Microphone Techniques for 3-Channel Stereo

3450 (S-3)

Michael A. Gerzon
Technical Consultant
Oxford, UK

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

AES

This preprint has been reproduced from the author's advance manuscript, without editing, corrections or consideration by the Review Board. The AES takes no responsibility for the contents.

Additional preprints may be obtained by sending request and remittance to the Audio Engineering Society, 60 East 42nd St., New York, New York 10165-2520, USA.

All rights reserved. Reproduction of this preprint, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

AN AUDIO ENGINEERING SOCIETY PREPRINT

Microphone Techniques for 3-Channel Stereo

Michael A. Gerzon

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

Abstract

The forthcoming advent of 3-channel stereo in HDTV and DAB sound means that recording engineers will have to relearn the art of stereo microphone technique for 3- or more loudspeaker reproduction. Known and new microphone techniques are described, including those using matrix methods, including a 3-loudspeaker technique using a Soundfield microphone, and techniques using spaced mono microphones and spaced pairs or triples of one-point stereo microphones. Mono and 2-loudspeaker stereo compatibility will also be considered.

0. INTRODUCTION

Over the years, an almost infinite variety of stereo microphone techniques have been developed for conventional 2-channel stereo, many examples of which are described in the collection [1] of papers on stereo. There is no universal consensus as to the "best" microphone techniques, since the trade-offs between different conficting factors in the stereo illusion of an original sound field is ultimately a matter of personal subjective judgement. However, it is undeniable that there is an enormous body of empirical experience with these techniques, which allows users to make a considered judgement as to the selection of a technique to be used to meet particular circumstances and requirements.

However, there is much less experience on microphone techniques for 3-channel stereo using three loudspeakers at left, center and right positions in front of the listener. While 3-channel stereo has been used for over 50 years in the film industry, experience there is confined to the reproduction in relatively large auditoria, whereas the near-future advent of domestic 3-channel 3-loudspeaker stereo systems in the home will be used under the very different conditions of domestic listening, which use smaller listening areas and generally less dominant acoustics.

Generally, the domestic environment is a much more critical one for subtleties of sound. The smaller relative time delays of speaker signal arrival at the listener permits better stereo imaging and a greater ability to resolve directional detail than in large auditoria.

Thus, for practical purposes, 3-channel stereo microphone techniques can be considered a largely new art, despite a history going back as far as 1933 [2]. There is a temptation to believe that 3-channel stereo techniques are a simple extension of the art familiar in the 2-channel case, but as discussed in this paper, we believe that this view may be mistaken. The 3-channel case involves some new factors and methods that are largely new - in particular the use of matrix methods that are not completely obvious or trivial.

The work in this paper is based both on fundamental researches on the psychoacoustics of the 3-loudspeaker stereo illusion, previously reported by the author in a series of papers (refs. [3]-[6]), and on a limited empirical experience of recording for 3-channel reproduction in a domestic environment.

This paper should not be regarded as any kind of definitive guide to 3-loudspeaker stereo recording techniques, since a much broader range of empirical experience is required than currently possessed by the author (or probably anyone else at this stage) to make such a claim. Rather, the aim is to introduce a range of unfamiliar ideas that may prove to be of benefit as the 3-channel stereo art develops, and to suggest approaches that may not otherwise occur to practical recording engineers.

Any understanding of 3-loudspeaker stereo microphone techniques is crucially dependent on the psychoacoustics of the 3-loudspeaker stereo illusion, a generally more difficult problem than for the 2-loudspeaker case. Therefore this paper starts with a relatively brief discussion of this before describing a variety of microphone techniques, ranging from the traditional three spaced mono microphones used in Bell Telephone Laboratories' 1933 experiments [2] to the use of a sound field microphone, via a wide range of hybrid techniques using matrixing and spaced pairs of stereo microphone pairs.

1. 3-LOUDSPEAKER IMAGING

Generally speaking, reproduction from 3 loudspeakers is capable of much more precise stereo imaging than is possible from 2 loudspeakers, and many of the defects of 2-loudspeaker stereo can be greatly reduced. But such an improvement can only be achieved if the feeds to the three loudspeakers are carefully designed to optimise image quality.

The aim of stereo is not merely to give an illusion of a sound emerging from any desired phantom image direction between the outermost loudspeakers, but to give the best possible quality of localization of such phantom images. Even for listeners at an ideal central "stereo seat", 2-loudspeaker stereo is known to give non-ideal quality of localization of phantom images. Not only are such images unstable in position as the listener moves (either from side-to-side or if the head is rotated), but different methods of localizing sounds used by the ears and brain actually give different results about the apparent direction.

In particular, it has long been known, as shown by various researchers [7]-[10], that the ears/brain localize high frequency sounds rather further away from the center of the stereo stage than low frequency sounds for two loudspeakers. This effect cannot be properly corrected even by a "shuffling" circuit to make width frequency dependent, as shown by Harwood [8], and this is believed to be because in fact more than one localization mechanism is used over quite a broad common frequency range centered around 700 Hz, as discussed in the author's theoretical paper

[11] on directional psychoacoustics.

The contradiction between different localization mechanisms, and the phantom image instability with listener position and orientation, result in a poor stereo illusion and a relatively high degree of listener fatigue with conventional two-loudspeaker stereo. Three--loudspeaker stereo offers the opportunity not only to make phantom images more stable with listener position, as shown by Theile [12], but also to make different localization cues more consistent, as discussed in the author's papers [3]-[6]. The effect of such more consistent cues is improved subjective fidelity, lower subjective coloration, a lower listening fatigue, and the ability to hear more subtle details of sound within an overall mix.

Three loudspeakers is still far too few to create a perfect directional illusion, so that there are still some trade-offs between different factors, such as image stability with listener position, versus consistency of localization according to different auditory localization mechanisms in the ears and brain. However, the compromise among these trade-offs with three loudspeakers is far better than with two.

However, achieving such good trade-offs requires care, since three loudspeakers also offer the possibility of greater confusion than two, by offering even more sets of contradictory cues if fed by inappropriate signals. There is a temptation just to stick up a lot more microphones to feed the extra channel, and without care this can result in very bad sound. (The author recalls one notorious promenade concert broadcast by the BBC in 1977 during a series of experimental quadraphonic broadcasts which miked the orchestra with a record 57 microphones - and undoubtedly gave the most muddled orchestral sound he has had the misfortune to hear!)

We now consider in more detail some of the factors involved in creating a stable and consistent phantom image via three loudspeakers

1.1 Speaker Layouts

We cannot consider 3-loudspeaker microphone techniques in isolation from the actual loudspeaker layout used for reproduction. Figures 1(a) to 1(d) show four loudspeaker layouts commonly used for monitoring stereo. Fig. 1(a) shows conventional 2-loudspeaker stereo subtending an angle $2\theta_2$ at the ideally-situated listener, fed with respective left and right feed signals L_2 and R_2, and figs. 1(b) to (1d) show three-loudspeaker stereo layouts subtending an angle $2\theta_3$ at the ideally-situated listener, fed with respective left, center and right feed signals L_3, C_3 and R_3. In figs. 1(a) to 1(c), all loudspeakers lie at identical distances from the ideally-situated listener, whereas in fig. 1(d) they lie on a straight line, with the central loudspeaker closer than the other two. In figs. 1(a) and 1(b), the loudspeakers all face the ideally-situated listener, whereas in figs 1(c) and 1(d), the outer pair are "toed in" such that their axes cross in front of the ideally-situated listener.

The results obtained depend on the layout used. We believe that it is important to standardize monitoring arrangements to be such that all loudspeakers are at identical distances from the listener, so that any use of phase-coherent signals between the speaker feeds are properly reproduced at an ideally situated listener, and this equal-distance monitoring convention will be assumed in the rest of this paper when discussing microphone techniques.

However, it is undeniable that 3-loudspeaker layouts having the central loudspeaker closer than the outer two, such as in fig. 1(d), are often more practical in a listening environment, so that it is also necessary to specify how listening via such layouts should be achieved. It is suggested that the shorter distance of sound travel from the central loudspeaker to the ideally-situated listener in such layouts be compensated by the use of a time delay in the central loudspeaker feed, as shown in fig. 2, to restore phase coherence, the delay being equal to the difference in distances divided by the speed of sound in air, which is around 340 m/s. Thus, for ½ m difference in distance, a 1.47 ms delay should be used to feed the center loudspeaker.

A small gain decrease to compensate for the closer loudspeaker placement could also be used, but this is found to be of much less importance, and can affect the perceived balance caused by reflected sounds in the listening environment. In general, it is found that gain adjustment is a much less satisfactory way of compensating for a closer central loudspeaker than time-delay adjustment, and should not be used as the main method of adjustment for serious listening and monitoring when a layout such as figure 1(d) is used where the central loudspeaker is closer than the other two. We have found that adjustment of path length differences and delays to an accuracy of as little as about 1 cm (a time difference of less than 30 μs) is worthwhile for optimum monitoring or listening at the ideally-situated listener.

The total angle $2\theta_2$ and $2\theta_3$ subtended by the loudspeaker layout also needs a degree of standardization, although this is less critical than equality of distance or time-delay compensation. For two-loudspeaker stereo, a total subtended angle $2\theta_2$ of 60° is a generally used standard for audio-only applications, and certainly should not be larger than this for reasons of imaging quality and stability. For video applications, $2\theta_2$ should preferably be between 45° and 60°. There is no universal agreement on the subtended angle $2\theta_3$ for the three-loudspeaker layouts, but a value of between 45° and 60° is widely used for video and film applications, whereas a value of between 60° and 90° is satisfactory for audio-only applications.

The wider the subtended angle, the poorer is the imaging stability and quality. Based both on the researches of Theile [12], and on the theoretical and empirical studies of the author reported in refs. [4] and [5], it is found that as a rough rule of thumb, the degree of apparent angular movement of a phantom image as the listener moves from side to side or as he/she rotates the head is proportional to the <u>square</u> of the subtended angle of the layout, so that widening the angular layout by a factor 1½ gives $(1\frac{1}{2})^2 = 2\frac{1}{4}$ times the amount of angular movement of the image for a given movement of the listener.

In general, image stability is more critical in video applications where sounds are associated with a visual image [13] than in audio-only applications, hence the preference for a narrower subtended angle for video than for audio-only applications. However, even in audio-only applications, image stability affects the size of the useable listening area and the imaging quality, and there is a trade-off between these and the benefits of image width and spaciousness. Taking all factors into account, a subtended monitoring angle of $2\theta_3 = 60^{\circ}$ is probably a good general purpose compromise, while a slightly larger angle of around 75° may be used in predominantly audio-only applications.

However, for monitoring purposes with audio-only applications, the precise subtended angle is probably not a critical factor in that monitoring decisions made for one angle will not prove to be grossly misleading for another, although there will be perceptible differences in the spaciousness and even subtleties of perceived tonal balance. Wider monitoring angles have the benefit of exaggerating any weakness in imaging quality of a recording technique, making them more analytical for refining microphone techniques.

The "toeing in" of the outer loudspeakers, as shown in figs. 1(c) and 1(d), and which can also be used for two-loudspeaker monitoring, uses the polar pattern of the loudspeaker to enlarge the useful listening area. This method has been well-known since the 1950's (e.g. see Snow [14]), but the degree of "toeing-in" that gives best results is very dependent on the precise characteristics of the loudspeakers, and must be determined empirically. For monitoring applications, it is also necessary to ensure that for an ideally-situated listener, the off-axis loudspeaker response heard by the listener from the outer loudspeakers remains a good match to the on-axis response heard from the central loudspeaker.

1.2. 3-Channel Panpot Laws

In his paper [5] on 3- and 4-channel stereo panpot laws, the author considered a variety of requirements on optimum panning of phantom sound images. These included not merely the requirement of achieving a good phantom image in every direction between the outer loudspeakers, but also requirements on improved imaging stability and quality, both for an ideally-situated central listener and for listeners to either side of such an ideal position. The optimization of such panpot laws is complicated by the fact that single sounds are rarely panned in isolation from other sounds in other directions. It is little good if one optimises the localisation of one sound position if the results in other positions present at the same time are contradictory. By way of example, there is a commercially-available process for reproducing two-channel stereo via three loudspeakers that gives excellent reproduction of central images all the way across a listening area, and which reproduces left center and right directions broadly correctly for an ideally situated listener, but which reproduces both left and right positions over to the nearest outermost loudspeaker for listeners to one side of the ideal listening position! Such an extreme geometric distortion of the stereo image is generally unacceptable for serious uses.

In refs. [4] and [5], the author took the view that a good panpot law for multi-loudspeaker stereo should not merely have good performance for each sound position taken in isolation,but that there should be, as far as practical, a consistency in the localization quality of different positions. In particular, it was felt important to minimize any geometric distortions of relative positions with a reproduced sound stage as the listener moves around the listening area, and this was ensured by requiring that the instability of images with listener movement (measured by the quantity r_E computed in those papers) should be similar for all sound positions.

On the basis of this requirement combined with the requirement that different auditory localization mechanisms used by the ears should give consistent results, the author showed in ref. [5] that the 3-loudspeaker panpot law shown in fig. 3 was generally optimal. This law is a little different from those proposed previously in that, as a sound is panned from left to right, the gain of the furthest speaker from the sound position becomes <u>negative</u> (i.e. of opposite polarity) to start with, returning to zero and then becoming positive only half way from edge to the center.

