoding of the 3 X 2 decoder. The output matrix in fig. 21 is exactly the same "transmission decoder matrix" as used in figure 13.

Thus, apart from the initial optional forward dominance, the 3-loudspeaker decoder for B-format shown in fig. 21 is essentially the optimum 3 X 2 stereo psychoacoustic decoder for the 2-channel stereo signal whose sum is a forward-facing hypercardioid and whose difference is a leftward-facing figure-of-eight, but supplemented by a frequency-dependent contribution from a rear-facing hypercardioid signal T_3.

While this 3-loudspeaker decoder for B-format does not give as good a localization quality as optimally-panned 3-channel stereo signals (see [9]), it gives significantly better quality than 3 X 2 decoding of 2-channel stereo, and is a worthwhile option for three-loudspeaker stereo listeners who have B-format surround-sound source material available.

With this decoder, a B-format surround program can thus satisfy the needs not only of surround-sound listeners as described in section 4 above, but also give good results for listeners equipped with just three frontal loudspeakers. In the past, three channel transmissions have had to be dedicated either to the surround-sound listener or to the 3-channel stereo listener, but not both.

By using a 4-loudspeaker decoder equivalent to following the above 3-loudspeaker decoder with a 4 X 3 stereo reproduction decoder matrix, it is also possible for 4-loudspeaker stereo listeners to reproduce B-format with improved image stability.

6. OTHER SPATIAL ASPECTS

This paper has primarily been concerned with decoders optimizing the quality of directional images of sounds, but passing reference has also been made to other aspects of the spatial illusion, notably distance. At least three aspects of spatial imaging other than simple direction are important: 1) the apparent distance of sounds, 2) the apparent angular size of sound sources, and 3) the "spaciousness" of sounds (often termed "spatial impression").

No one technology can be considered in pure isolation from the others, and in ref. [14], the author reported new methods of providing precise control of apparent distance by providing appropriate early reflection cues, which are used by the listener to locate in distance. Hitherto, such cues have only been derivable from certain "purist" microphone techniques listed in ref. [14], which include the Ambisonic B-format and 2-channel UHJ microphone techniques - one of the things that has kept UHJ recordings being made commercially when other surround-sound technologies from the 1970's fell into disuse.

As reported in [14] and [18], a technology has now been developed for synthesising distance cues by digitally simulating the cues used by the ears and brain to locate in distance. Since [18] was written, considerable progress has been made in optimizing these cues for studio use, departing from the early reflection patterns found in actual rooms so as to reduce the perceived comb-filter colorations caused by adding early reflections, while still preserving a reliable sense of distance. The modified natural cues also are designed to increase operational flexibility in practical studio use.

This technology, while first being made available for 2-channel stereo, has been conceived also for multispeaker stereo and surround sound. Thus distance cues will become generally available for use in directional encoding systems, allowing panning in distance as well as direction.

However, such distance cues place additional demands on psychoacoustic decoders in order that the information they carry not be disrupted. In particular, it is necessary that psychoacoustic decoders should preserve the timing and gains of the simulated early reflection cues, and the decoders described in this paper have specifically been designed to meet this requirement as well as possible. In particular, the exact energy preservation property of the $n_2 \times n_1$ stereo decoders meets this requirement, and the forward dominance gain correction in the new ambisonic decoders also is aimed at helping preserve distance cues.

In ref. [18], we also describe a technology for giving mixed mono source signals a realistic image size, and the B-format and 3-loudspeaker stereo versions of this technology have also been devised that ensure that the components of the spatially spread image are still accurately localized without fatiguing side-effects such as phasiness. Again, the full benefit of such image size technology will only be obtained if psychoacoustic decoding maximizing the accuracy and consistency of localization cues is provided.

It is also known that spatial impression is conveyed largely by the spatial distribution of early reflections, but also by the spatial incoherence of later reverberation. (There is in fact more than one factor making up the composite quality termed spatial impression). Here, the ability of a surround sound system to accurately convey the impression of early reflections from all round the listener, particular from side directions, is known to be important. Ambisonic decoders satisfy this requirement at low frequencies, although there is a conflict between designing decoders for the best portrayal of spatial impression and of direct image localization. Generally, good spatial impression requires a higher transition frequency than 400 Hz but reduces the optimum listening area for direct-sound localization. The improved frontal-stage imagery of the new decoders makes this trade-off less critical, and optimization for spatial impression is one of the factors being studied.

It is the author's experience that height (elevation) portrayal of early reflections and reverberation contributes markedly to spatial impression, and that horizontal surround-sound systems are inherently compromised in this respect, so that the trade-offs between spatial impression and other qualities for horizontal surround-sound are inevitably going to be somewhat personal and subjective.

7. CONCLUSIONS

This paper has provided a survey of psychoacoustic decoding technologies for directionally encoded sound, based on providing consistency of direction between different localization cues used by the ears below and above 700 Hz. After briefly showing how localization psychoacoustics leads to considering directional microphone patterns at the listener, the application of this theory in older Ambisonic decoders was reviewed, and the limitations of these older decoders described.

A new generation of psychoacoustic decoders was described, having in common the use of extra loudspeakers at or near the due front direction, so as to improve image stability for the important front direction. This is particularly important when sounds accompany associated visual images in TV and video applications.

Two main classes of decoders were considered: those for frontal stage stereo systems using three or more loudspeaksers, amd those for 360^{o} azimuthal portrayal of surround sound. Because of the different trade-offs in the two cases, the first class of decoders tends to be frequency dependent with a transition frequency around 5 kHz, whereas the second "ambisonic" class of surround-sound decoders tend to be frequency-dependent with a transition frequency around 400 Hz.

In general, the decoders for the stereo loudspeaker layouts tend to be fairly simple matrix algorithms, although time delays may be used to compensate for loudspeaker layouts in which loudspeakers are at an unequal distance from the center of the listening area. These stereo decoders generally preserve or improve the stereo effect heard, especially for three-loudspeaker decoding of ordinary two-channel stereo. An advantage of the frontal stage stereo decoders is that they are less obtrusive in domestic environments, and that they enlarge the listening area for good stereo, making stereo a more social listening experience. Their disadvantage is that they do not give the enhanced spatial qualities of good surround sound.

The new generation of ambisonic decoders combines improved frontal stage image stability with the benefits of a 360^{o} portrayal of the recorded environment. The new decoders, while offering a more flexible range of loudspeaker layouts than older Ambisonic decoders, have a rather more complicated signal processing architecture, and have proved to be much more complicated to design. They also give a slightly narrower portrayal of the frontal stage of directions than older Ambisonic decoders.

The paper also discussed a 3-loudspeaker stereo decoder for B-format signals, which permits the same 3 B-format channels to be used either for frontal stage stereo with more than 2 loudspeaksers or for full ambisonic surround sound via 5 or 6.

The paper also discussed new technologies for recording sounds in distance as well as direction, and also other spatial qualities, and described aspects of the psychoacoustic decoder design relevant to reproducing these spatial qualities well.

The combination of new psychoacoustic decoding technologies with the new studio technologies for enhanced spatial illusions means that we can look forward to greatly improved reproduction of spatial aspects of sound in the home in the near future. Learning how to use these new possibilities well has meant having to unlearn some older approaches that were developed when understanding of the psychoacoustics was much more limited.

The future extensions of these new decoding technologies to the full-sphere portrayal of surround sound means that in the further future, this technology will continue to offer continued improvements, and the beginnings to an accurate reproduction of the kind of experience one can hear live in a first rate acoustic.

8. ACKNOWLEDGEMENTS

I would like to thank Geoffrey Barton for his continued interest as well as many fruitful discussions, as well as his involvement in the development of the new technology. Thanks are also due to those, too numerous to mention without being invidious, who have stimulated the above ideas over the years with encouragement, interest or active help.

9. PATENT NOTE

Most of the decoders described in this paper are the subject of patents or pending patent applications. The newer decoders are the subject of applications filed by Trifield Productions Ltd.

10. REFERENCES

[1] A.D. Blumlein, British Patent 394,325 (filed 1931 Dec. 14), reprinted in (ed.) J. Eargle, Stereophonic Techniques, Audio Engineering Society, New York, 1986

[2] M.A. Gerzon, "Applications of Blumlein Shuffling to Stereo Micriphone Techniques," Preprint of the 93rd Audio Engineering Society Convention, San Francisco (1992 Oct. 1-4)

[3] M.A. Gerzon, "General Metatheory of Auditory Localisation," Preprint 3306 of the 92nd Audio Engineering Society Convention, Vienna (1992 March 24-27)

[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] M.A. Gerzon, "Optimum Reproduction Matrices for Multispeaker Stereo," J. Audio Eng. Soc., vol. 40 no. 7/8, pp. 571-589 (1992 July/Aug.)

[7] M.A. Gerzon & G.J. Barton, "Ambisonic Decoders for HDTV," Preprint 3345 of the 92nd Audio Engineering Society Convention, Vienna (1992 March 24-27)

[8] J.C. Steinberg & W.B. Snow, "Auditory Perspective - Physical Factors," Electrical Engineering, vol. 53 no. 1, pp. 12-15 (1934 Jan.), reprinted in (ed.) J. Eargle, Stereophonic Techniques, Audio Engineering Society, New York, 1986.

[9] M.A. Gerzon, "Panpot Laws for Multispeaker Stereo," Preprint 3309 of the 92nd Audio Engineering Society Convention, Vienna (1992 March 24-27)
[10] M. Gerzon, "Surround Sound Psychoacoustics," Wireless World, vol. 80, pp. 483-486 (1974 Dec.)
[11] M.A. Gerzon, British Patents 1,494,751, 1,494,752, 1,550,627 and 2,073,556. Corresponding U.S. patents: 3,997,725, 4,086,433, 4,081,606 and 4,414,430.
[12] M.A. Gerzon, "Periphony: With-Height Sound Reproduction," J. Audio Eng. Soc., vol. 21 no. 1, pp. 2 - 10 (1973 Jan.)
[13] M. A. Gerzon, "Practical Periphony," Preprint 15?? of the 65th Convention of the Audio Engineering Society, London (1980 Feb.)
[14] M.A. Gerzon, "The Design of Distance Panpots," Preprint 3308 of the 92nd Convention of the Audio Engineering Society, Vienna (1992 March 24-27)
[15] M.A. Gerzon, "Hierarchical Transmission System for Multispeaker Stereo," J. Audio Eng. Soc., vol. 40 no. 9, pp. ? - ? , (1992 Sept.)
[16] M.A. Gerzon, "Hierarchical System of Surround-Sound Transmission for HDTV," Preprint 3339 of the 92nd Convention of the Audio Engineering Society, Vienna (1992 March 24-27)
[17] M.A. Gerzon, "Compatibility of and Conversion Between Multispeaker Systems," Preprint of the 93rd Convention of the Audio Engineering Society, San Francisco (1992 Oct. 1-4).
[18] M.A. Gerzon, "Signal Processing for Simulating Realistic Stereo Images," Preprint of the 93rd Convention of the Audio Engineering Society, San Francisco, (1992 Oct. 1-4)

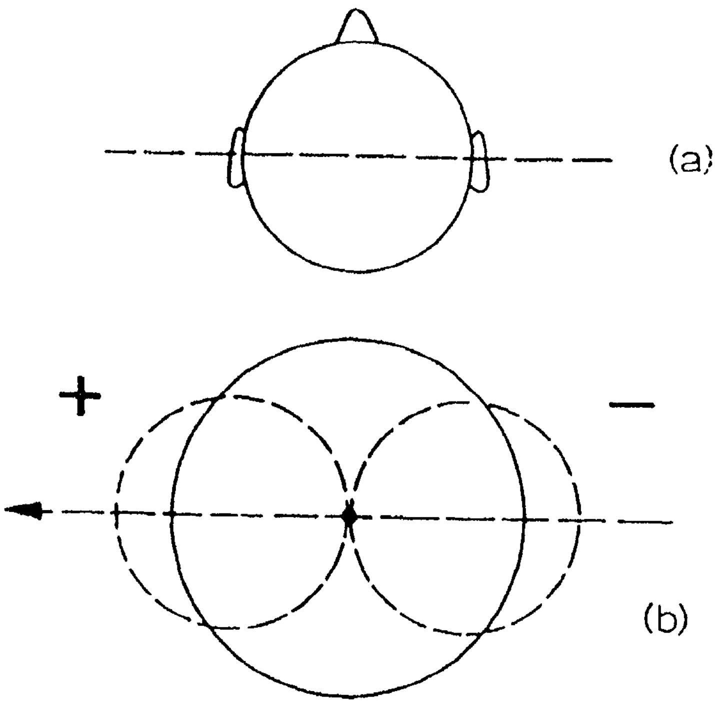


Fig. 1. Omnidirectional and velocity microphones (picture b) receiving the same low frequency information as the human hearing system (picture a).