However, this "optimal" 3-loudspeaker panning law does <u>not</u> give perfectly stable center-stage images, since it has has about -8.5 dB crosstalk from the center loudspeaker to each of the side loudspeakers. This is because of the requirement that central image localization quality should be consistent with the localization quality of other phantom image positions to either side of center. In practice, for a given loudspeaker layout subtended angle, such 3-loudspeaker phantom images near the center suffer from around one quarter the degree of instability (unwanted angular movement of the image with listener movement) than do similarly-positioned images via two-loudspeaker stereo.

The 3-loudspeaker panpot law of fig. 3 may be implemented on a mixing console to position closely-positioned mono microphone signals within a 3-channel stereo mix, but when people refer to "stereo microphone technique", they are usually referring to methods of directly transducing a live directional sound field, rather than of simulating such an effect via panpots. Nevertheless, the panpot law of fig. 3 is a reference point for evaluating the imaging qualities of such stereo microphone techniques.

At an opposite extreme to this "optimum" 3-channel panpot law is the familiar constant-energy or sine/cosine panpot law [15] used for two--loudspeaker stereo shown in fig. 4. This may be regarded as a (rather poor) 3-loudspeaker law in which the center speaker feed happens to be zero.

However, in ref. [4], the author described a psychoacoustically optimized method of feeding a three-loudspeaker stereo reproduction system with three signals L_3, C_3 and R_3 derived by a frequency-dependent 3×2 matrixing process from just the two signals L_2 and R_2 of conventional two speaker stereo, as shown in fig. 5. While such derived three--loudspeaker feeds are undoubtedly less good than optimised "genuine" three-loudspeaker stereo, the results obtained are generally quite good, and in the author's opinion frequently a great deal better than many less good "genuine" three-loudspeaker stereo microphone techniques. Such a method therefore deserves serious discussion in a paper devoted to three-loudspeaker stereo microphone technique.

The 3×2 conversion matrix at each frequency has the form shown in fig. 6, where an MS matrix is a 2×2 matrix network having the effect

$$M = 0.7071\ (L + R)$$

$$S = 0.7071\ (L - R) \qquad (1)$$

which takes left and right signals into M ("sum") and S ("difference") signals, with the inverse matrix

$$L = 0.7071\ (M + S)$$

$$R = 0.7071\ (M - S) \qquad (2)$$

also being an MS matrix having identical form.

The 3×2 conversion matrix converts the input 2-channel stereo into MS form, and then splits the M signal, via a constant-power sine/cosine pair of gains, into a component with gain $\cos\phi$ to feed the center loudspeaker feed C_3 and a component with gain $\sin\phi$ to feed an output MS matrix, with the difference signal S as the other input, to provide left and right signals L_3 and R_3 for the two outer loudspeakers. The choice of the angle ϕ depends on psychoacoustic considerations, but in general, the result is to give output speaker feed signals given by the equations:

$$L_3 = \tfrac{1}{2}(\sin\phi + 1)L_2 + \tfrac{1}{2}(\sin\phi - 1)R_2$$

$$C_3 = 0.7071\cos\phi(L_2 + R_2)$$

$$R_3 = \tfrac{1}{2}(\sin\phi - 1)L_2 + \tfrac{1}{2}(\sin\phi + 1)R_2\ . \qquad (3)$$

It can be shown that this gives exactly the same total energy from the three loudspeakers as was originally given from two, whatever stereo positioning signals may have been given, so that overall level balance of the original 2-channel recording is accurately maintained via three loudspeakers.

As shown in ref. [4], it is found that the optimum value of ϕ for good imaging quality depends on the frequency, with $\phi = 35.3^{\circ}$ a good choice at frequencies up to around 4 or 5 kHz, and $\phi = 54.7^{\circ}$ a good choice above about 5 kHz, with a smooth transition between the two values of ϕ near the transition frequency of 5 kHz. Fig. 7 shows how a crossover network can be incorporated in the matrix converter of fig. 6 to achieve such a frequency-dependent value of ϕ.

The resulting 3×2 matrix converter is given by the equations

$$L_3 = 0.7887\,L_2 - 0.2113\,R_2$$

$$C_3 = 0.5774\,L_2 + 0.5774\,R_2$$

$$R_3 = -0.2113\,L_2 + 0.7887\,R_2 \qquad (4)$$

at low frequencies below 5 kHz, for which $\phi = \arctan 2^{-\frac{1}{2}} = 35.3^{\circ}$, and

$$L_3 = 0.9082\,L_2 - 0.0918\,R_2$$

$$C_3 = 0.4082\,L_2 + 0.4082\,R_2$$

$$R_3 = -0.0918\,L_2 + 0.9082\,R_2 \qquad (5)$$

at high frequencies above 5 kHz, for which $\phi = \arctan\sqrt{2} = 54.7^{\circ}$.

Figure 8 shows the resulting 3-loudspeaker panpot law at low and high frequencies obtained when a two-channel recording following the sine/cosine panpot law of fig. 5 is matrixed into 3 loudspeaker feeds via eqs. (4) and (5) respectively.

At lower frequencies, below 5 kHz, the resulting panpot law is closer to the optimum 3-channel law of fig. 4 for near center positions, having -6dB cross-talk from the center to each of the two outer loudspeakers, whereas the high frequency law of fig. 8 is a better approximation to the optimum law for edge-of-stage sound positions. While the low and high frequency laws of fig. 8 are not ideal performers (as analyzed in detail in section 7 of ref. [4]), they indicate a minimum standard known to give reasonable results.

2. SPACED MICROPHONE TECHNIQUES

2.1 "Discrete" Spaced Technique

The oldest microphone technique for 3-loudspeaker stereo is the use of 3 separate omnidirectional microphones, typically spaced in a straight line at 3 or 4 m intervals, feeding separate loudspeakers [2]. Originally, the rationale of this technique was the "wavefront reconstruction curtain of microphones" theory, that saw the microphones and loudspeakers as sampling and recreating the original sound wavefronts passing through a notional "curtain" containing the microphones, but it was soon realized [14] that there were far too few microphones for such a wavefront reconstruction explanation of the operation to make sense. Rather, the spaced microphone technique worked [14,2] by psychoacoustic means where the apparent sound image was formed by a combination of amplitude and time-delay cues, based on the Haas or precedence effect [16], [17].

The problem here is that for each original sound source, there are three sound arrivals at the listener [14] generally arriving at time intervals spaced several ms apart. Such long time delays are too long to give sharply-defined well-fused images, and three separated sound arrivals rather than two tend to give extra spurious information not present in natural sounds. As shown in ref. [18], such extra time-delayed information not only affects the image postion of the sound, but also alters other perceived spatial qualities.

It is doubtful that, if one were designing from scratch a 3-loudspeaker panpot law using both amplitude and time delays from three loudspeakers to create illusory phantom images, that one would arrive at anything like the combination of amplitude gains and time delays given by the classic 3 spaced microphone technique.

Its continuing popularity must be seen as resulting from a combination of factors. It is an old technique with considerable experience in its use, and so is highly predictable for its practitioners. It uses omnidirectional microphone types, which are the microphone characteristics that are easiest to make with a high level of technical accuracy, which itself tends to give a less colored sound. The relatively close microphone placement used with this technique tends to reduce unwanted external acoustic interference, but the broad microphone spacing tends to mean that there is a fairly uniform coverage of a broad sound source, such as an orchestra, with few resulting balance problems (except in the case of a very deep orchestral stage, which requires the use of very high microphone positions to give reasonable coverage).

The added time delays with this technique also add a sense of spaciousness that, while not present in the original sound, is found to be pleasing to many. These time delays produce an "enhancement" similar to that produced in popular music production by some digital delay processing effects, although the fact that the delays vary with source position mean that the enhancement effect is less "mechanical" than with digital effects units. Finally, this spaciousness enhancement is heard not only by listeners at an ideal listening position, but across a very wide listening area.

Its main weakness is in the area of stereo directional imaging, where sounds tend to be concentrated in three areas around the three loudspeakers, with a rather vague wash of sound elesewhere. In particular, at higher frequencies above around 3½ kHz, one hears the sound as "splashing" from the three loudspeakers rather than as sharply localized at a single phantom image position. This is more distracting under domestic listening conditions than in larger auditoria.

Because it is not based on any well-founded principles of optimum directional sound imaging, there is considerable scope for variation of microphone positioning with this technique, not only in the spacing between microphones, but in the positioning of the center microphone in front of the line joining the left and right microphones. Such "forward" central microphone positions can help make central sound images more precise, depending on circumstances, by making their sound arrive at the center microphone sooner.

The advent of cheap high-quality digital delays has added a new element of post-production control to 3-channel spaced microphone techniques, in that it is now easy to vary the relative time delay (or advance) of the center microphone relative to that of the two outer microphones. For example, adding a time delay can ensure that for center-stage sounds, the arrivals from the three microphones are at a substantially identical time, which helps to ensure image coherence and reduce multiple sound arrival effects. It is certainly worth experimenting with adjusting time delay and gain of the central microphone on such recordings to optimize the subjective effect via the standardized monitoring loudspeaker layout.

The use of three spaced microphones can undeniably give very impressive and even spectacular results, but it must be recognized that at least a part of these are due to the addition of time-delay "effects" due to microphone spacing rather than to qualities present in the original sound field. This microphone technique is perhaps most justified when used with sound reproduction in auditoria, where similar large time delays are typically in any case produced by the different distances of the three loudspeakers from the listener.

2.2 Matrixed Spaced Techniques

It is not necessary to feed spaced microphones to independent loudspeakers, as was realized right from the earliest experiments in 1933 [2]. Rather, it is also possible to feed the microphones via a matrixing or cross--blending arrangement.

In the original 1934 paper [2], two such matrixing methods for feeding three loudspeakers were suggested. One method, the 2-2-3 method, fed the outputs of two widely spaced omnidirectional microphones to two outer loudspeakers and their average to the center loudspeaker. The other, 3-2-3, method, used three microphones as above, but derived two channels by mixing the center microphone signal into the left and right signals, with three-loudspeaker reproduction using the same process from the two channels as in the 2-2-3 case.

In the 1950's, Klipsch [19]-[21] performed an interesting series of experiments with these techniques, and made a remarkable observation which the present author was inclined to disbelieve until he duplicated it! The use of two widely spaced omnidirectional microphones, when reproduced via just two loudspeakers, is well known to give very poor phantom imaging, with just two predominant "pools" of sound at either side, a great deal of left-right "splash" of high frequency sounds, and a tendency on continuous sounds to give a great deal of random wandering of sound images. Klipsch claimed [19]-[21] that when such recordings were fed to three loudspeakers, with an average signal fed to the central loudspeaker, that a range of phantom image positions between the three loudspeakers became audible!

Klipsch found that the results were quite critically dependent on using an appropriate gain to feed the center loudspeaker, and that the stereo width becomes narrow unless a wide subtended loudspeaker angling $2\theta_3$ was used (equal to 90°), in conjunction with a "toeing in" of the outer loudspeakers. Klipsch used the layout of fig. 1(d), but we have found that his observations also apply to the layouts of figs. 1(b) and 1(c).

However, we have found that the loss of stereo width is greatly ameliorated, and the subjective results greatly improved over the loudspeaker layouts of figs. 1(b) and 1(c) with a spaced pair of omnidirectional microphones, if the three loudspeakers are instead fed via the matrix of figs. 6 or 7 and eq. (3). With a frequency-independent value of the parameter ϕ between 45^o and 55^o, or using the frequency-dependent 3×2 matrix of eqs. (4) and (5) and fig. 7, it is found that a spaced pair of omnidirectional microphones, or a spaced pair of boundary layer microphones. spaced around 3 to 6 m apart, can actually give reasonably good phantom imaging of sound sources across the stereo stage, with quite good image stability for central images. This is somewhat surprising, and not understood from a theoretical psychoacoustics viewpoint.

The resulting spaced omnidirectional technique retains much of the "spaciousness" and "width" of its two-speaker counterpart, but also retains the defect of a great deal of left-right "splash" of high frequency sounds above 3½ kHz.

The use of just two spaced microphones, rather than three, to feed three loudspeakers does have the advantage of providing less "muddle" due to the smaller number of time delays involved, and the use of the optimized 3×2 matrix decoder to provide these feeds is a significant improvement over the earlier Bell/Klipsch method.

For many practical applications, such as "atmosphere" or "audience" microphones, the use of a spaced omnidirectional or boundary layer microphone pair fed via an optimized 3×2 matrix decoder to three loudspeakers may be a very practical technique, notably for video and TV applications where low microphone visibility is also important. The writer has also found that a pair of boundary layer microphones taped to the sides of PA loudspeaker stacks, fed via a 3×2 matrix, can give excellent pick-up of the "backline" in rock recordings in smaller music venues, for mixing with the outputs from a PA mixing desk.

While the remaining left-right "splash" effect at high frequencies makes this technique suboptimal for the most purist applications, it does seem to work well as a part of a mix also involving other sound sources such as PA feeds or closely-microphoned actors, and can be practical in cases where there are severe restrictions on microphone positioning (e.g. in live recording of theatre or opera productions). Also, when a part of an overall mix, adjustment of the parameter ϕ provides a useful method of optimizing the effect of the mix in post-production.