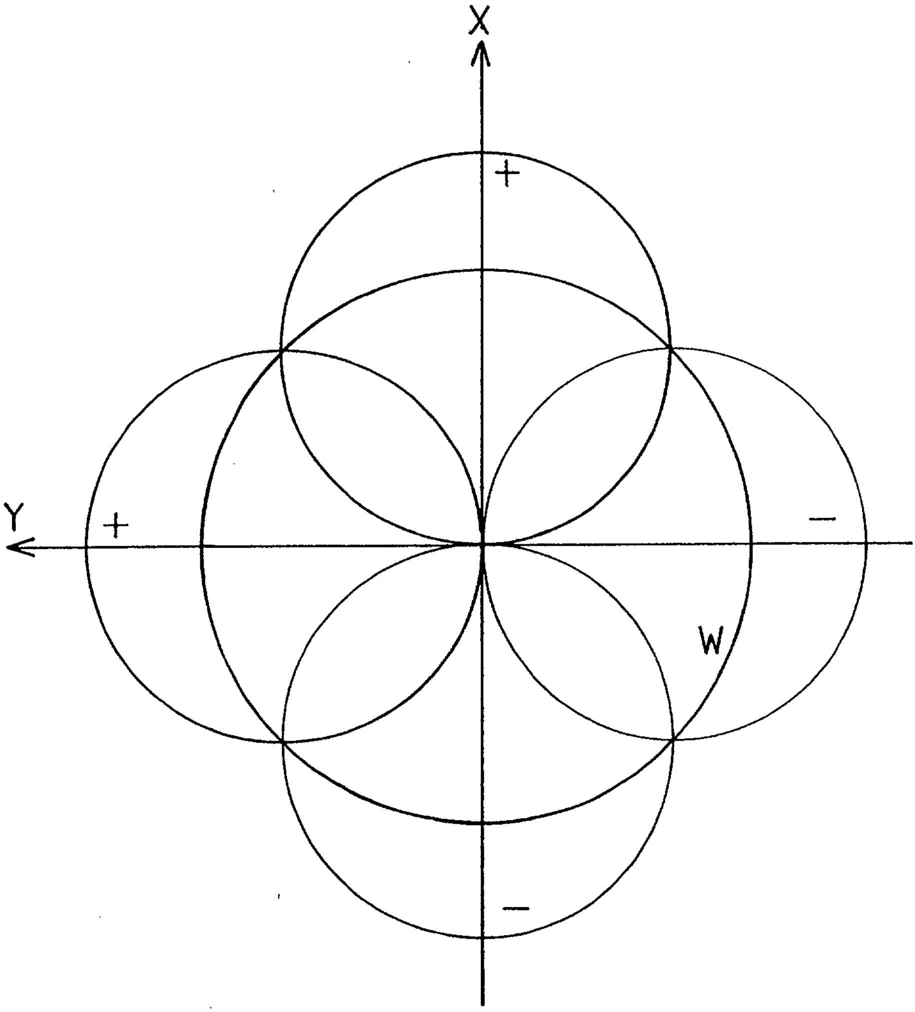


Figure 2. Polar diagrams in horizontal plane of B-format signals W (omnidirectional), X (forward figure-of-eight) and Y (leftward figure of eight), representing low frequency pick-ups derived by the ears of a listener as in fig. 1 whose head is allowed to rotate.

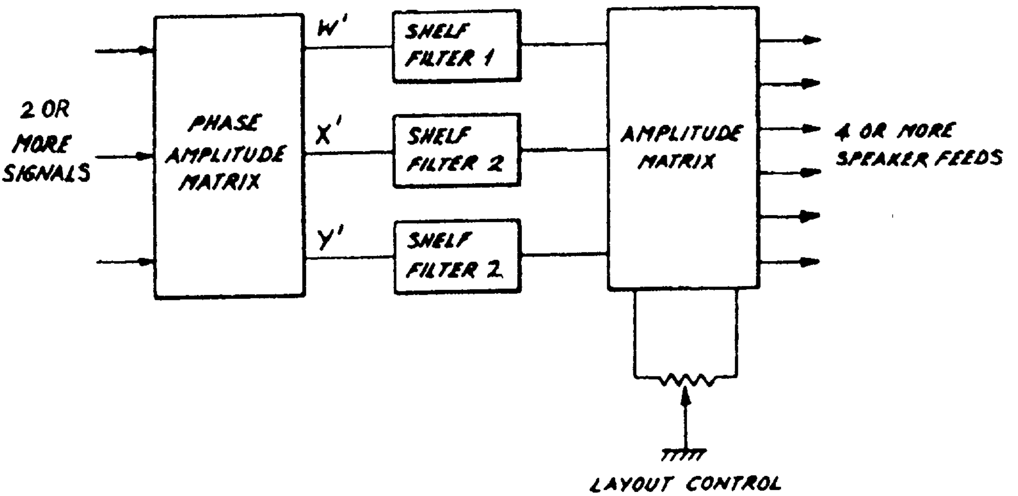


Figure 3. Basic schematic of available Ambisonic decoders

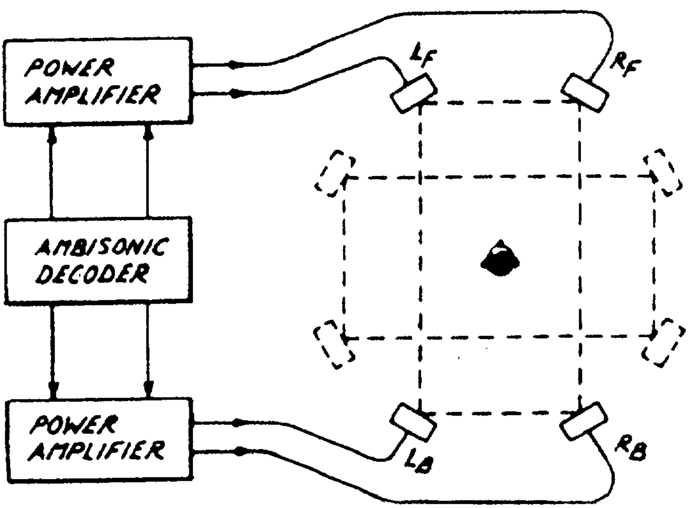


Figure 4. Rectangular speaker layouts for Ambisonic decoders

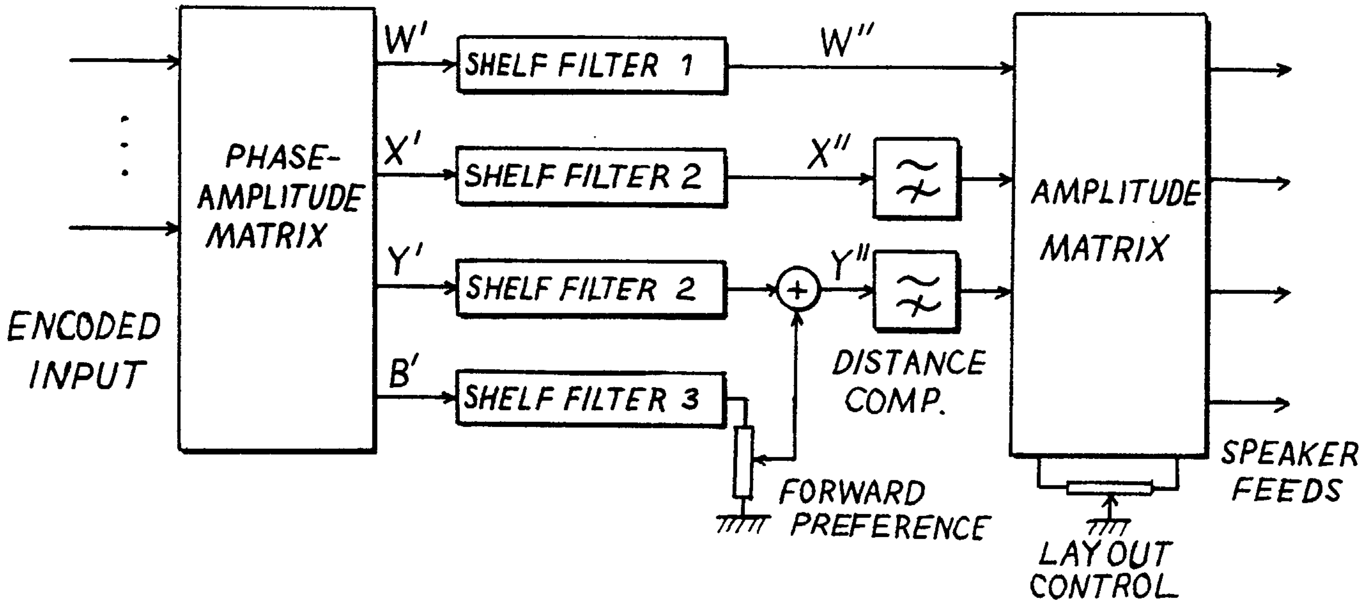


Figure 5. General architecture of horizontal Ambisonic decoder of the older generation.

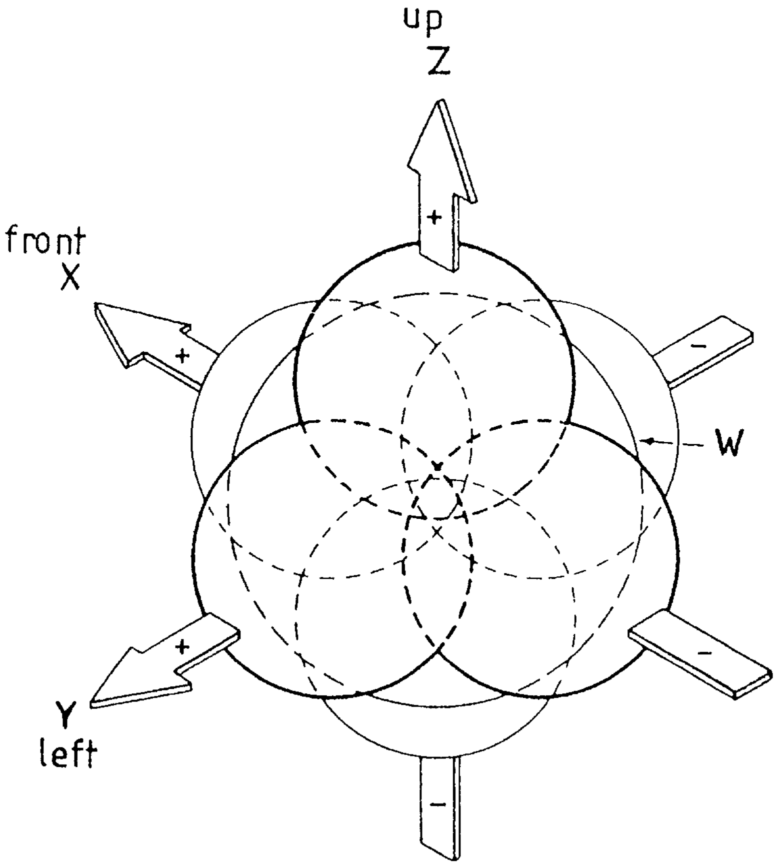


Figure 6. The directional gain patterns of with-height full-sphere (periphonic) B-format signals W, X, Y, Z.

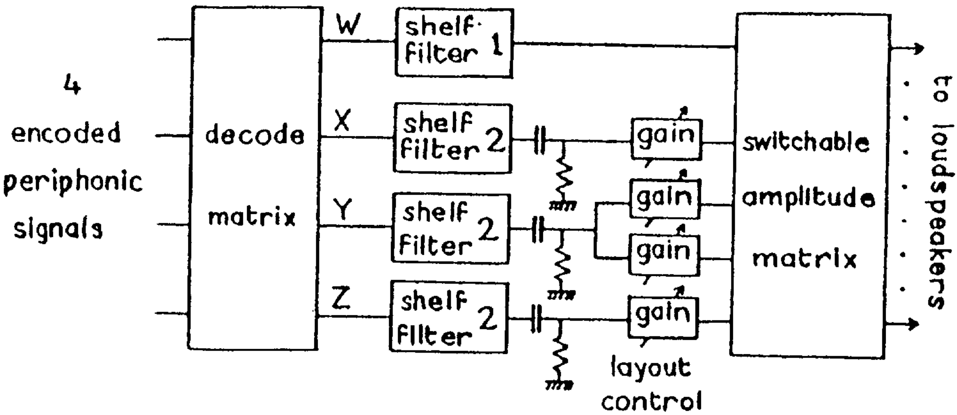


Figure 7. Block diagram of periphonic (full-sphere surround sound) Ambisonic decoder for 4-channel UHJ, and for intermediate B-format signals W, X, Y, Z.

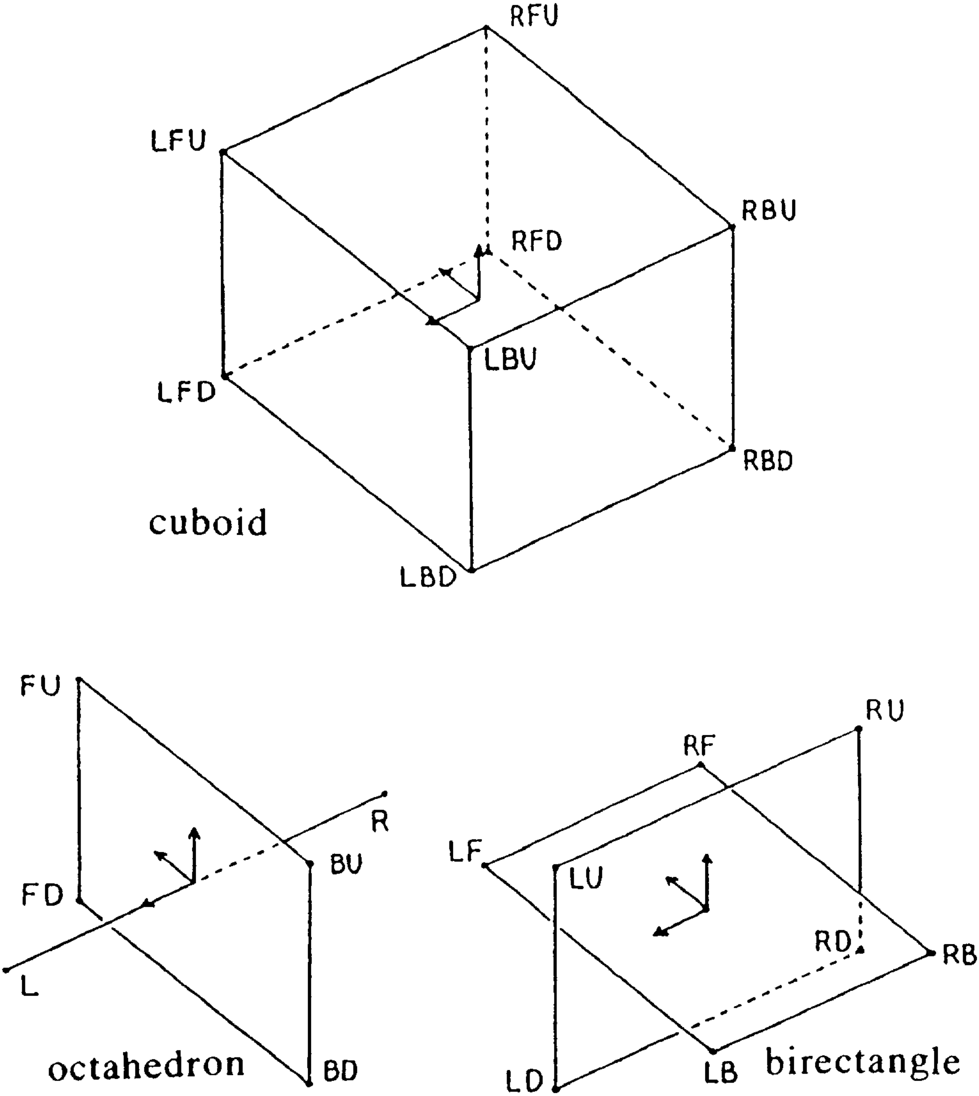


Figure 8. Three full-sphere loudspeaker layouts for Ambisonics, using cuboid, octahedron and birectangle layouts. L = left, R = right, U = up D = down F = front, B = back.