2.3 Spaced Stereo microphones

The main problem with traditional spaced microphone techniques is that they are reliant predominantly on rather long time delays to localize sounds. However, the use of widely spaced microphones does mean that the microphones can be used relatively close to a large band or orchestra without problems of balance, since a microphone is present close to each part of the band.

However, there is no law that says that spaced microphones need be mono microphones, and the writer has experimented with the use of spaced microphones each of which is a spatially-coincident stereo pair. There is nothing new in the idea of mixing together the outputs of two (or more) stereo microphones placed at different positions, but in ref. [22], the writer made what appears to be a novel observation that it is preferable that the individual stereo microphones in such use need not themselves be left-right symmetrical in nature. Rather, he suggested the use of a pair of stereo microphones, spaced apart by 3 to 6 metres, where the individual stereo microphones were highly assymetric, using a different polar diagram for left and right, but where overall left-right symmetry is restored by making the left-positioned stereo microphone pair a mirror image of the right-positioned microphone pair.

As shown in fig. 9, the left stereo microphone has two outputs, a "left" output denoted M_L and a "right" output denoted S_L. The right stereo microphone also has two outputs, a "right" output M_R with a similar polar diagram to M_L and a "left" output S_R whose polar diagram is the mirror image of that of S_L. Typically, as shown in fig. 9, M_L and M_R are cardioid microphones pointing towards the nearest edge of the sound source stage being recorded, with S_L and S_R being figure-of-eight microphones with axes at right angles to that of the cardioid and facing inwards towards the center of the sound stage.

Thus each stereo microphone consists of a cardioid and orthogonal coincident figure-of-eight, which are the M and S signals of the usual 2-channel MS stereo microphone technique (hence the notations M_L and M_R and S_L and S_R), but they are <u>not</u> used with MS matrixing in the present application.

In ref. [22], the author proposed deriving 2-channel stereo signals L_2 and R_2 simply by adding the two "left" microphone signals from the two stereo microphones together for L_2 and the two "right" microphone signals for R_2, i.e.

$$L_2 = M_L + S_R$$

$$R_2 = M_R + S_L \quad , \qquad (6)$$

which is simply done by mixing the outputs of the two stereo microphones. For such a 2-channel stereo recording, sounds from the left half of the sound-source stage arrive at the left stereo microphone before the right, and so by the Haas or precedence effect [16], [17], have a recorded stereo position determined predominantly by the two channel gains in the pick-up of the left stereo microphone. Similarly, sounds on the right hand half of the direct sound-source stage are predominantly imaged in stereo by the right stereo microphone pair. Sounds near the middle of the sound-source stage are imaged by both stereo microphones, which should ideally have an S signal gain to allow pick-up of a central sound source from the middle via each stereo microphone on its own.

The use of two stereo microphones in this manner allows the use of spaced microphones while having the virtues of amplitude-panning for determining stereo positioning via each of the two stereo microphones. This technique is found to give excellent phantom sound imaging, except for a slight image broadening and coloration at the very center of the sound-source stage due to interference effects from the two stereo microphones, and retains the virtues of allowing close placement and good balance of the traditional spaced omnidirectional technique. More surprisingly, as reported in ref. [22], it actually improves on the sense of spaciousness of the traditional technique, partly due to the antiphase pick-up of reflected sounds beyond the bounds of the direct sound-source stage, as shown in fig. 9.

As noted in ref. [22], the optimum gain of the S microphones is found to be that such that the on-axis gain of the S microphone is about $\frac{1}{2}\sqrt{3} = 0.866$ of the on-axis gain of the cardioid M microphones. due to the evenness with which the reverberant sounds are picked up.

For microphone spacing different to the width of the direct sound-source stage, as shown in fig. 10, the two stereo microphones may be angled so that the axes of the cardioids points towards the edge of the sound-source stage.

While this spaced stereo microphone technique is left-right symmetrical and works well when fairly closely positioned microphones are essential (such as when there are audience noise problems or for recording a backline in the presence of PA equipment, it is a 2-loudspeaker stereo technique as so far described. The question arises as to the optimum way of feeding the two stereo pairs to three loudspeaker channels L_3, C_3 and R_3. The "obvious" way of doing this - to feed the left stereo microphone to L_3 and C_3 and the right stereo microphone to C_3 and R_3 - results in the earliest-arrival pick-ups from the two stereo microphones being "squeezed" into a too-narrow stage, resulting in a hole in the middle of the reproduced stereo in which phantom imaging is largely absent.

What is needed is a way of feeding the three loudspeakers by the left stereo microphone (and a mirror-symmetrical way for the right stereo microphone) that: (i) preserves the total signal energy, and that (ii) reproduces the <u>left half</u> of the stereo stage picked up by the left stereo microphone with substantially optimal imaging quality via three loudspeakers, with little emphasis placed on the imaging quality of the right half of the sound stage, since this is predominantly picked up by the right stereo microphone. "Optimal imaging quality" here means roughly approximating to the 3-loudspeaker panpot law of figure 3.

This approximation cannot be exact, but the following are two possible compromises for feeding the microphone method of figs. 9 or 10 to three loudspeakers with good imaging quality and avoiding an excessive hole in the middle:

$$L_3 = M_L + 0.32\ S_R - 0.16\ S_L - 0.10\ M_R$$

$$C_3 = 0.20\ M_L + 0.96\ S_R + 0.96\ S_L + 0.20\ M_R$$

$$R_3 = -0.10\ M_L - 0.16\ S_R + 0.32\ S_L + M_R \qquad (7)$$

or, with less of a hole in the middle but slightly less stable near-center images,

$$L_3 = M_L + 0.4164\ S_R - 0.1434\ S_L - 0.1\ M_R$$

$$C_3 = 0.2\ M_L + 0.9252\ S_R + 0.9252\ S_L + 0.2\ M_R$$

$$R_3 = -0.1\ M_L - 0.1434\ S_R + 0.4164\ S_L + M_R\ . \qquad (8)$$

The above examples of microphone characteristics for the two stereo microphones and the matrixing into three loudspeaker feeds is only an example, and other examples are possible. However, these examples do illustrate that simply naively mixing two standard stereo microphones into three loudspeakers is almost certainly not optimum, both as regards image positioning and quality. Rather, the choice of the left and right polar diagrams and orientations of each stereo microphone, and the choice of matrixing so that the earliest sound arrivals approximate the optimum 3-loudspeaker panpot law of fig. 3, is generally nontrivial, and requires the use of equipment designed for this application.

Wherever matrixing techniques are used, it is difficult to adapt "standard" recording equipment for the purpose, and it is necessary to use equipment designed for this purpose, where the matrixing coefficients, both positive and negative, can be "dialled up" or recalled from program memory or plug-in resistor cards. The use of microphone techniques involving more than one stereo pair makes the availablity of such specialist matrixing equipment essential if optimum results are to be obtained. It is also necessary for recording engineers wishing to innovate in microphone techique (rather than to adopt someone else's technique "off the shelf") to study the psychoacoustics of 3-loudspeaker phantom imaging [4], [5] and to use the mathematics of matrixing equations, in order to develop an intuition as to the possibilities - which actually means that the recording "engineer" must acquire the attitude of an engineer in the technical sense - there are certainly many recording engineers with the appropriate combination of artistic and purely technical skills.

2.4 Three Stereo Microphones

One problem with the use of two stereo microphones is that directional sound sources near the center of the sound-source stage may be picked up by both stereo microphones off the axis of the sound source, resulting in a loss of tonal quality - this particularly affects instruments such as trumpets and saxophones. This suggests that three microphone positions should be used, with a central stereo microphone to cover the center of the sound-source stage. Unlike the left and the right stereo microphone, such a center microphone should, on its own, be left-right symmetric in order to preserve overall left-right symmetry.

However, the use of three stereo microphones means that there are now three sound arrivals from every sound source, which can result in a more muddled overall sound and the risk of a high degree of comb-filter coloration. Lack of experience with these techniques means that the author cannot recommend optimum techniques for three stereo microphones. The available parameters here are not only the polar diagrams of the three stereo pairs, and the matrixing of them into the three loudspeakers such that the first-arrival sounds are optimally localized according to the panpot law of fig. 3. Also, as in the case of the 3-omnidirectional spaced technique, one also has the options of varying the distance of the center stereo microphone in front of the line joining the outer two, and of subjecting the outputs of the center stereo microphone to a relative stereo time delay (or advance). The number of parameters involved is so large that there is room for many individual variations.

However, the use of three stereo microphones is already well away from a purist technique, and the validity and utility of such techniques may well prove to be controversial.

The use of two-output stereo microphones can also be replaced by the use of microphones giving as outputs three independent polar characteristics at a point - such as are produced from a sound field microphone [23], [24]. As we shall see in the next section, even feeding such sound field information into three loudspeakers is not a trivial thing, and the number of possible variations when two or three spaced sound field microphones is used is extremely large.

3. COINCIDENT MICROPHONE TECHNIQUES

3.1 The Sound Field Microphone

One of the basic techniques used for two-loudspeaker stereo has been coincident microphone techniques, whereby two directional microphones at a single point in space, pointing in different directions, are used to pick up a stereo effect. In effect such microphones act as a panpot in response to incident sound directions, by giving two channels that differ only in amplitude gain for each sound direction. While such techniques are now quite well understood for two-loudspeaker use (e.g. see [1], [25]), there are unexpected problems for three-loudspeaker stereo, as briefly noted by Meares in ref. [26].

Available microphones with high-quality broadband polar diagrams are all <u>first order</u> microphones, i.e. have polar diagrams that are combinations of pressure and velocity pick-up. All such polar diagrams at a point in space may be achieved by forming linear combinations of just four microphone pick-ups at that point, an omnidirectional (pressure) pick up and three orthogonal figure-of-eight (velocity) pick-ups pointing say forward, leftward and upward, as illustrated in fig. 11. These four signals, known as <u>B-format</u> (e.g. see [23], [27]), may be produced simultaneously by suitable microphone array systems, such as the commercially available AMS Soundfield microphone. If matrixing of these signals is to be performed to provide loudspeaker feed signals, it is highly desirable that the polar diagrams of these microphones should be consistent across a broad frequency range and that the microphones should be spacially coincident to within about 3 mm or so. This is not generally possible because of the physical size of low-noise microphone capsules, but the Sound Field microphone uses a special, tetrahedral arrangement of capsules with subsequent frequency-dependent matrix signal processing [23], [24] to achieve both broadband consistency and effective spacial coincidence of the output B-format signals.

For feeding loudspeakers in the horizontal plane, the vertical Z velocity signal of B-format is not used (except to achieve an upward or downward tilt of the overall microphone array [23]), so that all loudspeaker feed signals derived from a sound field microphone are linear combinations of the three horizontal signals W, X and Y of the respective omnidirectional, forward and leftward velocity pickups shown in fig. 11. However, as noted by Meares [26], if one simply derives three coincident directional microphone characteristics pointing forward to the left, front and right for the three loudspeaker feeds L_3, C_3 and R_3, the result has poor imaging quality.

It is typically found that either the degree of cross-talk of frontal sounds onto the two outer loudspeakers is large, giving very poor image stability for center-stage images, or else the stereo width of the image is very narrow, and that there seems to be no satisfactory trade-off between these problems. Indeed, without great care, it is found generally that the results are actually poorer than those obtained from a conventional 2-channel stereo coincident microphone technique fed to the three loudspeakers via the psychoacoustic 3×2 decoder of eqs. (4) and (5) and fig. 7.

Ideally, one would like the three microphone signals to respond to sound directions in front of a sound field microphone by giving gains following the optimum 3-loudspeaker panning law of fig. 3. Studies show that it is indeed possible to matrix the B-format signals from a sound field microphone to closely approximate the law of fig. 3 across about 95% of the stereo stage - but there is a snag. As reported in ref. [3], such a matrixing results in a microphone energy response that is actually 6 to 8 dB more sensitive to sounds arriving from the back of the microphone than from the frontal stage - and moreover, such back sounds are reproduced via three loudspeakers with an extremely bad imaging quality with severe destructive interference effects at the "ideal" listening position. For this reason, such a microphone technique, although it gives excellent imaging at the front, is practically almost useless.

In practice, one seeks a compromise, in which the matrixed signals from a sound field microphone do not have greater pick-up of sounds from the back than from the front, but where the results are nevertheless significantly better than for a 2-channel coincident microphone technique optimally matrixed for 3-loudspeaker reproduction. Such a technique, which requires the use of a special signal processing arrangement, has been described in ref. [6], but explaining it requires some more theory.

3.2 MST Matrixing

The theory required is that of the 3-loudspeaker counterpart to the theory of the MS matrix for 2-loudspeaker stereo, which we term the MST matrix, which was developed in ref. [28]. That paper was conceptually and mathematically difficult. While serious researchers are encouraged to study it and the companion paper [4], for most recording engineers, we simply summarize what we need here without fully explaining how it was arrived at.

The MS matrices of eqs. (1) and (2) earlier had three significant properties that make them useful for 2-channel stereo signal processing, namely:
(i) the MS matrix is its own inverse, i.e. the cascade of two MS matrices restores the original signals.
(ii) MS matrices are energy preserving, i.e. the total energy of the signals coming out equals the total energy of the signals going in, and
(iii) if left and right signals are interchanged before an MS matrix, the output signals are either unchanged (in the case of M) or are simply inverted in polarity (in the case of S).

The usefulness of processing signals in MS or "sum-and-difference" form after passing them into an MS matrix was realized by Blumlein as long ago as 1931 [29], and has been widely used since [30], [31], with MS processing having become almost the standard technique in stereophonic TV applications [31].

The MST matrixing is a three-channel counterpart, having the above three properties for 3-loudspeaker signals L, C and R (left. center and right). The matrix is given by the formulas

$$\begin{aligned} M &= 0.5000\ L + 0.7071\ C + 0.5000\ R \\ S &= 0.7071\ L - 0.7071\ R \\ T &= 0.5000\ L - 0.7071\ C + 0.5000\ R\ , \end{aligned} \tag{9}$$

and the inverse MST matrix is given by

$$\begin{aligned} L &= 0.5000\ M + 0.7071\ S + 0.5000\ T \\ C &= 0.7071\ M - 0.7071\ T \\ R &= 0.5000\ M - 0.7071\ S + 0.5000\ T\ . \end{aligned} \tag{10}$$

The three properties (i) to (iii) above still hold for the MST matrix, with the additional observation added to (iii) that T is also unchanged when left and right are interchanged.

Thus 3-loudspeaker stereo signals can be handled either in "LCR" form or in "MST" form, with identical conversion matrices between the two forms. It is convenient to think of 3-loudspeaker stereo coincident microphone techniques in terms of the polar diagrams of their MST signals, instead of their LCR signals, just as it is often helpful to think of two-channel stereo microphone characteristics in MS rather than LR form (see [25], [30]), knowing that one can be converted to the other by the matrixing (9) and (10).

3.3 Relationship to 3×2 Decoding

A valuable aspect of MST matrixing is that it provides another view of the psychoacoustic 3×2 decoders described above with reference to eqs. (3) to (5) and figs. 6 and 7, for 3-loudspeaker reproduction from 2-channel stereo. For a value of ϕ near 45°, it was shown in ref. [28] that the 3×2 decoder of fig. 6 and eq. (3) can be implemented as an MS matrix for the input 2-channel left and right signal, with M and S fed to an output MST matrix, where the M and S signals are simply those derived from the two input channels, and where the T signal equals M given the (small) gain $\tan(\phi - 45^{\circ})$, i.e.

$$T = (\tan(\phi - 45^{\circ}))M . \tag{11}$$

In particular, in the frequency-dependent decoder, where $\phi = 35.3^{\circ}$ well below 5 kHz and $\phi = 54.7^{\circ}$ above 5 kHz, the gain of M in the T channel is $-\tan 9.7^{\circ}$ below 5 kHz and $+\tan 9.7^{\circ}$ above 5 kHz, so that the T channel is effectively filtered by an all-pass network having gain -1 below 5 kHz and gain +1 above, followed by an extra gain of 0.172 = tan 9.7°. Thus the psychoacoustically optimized 3×2 decoder of fig. 7 and eqs. (4) and (5) can alternatively be implemented as shown in fig. 12.

For other fixed values of ϕ not differing too greatly from 45° (say between 25° and 65°), the 3×2 decoder of fig. 6 and eq. (3) can alternatively be implemented using an MS matrix and and MST matrix as shown in fig. 13; for $\phi = 45^{\circ}$, the T signal input becomes zero and may be omitted.

Thus 3-loudspeaker reproduction of 2-channel stereo essentially involves using the same M and S signals, with only a small T signal 15 dB or so below the M signal in level.

3.4 Relationship to Optimal Panning

In ref. [5], the values of M, S and T were calculated for the optimal 3-loudspeaker panpot law of fig. 3 (in table 7 of [5]), and it was shown that the gain of T was below -10 dB across 95% of the stereo stage (increasing to - 6 dB only at the extreme edge of the stage), having a gain of -0.276 at the center of the stereo stage, falling to 0 at about 71% of the way from the center to the edge of the stereo stage, and increasing to about + 0.296 at 95% of the way from center to edge.

Thus, it will be seen that the optimum 3×2 psychoacoustic decoder for 2-channel stereo attempts to more closely approximate the central imaging properties of the optimal panpot law below 5 kHz, and the imaging near the edges of the stage above 5 kHz.

3.5 B-format decoding.

A simple frequency-independent 3-loudspeaker stereo decoder for B-format signals based on the observations of subsections 3.3 and 3.4 above uses a forward-facing hypercardioid signal for M, a leftward-facing figure--of-eight for S, and a backward-facing hypercardioid for T, whose small negative-polarity forward-facing "rear" lobe provides the negative gain for the T-channel signals for nearly-due-front-sounds, and a small positive gain for sounds in the front stage further left or right.

More precisely, a uniform energy pick-up from all directions can be provided, and the frontal-stage directions can be made to follow a panning law that, for front-center sounds agrees with the results of the low-frequency 3×2 decoder, and for azimuths $\pm 60^{\circ}$ from due front agrees with the high-frequency 3×2 decoder for sounds at left or right. Thus, at all frequencies, this microphone array combines the imaging vitues of the low-frequency 3×2 decoder for parts of the sound stage for which it is best with the virtues of the high-frequency 3×2 decoder for parts of the stage at which it is best.

Denoting B-format signals by W for a signal with gain 1 in all directions, X for a signal with gain $2^{\frac{1}{2}}\cos\theta$ for sounds from an azimuth θ measured anticlockwise from due front, and Y for a signal with gain $2^{\frac{1}{2}}\sin\theta$, as illustrated in fig. 14, the M, S and T signals are derived by the matrix

$$M = 0.7071(W + X)$$

$$S = Y$$

$$T = 0.7071(W - X) \qquad (12)$$

which is an orthogonal matrix, and hence energy-preserving. The basic B-format decoder for 3 loudspeakers thus provided comprises a conversion matrix converting B-format to signals M, S, T as in eqs. (12), followed by an MST matrix to provide 3 loudspeaker feeds L_3, C_3 and R_3 as shown in fig. 15.

While this frequency-independent decoder for B-format is better than 3×2 decoding from a 2-channel stereo microphone, the panning law it produces across the frontal stage is still less good than the optimum law of fig. 3. Improving on it without increasing the gain of sounds arriving from the rear requires the use of frequency-dependent matrixing similar to that used in the optimum 3×2 decoder of fig. 12.

The T signal produced in fig. 14 or via eq. (12) has a very low gain across the frontal stage of azimuths within $\pm 60^{\circ}$ of due front, being about 15 dB or more below the M signal; despite this low gain, it produces a significant improvement in image quality. Comparing fig. 12, the optimum 3×2 decoder, with fig. 15 reveals that the two are similar, where fig. 15 uses an M signal that is a forward-facing hypercardioid with nulls 135° off-axis and an S signal that is a sideways figure-of-eight, but a T signal that is, unlike fig. 12, not derived from M.

A frequency-dependent improvement on fig. 15, making the sounds near azimuth $\theta = 0^{\circ}$ conform more closely to the panpot law of fig. 3 below 5 kHz and sounds near azimuths $\pm 60^{\circ}$ conform more closely to the panpot law of fig. 3 above 5 kHz, introduces a frequency-dependent rotation matrix in the M and T signal paths of fig. 15, as shown in fig. 16. The rotation matrix itself is an extension of the arrangement found in fig. 12, using first-order all-pass networks with gain -1 below 5 kHz and gain + 1 above 5 kHz, shown in fig. 17. The effect of the rotation matrix is that the decoder has the same form as the 3×2 decoder of fig. 12 if the T input to the rotation matrix in fig. 16 is suppressed, and restoring the T input has the effect of increasing the magnitude of T further at the center of the image below 5 kHz and at the edge of the image above 5 kHz, thereby improving the image quality further over that of the frequency-independent decoder of fig. 15.

The mathematical form of the rotation matrix is ideally

$$M_{out} = \cos(\phi-45^o)\, M_{in} - \sin(\phi-45^o)\, T_{in}$$

$$T_{out} = \sin(\phi-45^o)\, M_{in} + \cos(\phi-45^o)\, T_{in} \quad , \qquad (13)$$

and in the approximation of fig. 17, applicable when $\phi-45^o$ has small values (say of magnitude less than 15^o), is of the practical form

$$M_{out} = M_{in} - \tan(\phi-45^o)\, T_{in}$$

$$T_{out} = \tan(\phi-45^o)\, M_{in} + T_{in} \quad , \qquad (14)$$

obtained by dividing eqs.(13) by the factor $\cos(\phi-45^o)$ which approximately equals 1.

Since all the matrices in fig. 16 are orthogonal (and hence preserve energy), the output signals fed to the loudspeakers still have a uniform energy response in all horizontal directions, so that the frequency--dependence does not affect flatness of total frequency response into the room from the loudspeakers.

3.6 Properties of B-format decoder

The microphone technique obtained by feeding B-format from a sound field microphone into the 3-loudspeaker decoders of fig. 15 (frequency--independent case) or fig. 16 (frequency-dependent case with improved imaging) has been analyzed using methods similar to those of Julstrom [25]. The microphone technique has an omnidirectional energy response in the horizontal plane, although its gain falls for sounds from non-horizontal directions, with a gain of -0.79 dB for elevation angles of $\pm$ 30°, -1.76 dB for $\pm$ 45° and -4.77 dB for $\pm$ 90° (i.e. above or below the microphone). Such near-omnidirectionality helps to give this microphone technique a good portrayal of sound-source distance, since it preserves both the relative time and (within 1 dB for elevation angles between -33.7° and +33.7°) relative gains of early reflections, which are now known to be the main cue for distance perception (see ref. [32]).

The calculated directivity factor of this microphone technique is 9/7 = 1.286, because the velocity components of the pick-up reject energy arriving from above and below, giving a distance factor (see Julstrom [25] section 3.2.1 for the 2-channel case) equal to 1.134, which is slightly worse than common 2-channel coincident microphone techniques.

The effective angular coverage of these 3-loudspeaker microphone techiques of figs. 15 or 16 for sounds within the stereo stage (which may be defined as azimuths for which

$$|L_3 - R_3| \leq |L_3 + C_3 + R_3| \qquad) \tag{14}$$

is 141.06°, i.e. azimuths between -70.53° and +70.53°, which is comparable to the angular coverage of popular 2-channel coincident stereo microphone techniques. Outside these angular limits, sounds will appear to come from beyond the outer loudspeakers at low frequencies, becoming unlocalizable for sounds originating from rear azimuth arrival angles.

It will be noted from the polar diagrams shown in fig. 15 that the frequency-independent B-format 3-loudspeaker decoder has left and right loudspeaker signals that are sideways facing hypercardioids with nulls 135° off-axis, and the center loudspeaker feed is a forward figure of eight. This means that the energy of pick-up in the three loudspeakers is identical for front-stage and rear-stage sounds arriving at the microphone, with the polarity of the center loudspeaker inverted for rear stage sounds.

The frequency-dependent decoder has an M polar diagram below 5 kHz that is a forward-facing cardioid, whereas the M polar diagram above 5 kHz becomes a hypercardiod with nulls about 120° off-axis. Conversely, the T polar diagrams are hypercardiod below 5 kHz and cardioid above.

The frequency-dependent 3-loudspeaker signals result in frequency-dependent polar diagrams for the mono and 2-channel stereo fold-down of the 3-speaker signals [28]. However, this frequency-dependence is not serious, resulting in a variation in frequency response of about 0.51 dB at center-front, no variation for sounds arriving from azimuths ±45°, and variations of under 1 dB in frequency response across the frontal stage.

3.7 Variable B-format techniques

The microphone technique of figs. 15 or 16 from B-format from a sound field microphone has the disadvantage that it is inflexible in use, having a fixed directional pick-up pattern and reproduced width. In practical use, one wishes to have the option of adjusting both the stereo stage width and the relative sensitivity of rear pick-up so as to provide the option of reducing unwanted noises from the rear of the microphone.

Width adjustment in the 3-channel case is not simply a matter of altering the gain of the S signal, since this also affects the panning law and generally degrades imaging quality near the edges of the stage. Similarly, varying the M polar diagram risks degrading imaging quality in some parts of the reproduced stereo stage. The alterations must be done in a carefully controlled way in order to avoid or minimize degradation of stereo imaging quality. The method of doing this is shown in figure 18.

Figure 18 involves two adjustments, "forward dominance" acting on the B-format input signals, and adjustment of the gain of the T channel before the rotation matrix.

The T-channel gain adjustment is the easiest to understand. If the T-channel gain is reduced to zero, then the decoder simply becomes the psychoacoustically-optimized 3×2 decoder of fig. 12 for that two-channel stereo microphone technique whose M-signal is a forward-facing hypercardioid (with nulls 135° off-axis) and whose S signal is a leftwards facing figure-of-eight. The reduction of the gain of the T-channel hardly affects the energy gain of the front-stage pick-up, since the T-signal has very little energy-gain across the frontal stage, and it also hardly affects the portrayed stereo width of the frontal stage for a similar reason.