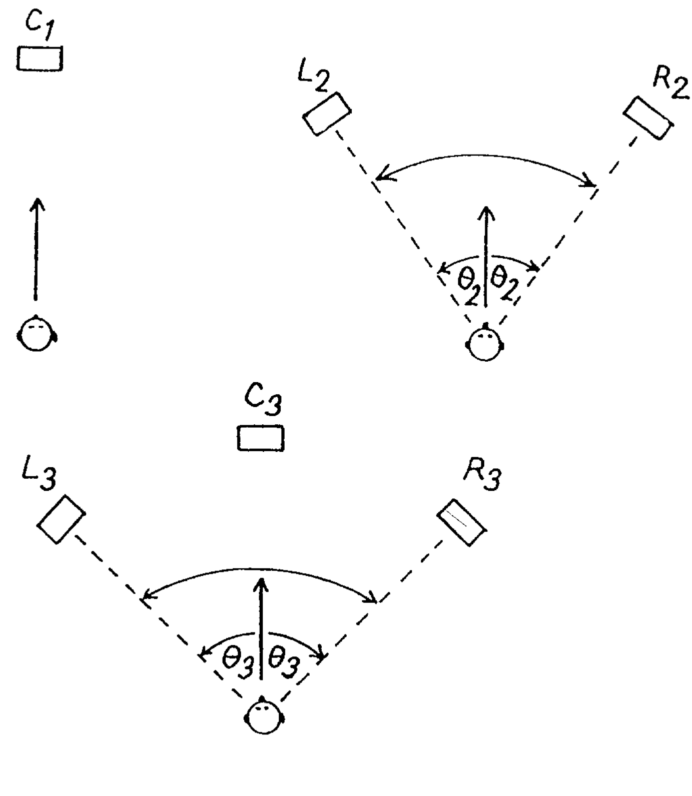
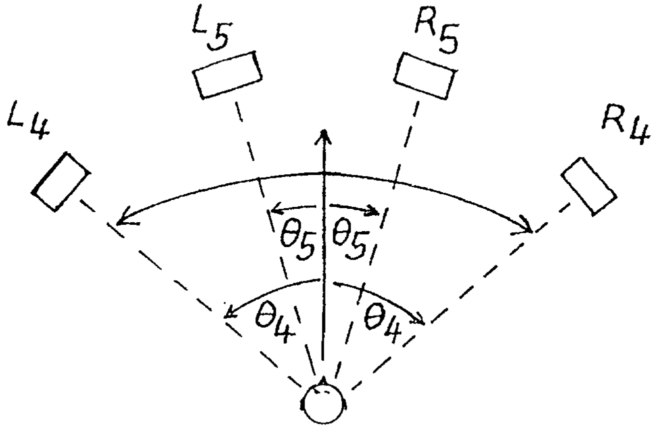


Figure 9.



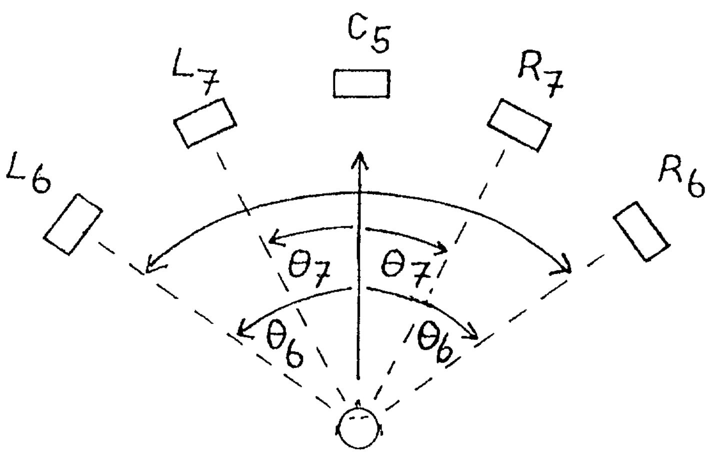


Figure 9 continued. Frontal stage stereo loudspeaker layouts with one to five loudspeakers, showing angles and notations for loudspeaker feeds. All loudspeakers are equally distant from a central listener and face towards the listener.

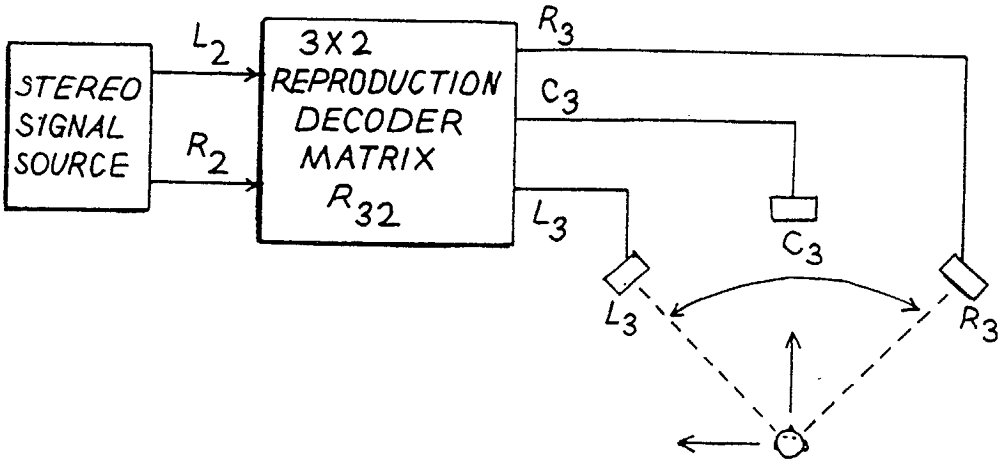


Figure 10. The conversion of 2-channel stereo source signals to loudspeaker feed signals for three loudspeakers.

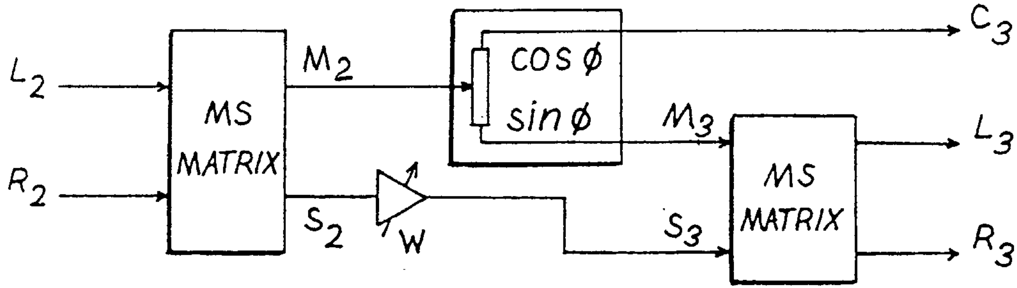


Figure 11. Basic form of 3 × 2 reproduction decoder, with optional width gain w adjustment.

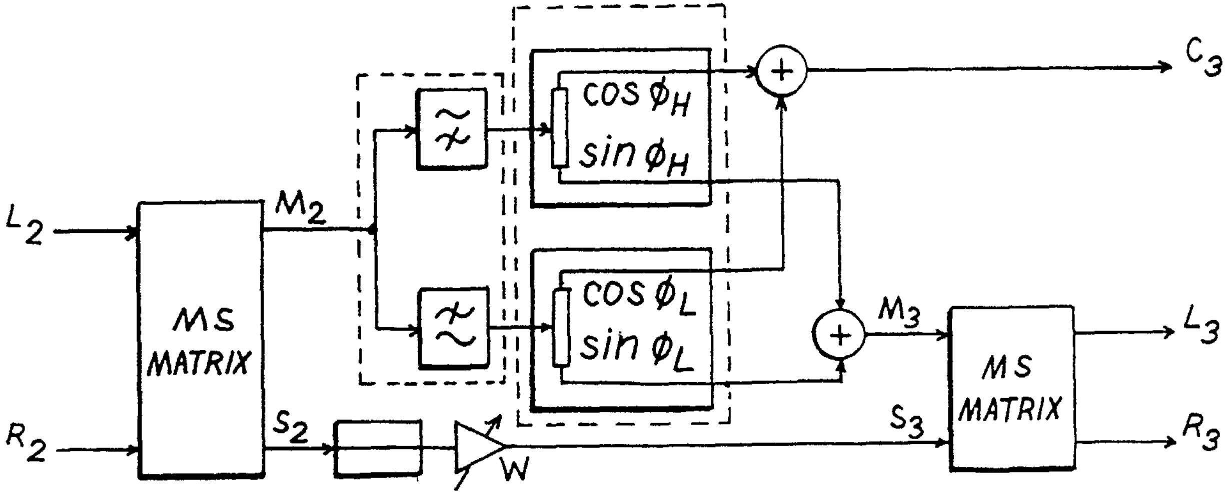


Figure 12. Frequency-dependent form of the 3 × 2 decoder of figure 11, with crossover network in sum (M_2) signal path and possible all-pass phase compensation for crossover network in the difference (S_2) signal path.

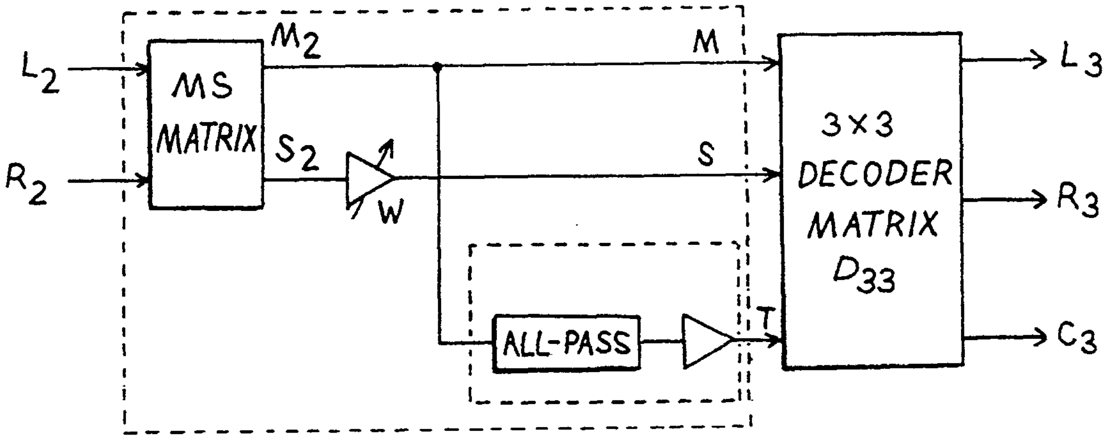


Figure 13. Alternative architecture for the frequency-dependent 3 × 2 decoder of figure 12.

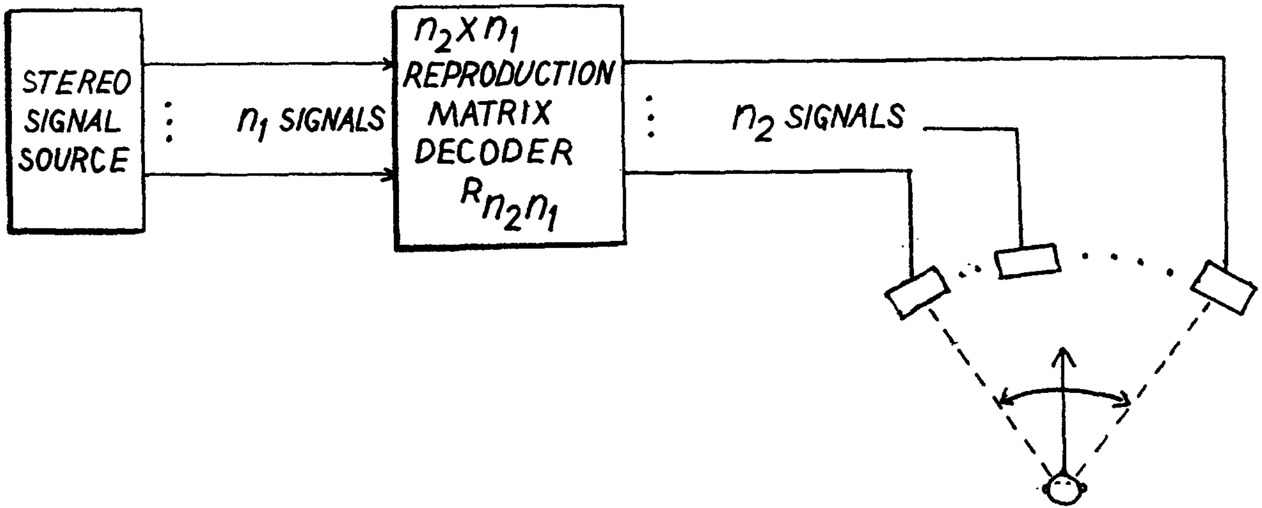


Figure 14. Schematic of reproduction matrix decoder for converting a stereo signal intended for a number n_1 of loudspeakers for reproduction via a latger number n_2 of stereo loudspeakers.

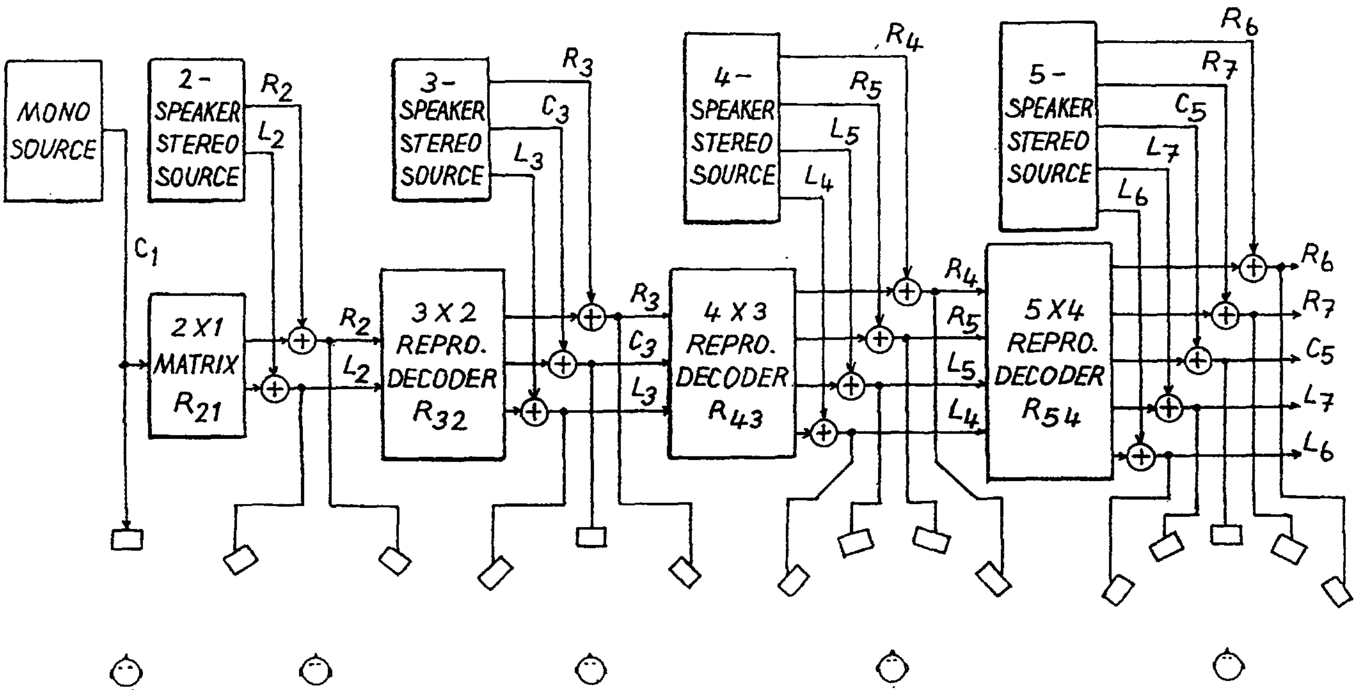


Figure 15. Showing how a psychoacoustic $n_2 \times n_1$ reproduction matrix decoder can be derived as the result of cascading psychoacoustic $(n+1) \times n$ decoders.