Thus the main effect of fading out the T-signal is two-fold:
(i) the image localization quality of frontal-stage sounds is degraded somewhat to that given by a 3×2 decoder, and
(ii) the rear-stage pick-up is reduced.

Thus intermediate values of the T-channel agin between 0 and 1 have the effect of reducing rear-stage pick-up (roughly proportionately to the gain of the T-channel), with an intermediate localization quality that is best if the T-gain is near 1.

Forward dominance is a transformation of B-format signals that has the effect of converting a B-format encoded sound field into another B-format sound field whose encoded azimuths are altered and whose encoded gains are also altered, in a manner depending on the original azimuth. Thus preceding a decoder by a forward-dominance adjustment of the B-format sound field before 3-loudspeaker decoding does not affect the imaging <u>quality</u> of the decoded soundfield, but only the apparent <u>positions</u> of sounds within the sound stage and their relative gains.

Full details about the forward-dominance transformation are given in refs. [27], [33] and [23], which we only summarize here. The forward dominance transformation is given by

$$
\begin{aligned}
W' &= \tfrac{1}{2}(\lambda + \lambda^{-1})\,W + 8^{-\frac{1}{2}}(\lambda - \lambda^{-1})\,X \\
X' &= 2^{-\frac{1}{2}}(\lambda - \lambda^{-1})\,W + \tfrac{1}{2}(\lambda + \lambda^{-1})\,X \\
Y' &= Y
\end{aligned}
\tag{15}
$$

for a positive parameter λ, which equals 1 for no change. The gain of sounds at the front is increased by a factor λ, whereas those at the back are multiplied by a factor $1/\lambda$, causing a relative gain of the back sounds (realtive to the front) of $1/\lambda^2$. For λ greater than 1, the encoded azimuth of all sounds is shifted towards the front, as illustrated in fig. 19 for the case $\lambda = \sqrt{2}$, and for λ less than, the encoded azimuth is is moved towards the back in a similar way.

Thus using forward dominance with $\lambda > 1$ both reduces rear pick-up and gives a narrower reproduced stereo stage. For example, with $\lambda = \sqrt{2}$, the rear pick-up is reduced by 6 dB relative to frontal stage pick-up, but the angular stage width needed to fill the stereo stage is increased from 141° to a very wide 180°. The case $\lambda < 1$ has a converse effect. For example, $\lambda = 0.7071$ increases rear pick (relative to front) by 6 dB, which is generally unacceptable unless the T gain is reduced by at least 6 dB in the decoder of fig. 18, but the angular coverage (i.e. the sound source stage angle required to fill the reproduced stereo stage) is reduced from 141^{o} to 106^{o}, which results in a wider reproduced stereo.

Thus the B-format to 3-loudspeaker decoder of fig. 18 provides a flexible range of adjustments that permit the B-format output of a sound field microphone to be reproduced via 3 loudspeakers with good imaging quality and adjustable width and back pick-up gain.

Attempting to achieve variable decoding simply by directly deriving microphone signals pointing in different directions from B-format results in it being very difficult to adjust the resulting parameters to ensure good localization across the stereo stage, and is practically impossible with the pressures of actual recording. The use of B-format to provide 3-loudspeaker coincident microphone technique is only feasible if a dedicated processor of the form shown in fig. 18 is used, where the required adjustments are provided without undesirable effects on localization quality.

More details of this technique are provided in ref. [33], and it is the subject of patent applications.

4. OTHER TECHNIQUES

4.1 Near-Coincident Techniques

The above work has shown that in general, if one wishes to take advantage of the improved imaging performance of 3-loudspeaker stereo, implementing coincident microphone techniques is difficult, involving the use of dedicated signal processing. The reason is that a "randomly" selected technique has poor performance, and it is difficult to locate those relatively few techniques that work well among them unless the equipment is designed only to produce techniques with good localization quality.

This difficulty with coincident techniques also makes it difficult to optimize those 3-channel microphone techniques using directional microphone capsules spaced relatively closely together, say with spacings of between 5 and 30 cm. In 2-channel stereo recording, such techniques have proved popular (e.g. see various papers in ref. [1]), and it is known that such small spacings can actually improve stereo image quality for listeners in the ideal stereo seat through simulating interaural phase and amplitude differences across a broader range than coincident microphone techniques [34], [22].

However, any three-channel analogues of these techniques must use a strictly limited subset of possible polar characteristics in order that the 3-loudspeaker imaging quality be optimized, and finding the optimum spacing and polar diagram characteristics involves studying a much larger number of parameters than in the 2-channel case. Such studies have not yet been undertaken. It seems a reasonable assumption that the polar diagrams of the three microphones will be similar (although probably not quite identical) to those produced by the methods of figs. 15 to 18 in the coincident case, and this immediately suggests further problems.

For example, the polar diagrams produced by the method of fig 15 include two hypercardioids with nulls 135° off-axis, so that two of the microphones need to have these polar diagrams, with low colouration and reasonably accurate polar diagrams even at the sides of their response. we are unaware of any microphone having the required characteristics currently on the market, so that the only way currently of implementing such a nearly-coincident microphone technique is to use a pair of soundfield microphones, one each to provide a suitable mono feed signal for the left and right loudspeakers, plus a third forward-facing mono microphone for the center loudspeaker. Using a sound field microphone to derive a mono speaker feed is not only a very expensive way of doing things, but the physical size of a sound field microphone makes such a closely-spaced array liable to acoustic obtruction effects.

The imperfections of the side response of the microphones used to implement the left and right pick-ups of a closely-spaced 3-channel microphone technique make this approach somewhat problematical at present, and in particular, there is an urgent need, for a number of applications other than the one here, for a microphone with an accurate broadband polar diagram that is a hypercardioid with nulls 135° off-axis. Currently, the only characteristics available with accurate broadband polar diagrams are omnidirectional, subcardioid, cardioid, hypercardioid with nulls 120° off-axis and figure-of-eight.

If suitable microphones become available, one can experiment not only with the left-right spacing, but also with the spacing of the center microphone in front of the line joining the left and right microphones, and with the use of a short time delay (probably under 1 ms) in the center or in the two outer microphone feeds.

Studies need to be done comparable to those of ref. [34] as to what spacings and time delays provide the most accurate recreation at the ears of an ideally-situated listener of natural interaural phase and amplitudes encountered with natural sound sources. This requires further work based either on interaural phase and amplitude data measured on actual heads, or based on solid-sphere theoretical models of the head as used in [35], [36].

4.2 Hybrid techniques

As in the 2-channel case, there is scope for using a wide variety of hybrids of known techniques, and we do not feel the need to go into these in much detail, as the use of such hybrids is a matter of the taste and temperament of the recording engineer. Such hybrids basically involve the mixing together of the speaker feed signals produced by two (or more) separate 3-loudspeaker microphone techniques.

One popular hybrid technique in 2-channel use is the mixing of a main coincident stereo pair with an "outrider" widely spaced pair of omnidirectional microphones, where usually, the spaced pair is mixed at a lower level, between -10 and -20 dB down, to provide a sense of "width" or "space" while still leaving the main imaging to the coincident pair. Analogous methods exist for 3-channel stereo, where the center stereo microphone may be a 2-channel microphone fed into a psychoacoustic 3×2 matrix or a sound field microphone fed to a matrix such as that of fig. 18, and where the spaced pair may be fed directly to the outer two loudspeakers, since their function is not to provide direct sound imaging in this application. Alternatively, the spaced pair may be fed into a matrix such as that of fig. 6 with ϕ between 55^{o} and 90^{o}, or to a matrix such as

$$L_3 = L - 0.1R$$

$$C_3 = 0.2(L+R)$$

$$R_3 = -0.1L + R \,, \qquad (16)$$

where L and R are the left and right spaced microphone signals, which has the effect of subjectively panning their sounds about 95% of the way to the two sides of the stereo stage (according to the panning law of fig. 3) while helping to "pin down" their output in the center loudspeaker.

In general, with hybrid techniques where the main burden of stereo imaging is performed by one component of the mix, it is acceptable to matrix another component of the mix with suboptimal stereo imaging, choosing the matrix rather to enhance its other qualities such as the "spaciousness" of the spaced pair in the above case of outrider microphones.

In general, as in the case of 2-channel stereo, there may be a conflict between goodness of stereo imaging and an overall sense of "spaciousness", and tradeoffs between the two may be made. However, in the 3-channel case, this tradeoff should not be such as to make the stereo localization quality worse than for the 2-channel case, since otherwise there is no point in using the third channel! In general, unless there is a specific requirement for image delocalization [37], one may seek to aim for a direct-sound imaging quality and stability approaching that of the optimum 3-channel panpot law of fig. 3 for at least the main direct sound sources.

4.3 Baffling techniques

Another class of techniques used in 2-channel stereo is the use of microphones placed on a relatively small baffle, such as a dummy head, a solid sphere [38] or to either side of a plane baffle [29]. with these techniques, the acoustic obstruction provided by the baffle helps provide left/right separation at higher frequencies, and Blumlein Shuffling [29], [38] may be used to provide separation at lower frequencies.

Similar baffling techniques may be used in the 3-channel case to increase the poor separation between nearly-coincident microphones having conventional polar diagrams discussed above. It is not, however, at all obvious what the best form of baffling is. One possibility might be to use a hard sphere about 20 cm diameter as used by Theile [39], but to use three embedded omnidirectional capsules on its surface rather than two, with capsules placed at azimuths +120°, 0° (due front) and -120° for left, center and right loudspeaker feeds. A 3-channel analogue of Blumlein shuffling can cause this technique to merge at low frequencies (below a few hundred Hz) into that of fig. 15. As in the case of other possible 3-channel microphone techniques, this involves the use of dedicated signal processing.

5. CONCLUSIONS

All but the most naive microphone techniques for 3-loudspeaker stereo appear to need the design of dedicated matrix signal processing, often of a frequency-dependent character. This is partly because the number of available parameters in the choice of a microphone technique is so large that there is a much higher risk in the 3-channel case of "missing" the good techniques in a wealth of bad ones.

Very little dedicated equipment for this application is currently available, and as a result, nearly all 3-channel recordings heard by the author fail to have the quality of stereo imaging that he knows is possible. Often the results are very crude, largely due to the lack of appropriate production tools, which are often 2-channel tools poorly adapted to 3-channel use as noted by Meares [26]. This crudity of effect was less important in large auditorium environments where the large seating area and the large acoustic combined to mask such problems, but in the domestic environment, listeners are already used to hearing subtly-composed stereo stages even for 2-channel material.

A typical 3-channel recording today will tend to have sounds panned by a pairwise panpot [5], i.e panned with a sine/cosine law between an adjacent pair of loudpeaker feeds, mixed with sounds from one or more 2-channel stereo microphones, the left and right channels of each of which will again be panned by a pairwise panpot [26]. Such a feed of stereo microphone signals into 3 channels is markedly suboptimal, and is generally bettered by the use of dedicated 3×2 matrices.

Although this paper has outlined some of the matrix signal processing that may help to make 3-channel microphone techniques less crude, understanding their use involves much more theory than for the corresponding 2-channel methods. In order to get the best possible results, it is desirable both that appropriate matrixing equipment become available and that recording engineers become much more technically informed about the design of microphone techniques at a purely technical engineering level. The extra degrees of freedom of 3-channel microphone technique have a lot more potential for going wrong than in the 2-channel case.

In particular, this paper has outlined the requirements for improved image quality, summarized by the panpot law of fig. 3, the use of 3×2 matrices for feeding 2-channel microphone techniques into 3 loudspeakers, the use of a 3-channel analogue, the MST matrix, to 2-channel MS matrixing methods, and the design of decoding matrices to feed B-format signals into 3 loudspeakers, with a B-format equivalent of width control and of front/back directivity control.

As this paper has noted, there is a lot of room for experimentation with microphone techniques for three loudspeakers, but that there are serious problems in developing these techniques if done without a good understanding of the unique problems of the 3-loudspeaker case, and good signal-matrixing tools are required.

This paper has also emphasized the importance of adopting standardized monitoring loudspeaker arrangements for 3-loudspeaker stereo, using arrangements that ensure exactly the same times of arrival at an ideally situated listener from all three loudspeakers, using delay compensation as in fig. 2 if necessary. Without standardized arrangements, there will be no basis for agreeing on the effects of different recording techniques. With digital recording and transmission media, time delays are now cheap, and any relative time delay used between the three loudspeakers should be a conscious recording decision, and not an accident of particular monitoring arrangements.

The emphasis in this paper on the engineering design aspects of 3-channel microphone techniques is not intended to over-ride artistic judgements, but simply to provide recording tools that allow those artistic judgements to be made without the need to struggle with purely technical aspects. An infinite degree of freedom in implementing microphone techniques means an almost zero chance of happening to find the ones that perform particularly well, because of the large number of free parameters.

While this paper has presented many new microphone techniques, notably the 3-loudspeaker coincident technique of figs. 15 to 18 and the spaced stereo microphone technique of figs. 9 and 10 (with eqs. (7) or (8)), it has also pointed to many unsolved problems, requiring solid engineering design work, in implementing other classes of microphone techniques such as baffled microphones or techniques using a small spacing. We would encourage technically literate recording engineers and equipment manufacturers to develop technical solutions to these problems, so that the full potential of 3-loudspeaker stereo, with improved imaging quality and stability across a large listening area, can be achieved with a range of recording approaches.