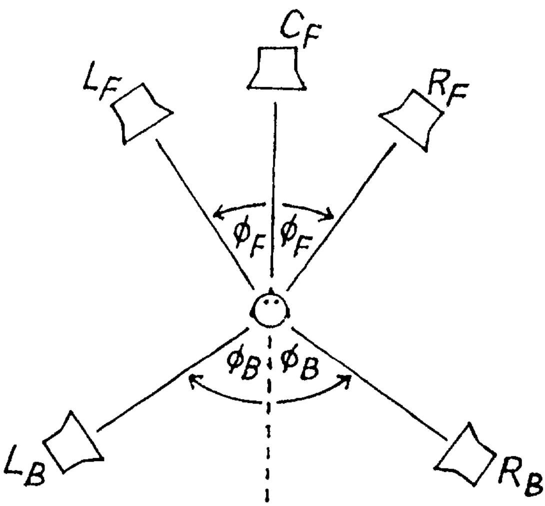


Figure 16. Five-loudspeaker layout for surround sound reproduction, supplementing a trapezium of loudspeakers with a front center loudspeaker.

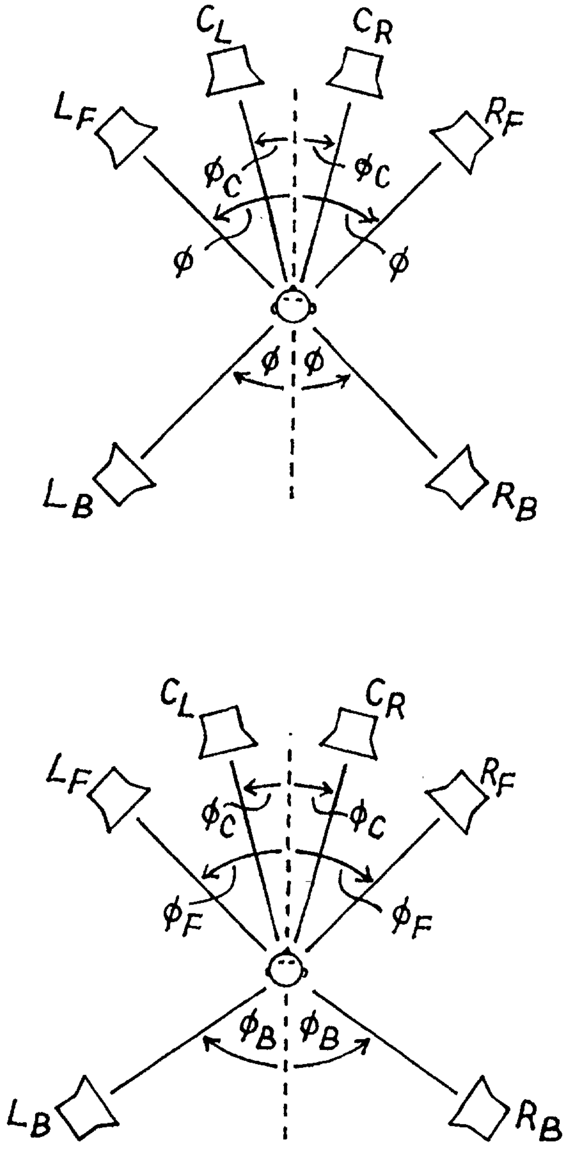


Figure 17. Six-loudspeaker surround-sound loudspeaker layouts based on adding two frontal loudspeakers to a respective rectangle and trapezium of loudspeakers.

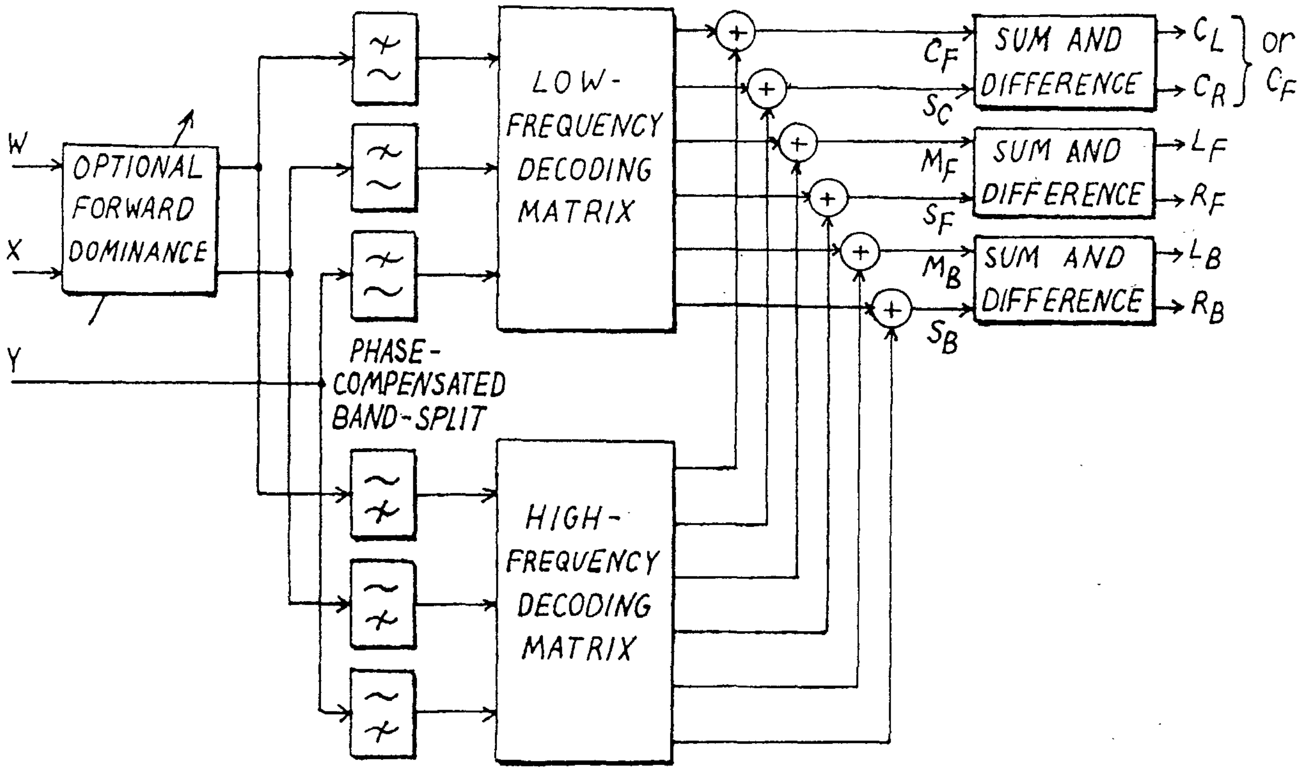


Figure 18. Architecture of B-format psychoacoustic surround-sound decoder for the loudspeaker layouts of figs 16 or 17.