6. PATENT NOTE

A number of the signal processing algorithms and methods described in this paper are the subject of patent applications by the author assigned to Trifield Productions Ltd.

7. ACKNOWLEDGEMENTS

I would like to thank David Meares and David Kirby of the BBC for discussions in which they gave me many insights into practical operational problems, and to Geoffrey Barton for discussions and support.

8. REFERENCES

[1] ed. J. Eargle, Stereophonic Techniques, Audio Engineering Society, New York, 1986

[2] J.C. Steinberg & W.B. Snow, "Auditory Perspective - Physical Factors," Electrical Engineering, vol. 53 no. 1, pp. 12-15 (1934 Jan.) reprinted in ref. [1].

[3] M. Gerzon, "Three Channels. The Future of Stereo?," Studio Sound, vol. 32 no. 6, pp.112-125 (1990 June)

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

[5] M.A. Gerzon, "Panpot Laws for Multispeaker Stereo," preprint 3309 of the Audio Engineering Society 92nd Convention, Vienna (1992 Mar. 24-27)

[6] M.A. Gerzon, "Hierarchical System of Surround Sound Transmission for HDTV," preprint 3339 of the Audio Engineering Society 92nd Convention, Vienna (1992 Mar. 24-27), especially section 5.

[7] H.A.M. Clark, G.F. Dutton & P.B. Vanderlyn, "The 'Stereosonic' Recording and Reproducing System," J. Audio eng. Soc., vol. 6 no. 2, pp. 102-117 (1958 April); reprinted in ref. [1].

[8] H.D. Harwood, "Stereophonic Image Sharpness," Wireless World, vol. 74, pp. 207-211 (1968)

[9] J.S. Bower, "The Subjective Effects of Interchannel Phase-Shifts on the Stereophonic Image Localisation of Narrowband Audio Signals," BBC Research Department Report BBCRD 1975/28 (1975 Sept.)

[10] D. Griesinger, "Spaciousness and Localization in Listening Rooms and their Effects on the Recording Technique," J. Audio Eng. Soc., vol. 34 no. 4, pp. 255-268 (1986 Apr.)

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

[12] G. Theile, "On the Performance of Two-Channel and Multi-Channel Stereophony," preprint 2887 of the Audio Engineering Society 80th Convention, Montreux, (1990 Mar. 13-16)

[13] S. Komiyama, "Subjective Evaluation of Angular Displacement Between Picture and Sound Directions for HDTV Sound Systems," J. Audio Eng. Soc., vol. 37 no. 4, pp. 210-214 (1989 Apr.)

[14] W.D. Snow, "Basic Principles of Stereophonic Sound," J. SMPTE, vol. 61, pp. 567-589 (1953 Nov.); reprinted in ref. [1].

[15] R. Orban, "Notes on Panpots," J. Audio Eng. Soc., vol. 19 no. 11, pp. 936-939 (1971 Nov.)

[16] H. Haas, "The Influence of a Single Echo on the Audibility of Speech," J. Audio Eng. Soc., vol. 20 no. 2, pp. 145-159 (1972 Mar.)

[17] H. Wallach, E.B. Newman & M.R. Rosenzweig, "The Precedence Effect in Sound Localization," Am. J. Psychol., vol. 52, pp. 315-336 (1949); reprinted in J. Audio Eng. Soc., vol. 21, pp. 817-826 (1973 Dec.)

[18] F.E. Toole, "Loudspeakers and Rooms for Stereophonic Sound Reproduction," The Sound of Audio, Proc. AES 8th Int. Conf., Washington DC (1990 May 3-6), pp. 71- 91; especially section 2.4.

[19] P.W. Klipsch, "Stereophonic Sound with Two Tracks, Three Channels by Means of a Phantom Circuit (2PH3)," J. Audio Eng. Soc., vol. 6 no. 2, pp. 118-123 (1958 Apr.); reprinted in ref. [1].

[20] P.W. Klipsch, "Stereo Geometry Tests," IRE Trans. Audio, vol. AU-10, pp. 174-176 (1962 Nov./Dec.); reprinted in ref. [1].

[21] P.W. Klipsch, "Signal Mutuality in Stereo Systems," IRE Trans. Audio, vol. AU-8 no. 5, pp. 168-173 (1960 Sept./Oct.); reprinted in ref. [1].

[22] M.A. Gerzon, "Stereo Recording of Live Amplified Music," Proceedings of "Reproduced Sound 3", Institute of Acoustics Conference, Windmere, England (1987 Nov. 5-8), pp. 69-87.

[23] K. Farrar, "Soundfield Microphone," Wireless World, vol. 85 pp. 48-50 (1979 Oct.) and pp. 99-103 (1979 Nov.)

[24] M.A. Gerzon, "The Design of Precisely Coincident Microphone Arrays for Stereo and Surround Sound," Proc. of the Audio Engineering Society 50th Convention, London (1975 March)

[25] S. Julstrom, "An Intuitive View of Coincident Stereo Microphones," J. Audio Eng. Soc., vol. 39 no. 9, pp. 632-649 (1991 Sept.)

[26] D.J. Meares, "High Definition Sound for High Definition Television," Proc. AES 9th International Conference "Television Sound Today and Tomorrow," Detroit Michigan (1991 Feb. 1-2), pp. 187-215; especially p. 192 last paragraph.

[27] M.A. Gerzon and G.J. Barton, "Ambisonic Decoders for HDTV," preprint 3345, Audio Engineering Society 92nd Convention, Vienna (1992 Mar. 24-27), especially section 2.

[28] M.A. Gerzon, "Hierarchical Transmission System for Multispeaker Stereo," preprint 3199 of the Audio Engineering Society 91st Convention, New York (1991 Oct. 4-8); also J. Audio Eng. Soc., vol. 40, to be published probably 1992 Sept.

[29] A.D. Blumlein, British Patent 394325 (filed 1931 Dec.14); the main part of this is reprinted in ref. [1].

[30] W.L. Dooley & R.D. Streicher, "M-S Stereo: A Powerful Technique for Working in Stereo,", J. Audio Eng. Soc., vol. 30 no. 10, pp. 707-718 (1982 Oct.); reprinted in ref. [1].

[31] H. Evers, "The Use of MS Technique and MS Signals in Swedish Television," Proc. AES 9th International Conference "Television Sound Today and Tomorrow," Detroit Michigan (1991 Feb. 1-2) pp. 127-132.

[32] M.A. Gerzon, "The Design of Distance Panpots," preprint 3308 of the Audio Engineering Society 92nd Convention, Vienna (1992 Mar. 24-27)

[33] M.A. Gerzon, "Hierarchical System of Surround Sound Transmission for HDTV," preprint 3339 of the Audio Engineering Society 92nd Convention, Vienna (1992 Mar. 24-27)

[34] L.T. Olson, "The Stereo-180 Microphone System," J. Audio Eng. Soc., vol. 27 no. 3, pp. 158-163 (1979 March); reprinted in ref. [1].

[35] D.H. Cooper & J.L. Bauck, "Prospects for Transaural Recording," J. Audio Eng. Soc., vol. 37 no. 1/2, pp. 3 - 19 (1989 Jan./Feb.)

[36] Firestone,

J. Acoust. Soc. Am., (1929)

[37] M.A. Gerzon, "Signal Processing for Simulating Realistic Stereo Images," preprint presented at the Audio Engineering Society 93rd Convention, San Francisco (1992 Oct. 1-4)

[38] M.A. Gerzon, "Applications of Blumlein Shuffling to Stereo Microphone Techniques," preprint presented at the Audio Engineering Society 93rd Convention, San Francisco (1992 Oct. 1-4)

[39] G. Theile, "On the Naturalness of Two-Channel Stereo Sound," Proc. AES 9th International Conference "Television Sound Today and Tomorrow," Detroit, Michigan (1991 Feb. 1-2), pp. 143-149.

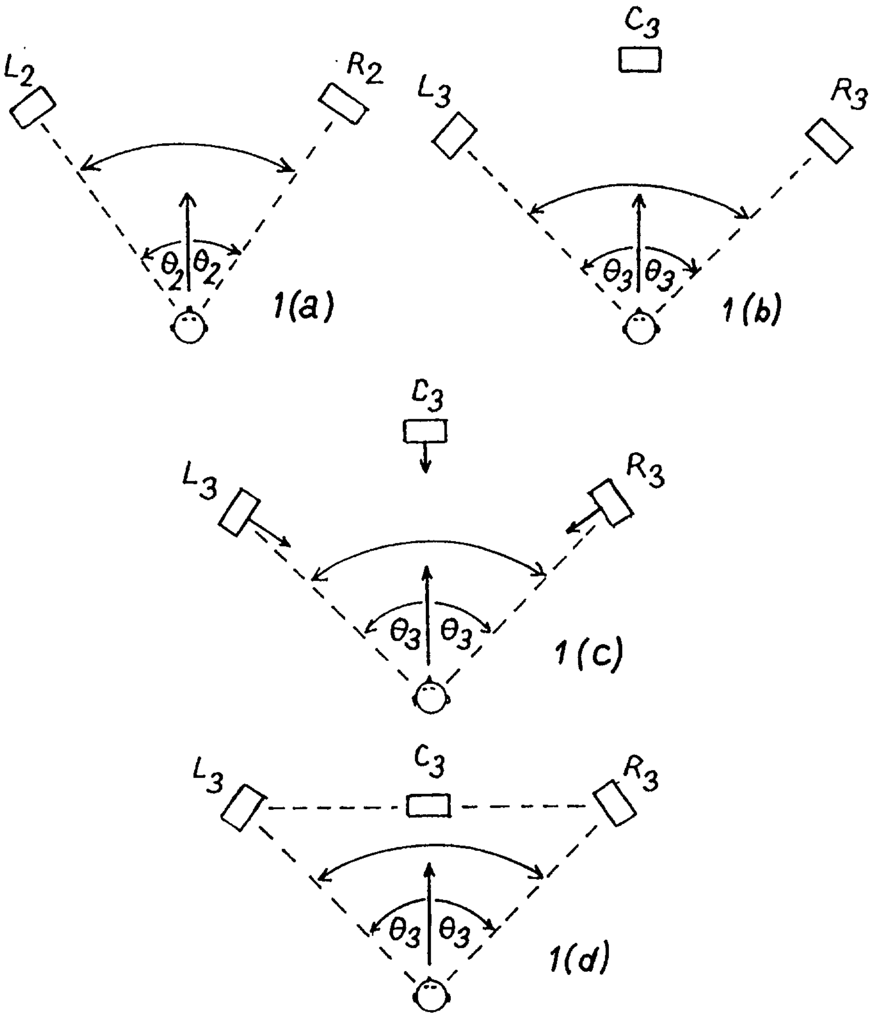

Figure 1. Two- and three-loudspeaker stereo layouts. 1(a) two loudspeakers. 1(b)-1(d) three loudspeakers. 1(a)-1(c) loudspeakers equidistant from ideally situated listener. 1(c) & 1(d) toed-in outer loudspeakers.

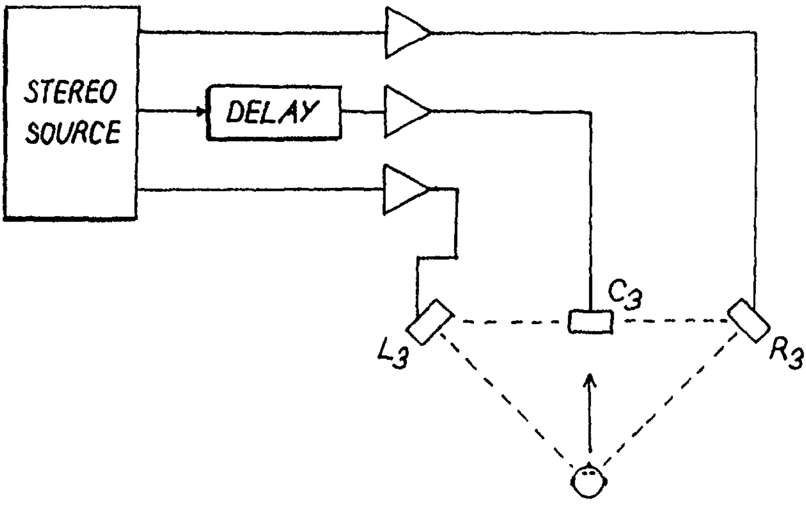

Figure 2. Use of time delay to compensate for unequal distances of loudspeakers from an ideally-situated listener.