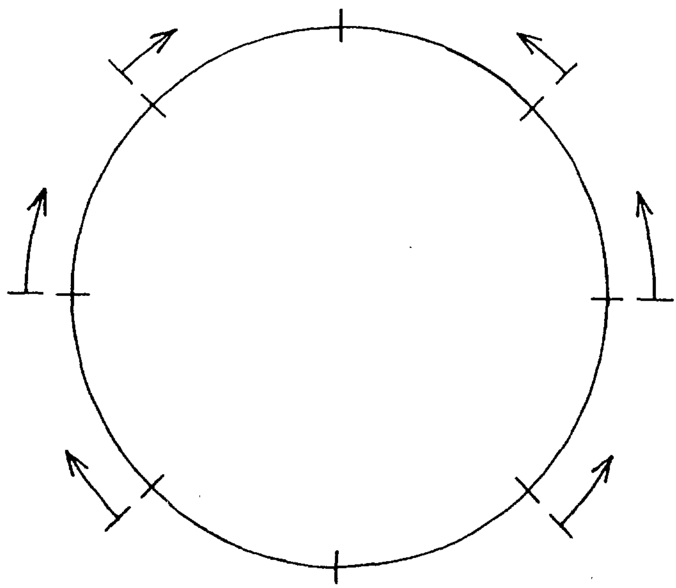


Figure 19. Effect of forward-dominance transformation of B-format signals on encoded direction.

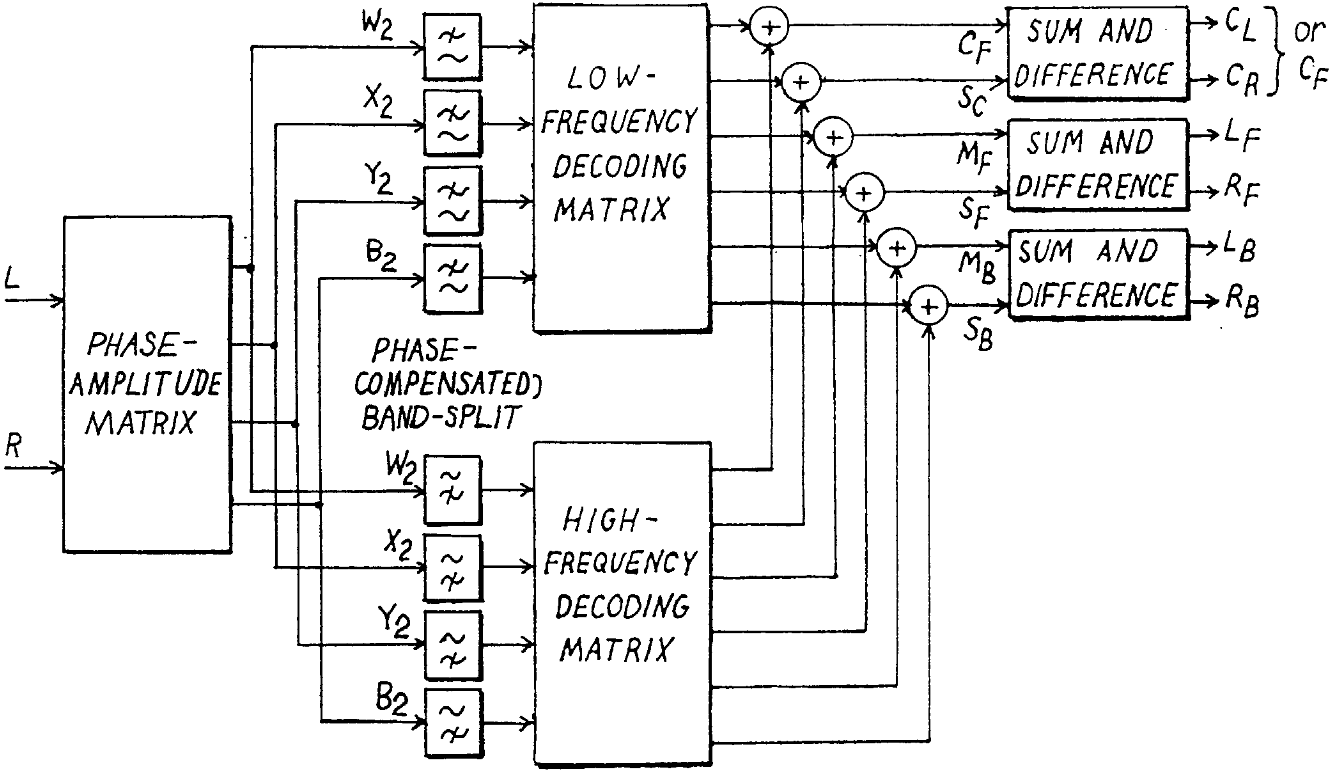


Figure 20. Architecture of psychoacoustic surround-sound decoder for 2-channel directionally encoded input signals for the loudspeaker layouts of figs. 16 or 17.

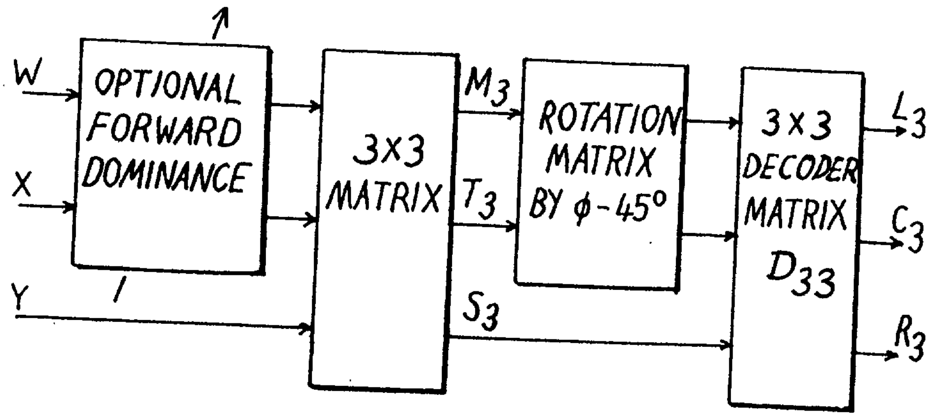


Figure 21. 3-loudspeaker stereo decoder for B-format signals.

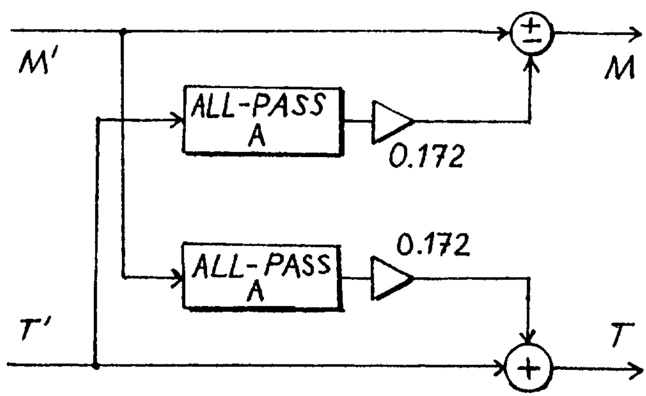


Figure 22. Form of the frequency-dependent rotation matrix used in the three-loudspeaker stereo decoder for B format of figure 21.