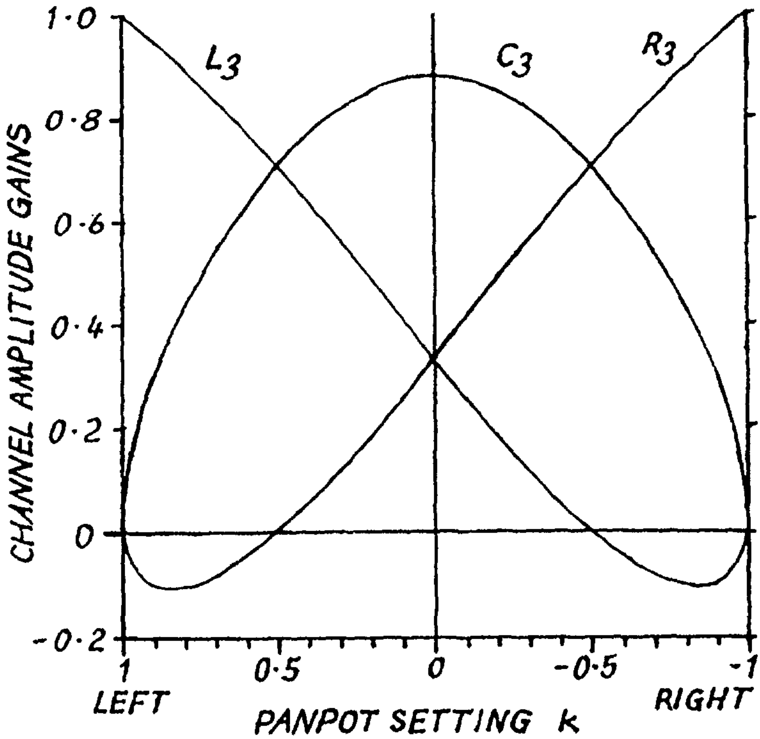

Figure 3. "Optimum" 3-channel panpot law gains for 3-loudspeaker stereo feeds L_3, C_3 and R_3, ensuring consistency of different localization cues and minimizing geometric image distortions for noncentral listeners.

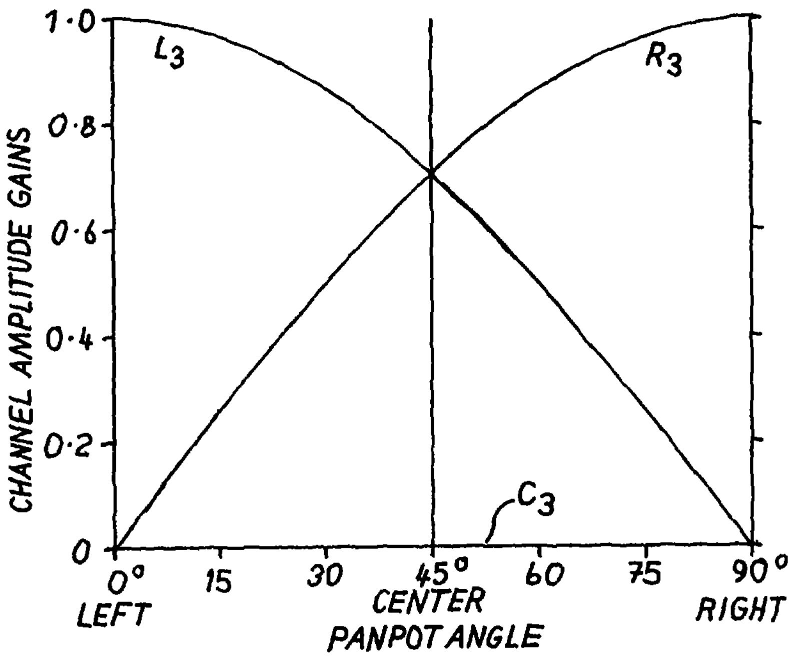

Figure 4. Channel gains for conventional sine/cosine constant-power sine-cosine panpot law for two-loudspeaker stereo, fed to three loudspeakers; the center loudspeaker C_3 gain is zero.

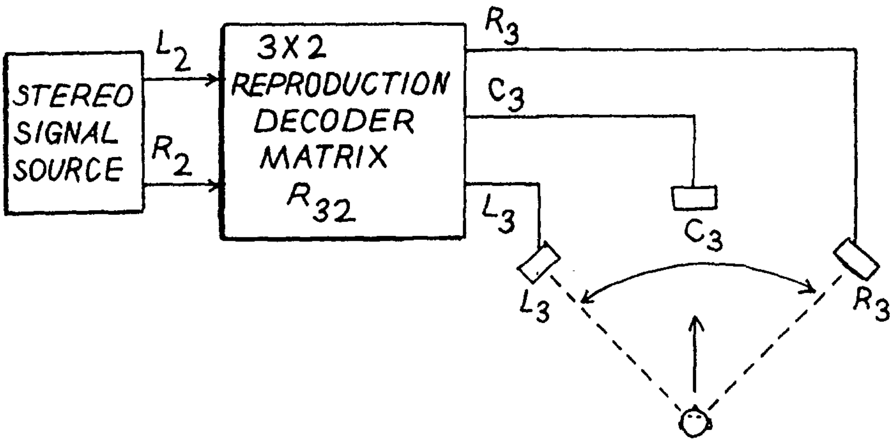

Figure 5. The reproduction of a 2-channel stero source via three loudspeakers, using a 3×2 psychoacoustic decoder.

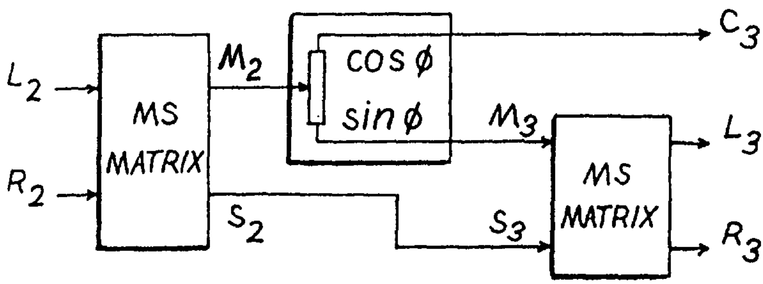

Figure 6. Form of 3×2 decoder for three-loudspeaker reproduction of two-channel stereo using sine and cosine gains dependent on an angle parameter ϕ.

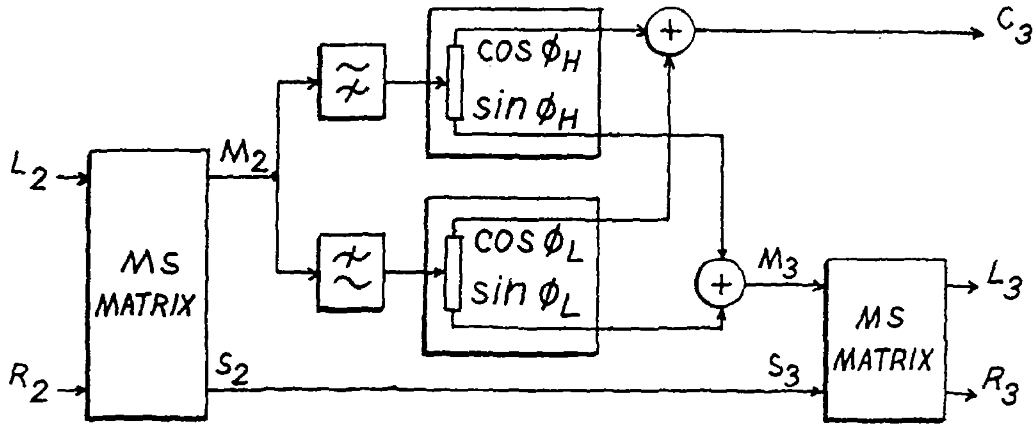

Figure 7. Frequency-dependent version of the 3×2 decoder of fig. 6, using low and high pass filters preceding different values, ϕ_L and ϕ_H, of the decoder parameter ϕ at respective lower and higher frequencies around 5 kHz. Recommended values are $\phi_L = 35.26^\circ$ and $\phi_H = 54.74^\circ$.

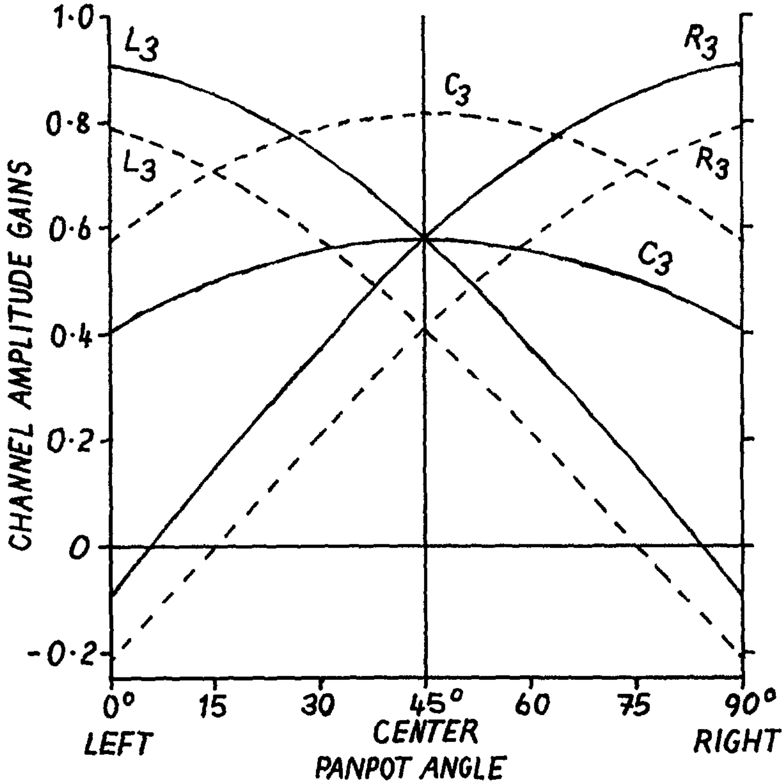

Figure 8. Channel gains for three loudspeakers L_3, C_3, R_3 when a constant-power sine-cosine two-channel panpot is fed to a psychoacoustically-optimized 3×2 decoder. Chained lines: low frequencies below 5 kHz, with $\phi = 35.26°$; solid lines: high frequencies above 5 kHz, with $\phi = 54.74°$.

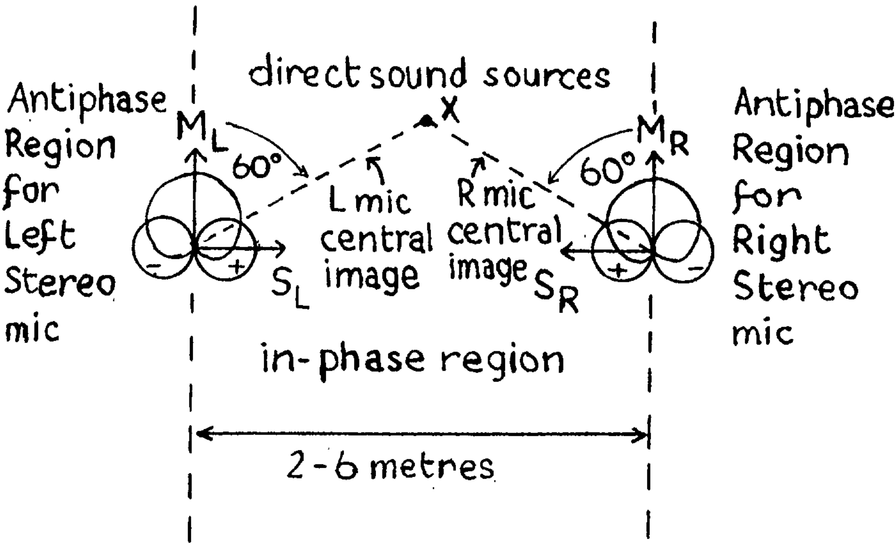

Figure 9. The use of a spaced pair of one-point stereo microphones, each of which is assymmetric, to cover a broad direct-sound-source stage. The gains of the S_L and S_R figure-of-eight microphones should ideally be such that a central sound source X in the stage is picked up as a central image, with equal channel gains, by both stereo microphones. Adapted from ref. [22].

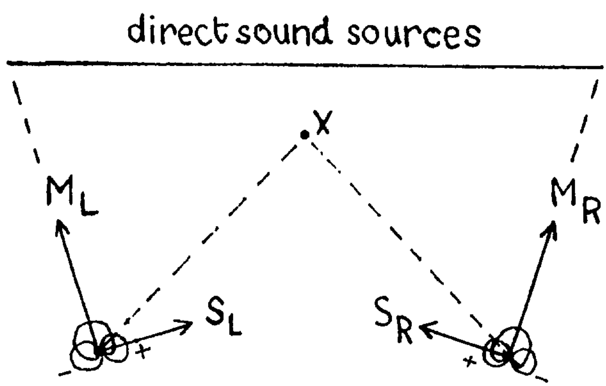

Figure 10. Illustrating the "splaying out" of the two stereo microphones, with M_L and M_R pointing at the edges of the stage, for a sound stage wider than the microphone spacing.

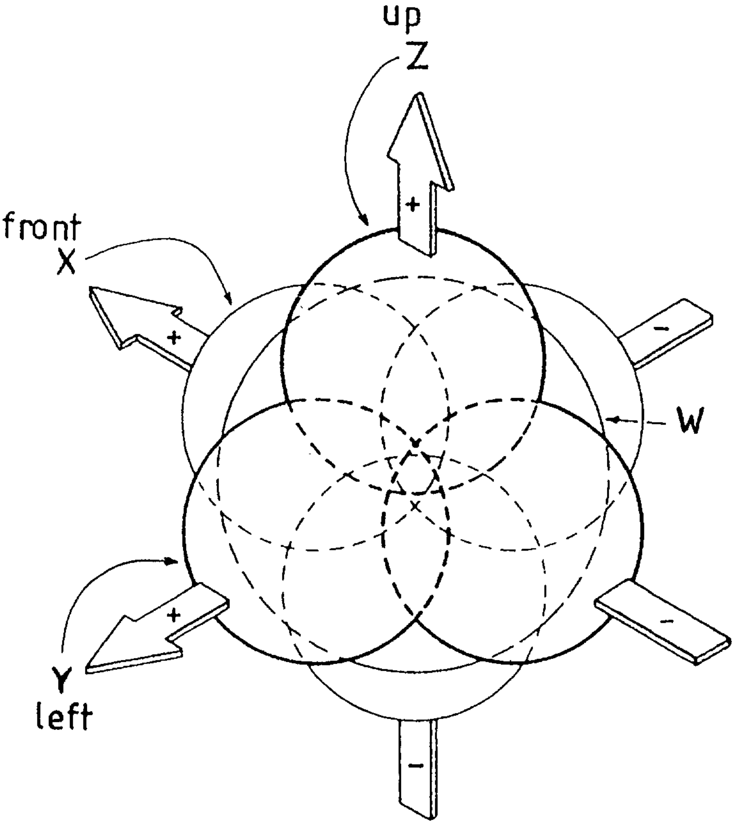

Figure 11. Polar diagrams of the B-format signals W (omnidirectional) and X, Y, Z (figure-of-eight with peak gain $\sqrt{2}$ = 1.4142) in 3 dimensions.

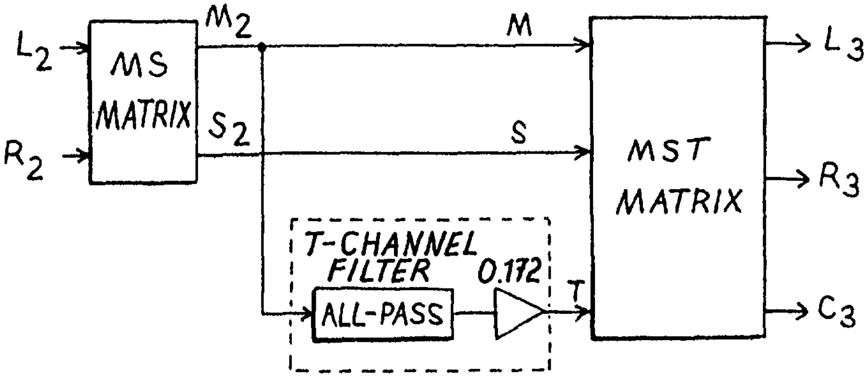

Figure 12. Psychoacoustically optimized 3×2 decoder of fig. 7 implemented as an MST decoder with a synthetic T channel produced by filtering the M signal; adapted from fig. 14 of ref. [28].

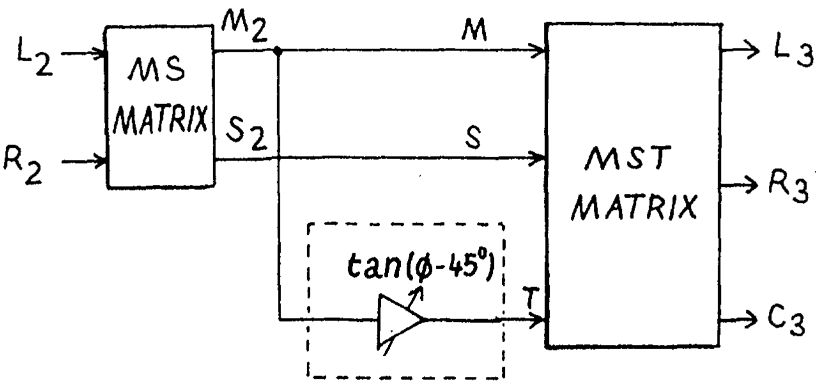

Figure 13. Version of the 3×2 decoder of fig. 6 implemented in MST form using a synthetic T channel; this is only an approximation to the results of fig. 6, but the approximation is very good for $|\phi - 45^\circ| \leqslant 15^\circ$

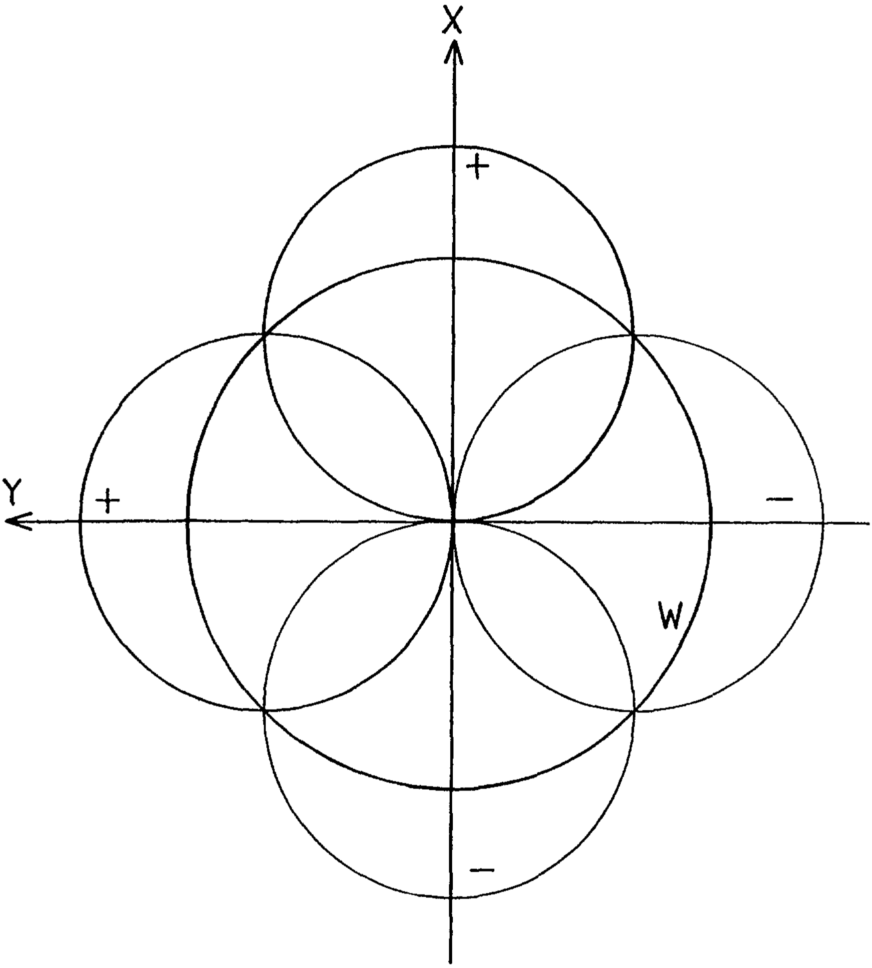

Figure 14. B-format directional polar diagrams in the horizontal plane as a function of azimuth angle.

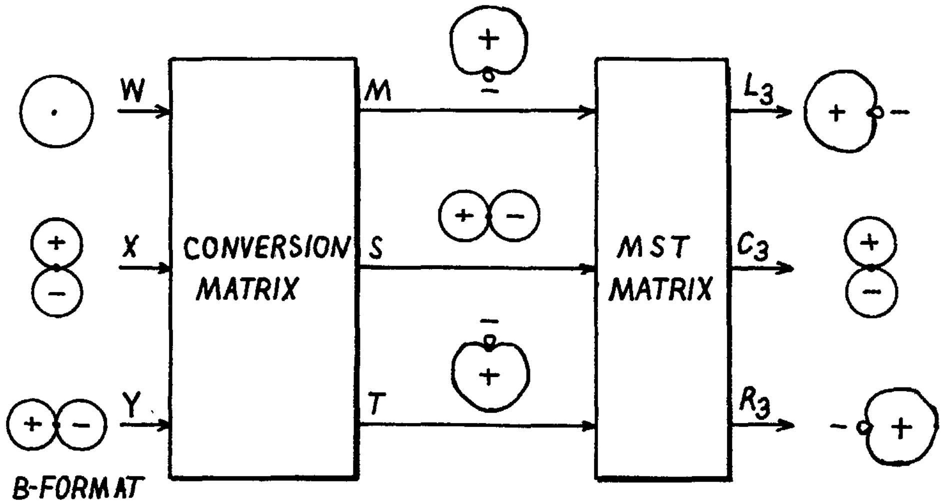

Figure 15. 3-Loudspeaker decoder for B-format input, using a conversion matrix (as in eqs. (12)) from B-format to MST signals and an MST matrix to produce loudspeaker feed signals, showing the polar diagrams of the signals.

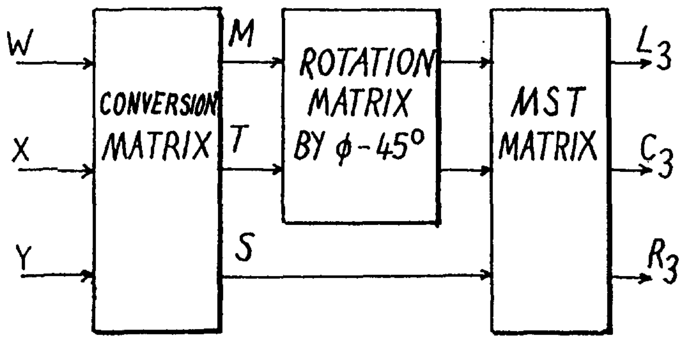

Figure 16. Use of frequency-dependent rotation matrix by angle ϕ-45^{o} in the M and T signal paths in the decoder of fig. 15. Typically, ϕ = 35^{o} below 5 kHz and 55^{o} above 5 kHz.

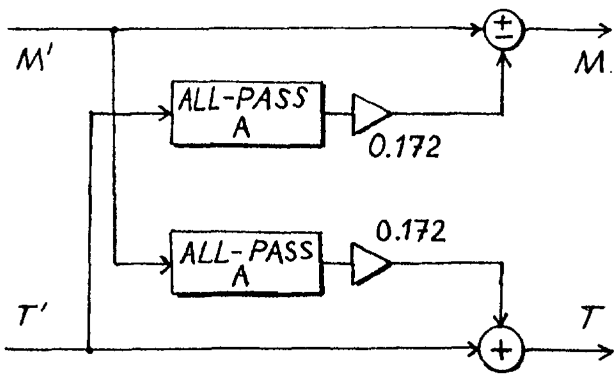

Figure 17. Implementation of a frequency-dependent rotation matrix for the decoder of fig. 16 using identical first-order all-pass networks with gain -1 below 5 kHz and +1 above.

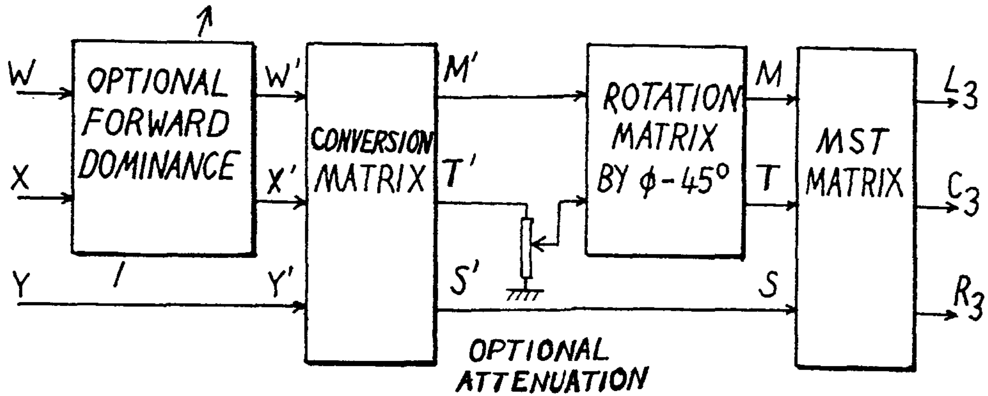

Figure 18. 3-loudspeaker decoder for B-format, showing T-channel gain adjustment before the frequency-dependent rotation matrix and forward dominance adjustment of the B-format input.

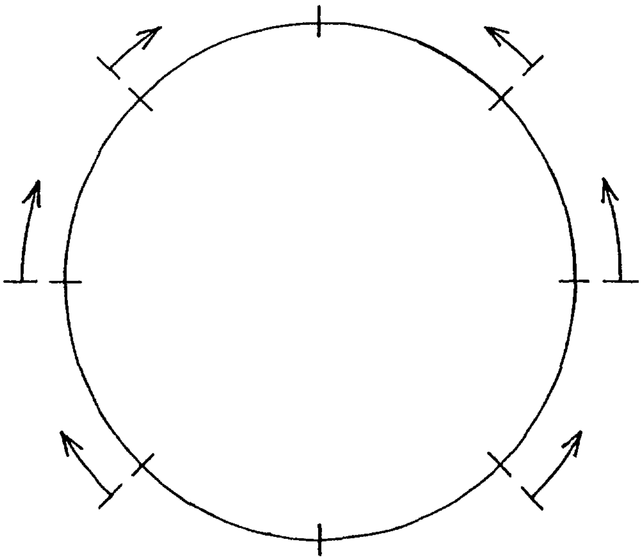

Figure 19. The effect of $\lambda = \sqrt{2}$ forward dominance on the encoded azimuths of B-format signals in the horizontal plane.