Reproduced from
Studio Sound, June 1990

Three Channels: The Future of Stereo?

Michael Gerzon takes a fresh look at 3-channel stereo in the light of recent high definition TV technology

The development of widescreen and HDTV systems for television has renewed interest in 3-speaker/3-channel stereo, supplementing the traditional left and right speakers with a central one in the middle of the screen. Such 3-channel stereo systems are far from new – they were first investigated at Bell Telephone Labs in the USA in the early '30s, and have been used by the film industry since Walt Disney's *Fantasia* in 1939.

The prime motivation for using an additional channel for a centre speaker is to 'lock' central sounds, such as on-screen dialogue, into a stable physical relationship with the screen. With 2-speaker stereo, the nominally central image is liable to wander around according to the location of the listener/viewer.

Domestic 3-speaker stereo operating from three channels is potentially capable of far more than merely 'locking' central dialogue into place. The extra speaker and channel can be used to improve the quality of other non-central image locations to be markedly better than from 2-speaker stereo. Via such improved phantom imaging at other locations, 3-channel stereo has the potential for a marked improvement even in audio-only applications.

Yet there is a strange paradox here. Despite the fact that 2-speaker stereo (deriving from Alan Blumlein's famous 1931 patent) and 3-channel/3-speaker stereo (deriving from the work of Snow, Fletcher and Steinberg at Bell Telephone Labs in 1933) are about equal in age, no one to my knowledge has seriously tried to design 3-channel stereo to optimise phantom images in the way that Blumlein and others have done with 2-speaker stereo. Snow and others did design 3-channel panpots in 1934 to create phantom 3-speaker images. Such panpots are still widely used in the film industry but their work was on a much more empirical basis than Blumlein's and no systematic psychoacoustic optimisation analogous to Blumlein's work with two channels was done.

Here we outline recent work done by the writer on the optimisation of 3-channel/3-speaker stereo systems to realise the potential for improved phantom images and superior quality from such systems. This work has used much of the psychoacoustic theory of directional sound localisation, originally developed in connection with Ambisonics but optimised here for frontal stage sound. This work has involved not merely understanding and optimising 3-speaker results, but also the development of simple practical technology and system design for the use of this know-how in the sound-mixing studio environment.

In previous work, the theoretical potential for 3-speaker stereo to improve on the phantom-image illusion of 2-channel stereo has not been realised in practical systems. In fact, some 3-channel panpots give inferior images near the central position compared to 2-speaker stereo for a central listener. This is very important when on-screen visual action is to be matched in sound position. This inferior imaging was not too important in large screen film presentation in auditoria, as creating localisation illusions in large auditoria is more hit-and-miss than in the home, due to the longer paths and larger time delays from the loudspeakers to the listener.

It is often forgotten that the Bell work on 3-channel spaced-microphone stereo was specifically aimed at reproduction in auditoria, not in the home. As a result, enthusiasts for Blumlein-type stereo have often unfairly criticised the Bell work as 'not properly understanding stereo' but they forget that Bell were not trying to solve the same problem (reproduction of phantom images in smallish rooms with reasonably centrally-placed listeners) as Blumlein.

While the Bell approach may be appropriate for large auditoria, including widescreen film presentation, it is undoubtedly not the optimum for domestic 3-speaker/3-channel reproduction. Such domestic systems can only realise the potential improvement of phantom image performance by careful design work extending Blumlein and others' work on 2-speaker 2-channel stereo.

Domestic uses

It now seems likely that 3-channel stereo will be used for the reproduction of frontal stage sounds in connection with domestic widescreen and HDTV systems. Extra channels for this are available in the various D-MAC and other broadcast systems across the world – and 'sub-band' systems can be used to add additional data-compressed channels compatibly to almost any digital sound broadcasting system, conveying the additional information via the least significant bits. The obvious advantage is that central dialogue is locked to the middle of the screen irrespective of listener/viewer position.

At first sight, one requires little more from a 3-channel TV audio system than a conventional 2-speaker stereo from the outer speakers for ambiance and sound effects, plus a central 'dialogue' channel. This is because of the various typical problems encountered in matching sound to picture.

Over the decades, we have learned to accept conventions of visual presentation that involve constantly changing angles and perspectives on a scene unlike anything encountered in real life. If one attempts to change the stereoism of the sound to match the picture, the ears are much less tolerant of sudden changes of position than the eye and in general, it is better to present a fixed stereo sound image even if the picture is constantly cutting between different viewpoints. (An exception to this rule is that it is sometimes desirable to fade up the level of sounds in the mix when the object producing the sound is subject to a camera close-up.)

The result of this need to keep the sound image fixed is that there is bound to be a mismatch between the sound and the visual positions of objects. Research has suggested that mismatch is audible over 4°, although up to about 11° need not be too objectionable. Thus such mismatches can often be kept to within the 'acceptable' region by piling up all the on-screen sounds at the centre position.

At first sight (and hearing?) all this makes true phantom imaging between the three loudspeakers pretty much irrelevant for TV use. But this is not so for a number of reasons we have so far not considered.

Matching the position of sound and vision is not the only reason for having stereo. In audio-only use, the main advantage of stereo (and probably the reason why it has become the standard domestic audio system) is not the direct ability to localise sounds at particular positions. Rather, if sounds are localised at different positions, it becomes much easier for the ears and brain to separate out the different sounds in a mix. The intelligibility improves, listener fatigue reduces and the subjective distortion is substantially reduced. The better the illusory quality of phantom images, the more these other advantages of the ears' ability to 'listen through' the sounds are obtained.

In conventional 2-speaker stereo, one of the simplest tricks to improve the subjective quality of a mix is instead of piling up several sounds into precisely the same stereo position, giving each sound a slightly different position. The same is true for stereo TV production and there is a strong argument not to pile up all on-screen sounds precisely at the centre of the screen but to distribute them across a narrow stage around the centre. But to do this well requires good quality phantom imaging near the centre of the screen, not just at the centre.

A typical example is the broadcasting of a TV quiz show. It is generally better if each contestant's voice is in a slightly different position from the others, even if this does not match the visual image. This way, if two contestants speak at the same time, it is easier for the viewer/listener to follow what both are saying. A similar consideration applies to interviews, broadcasts of drama, and of musical groups. The ability to form convincing phantom images greatly helps the intelligibility of stereo sound.

A second reason for good phantom imaging is that, especially with HDTV (or even with widescreen conventional-resolution TV with displayed-image enhancement), the need for rapid cutting of camera angles may well diminish in some kinds of programmes. A fixed-angle presentation usually loses a lot of important fine image detail, as compared to what one can see in real life. Close-ups often make up for this loss of resolution. For example, an overall camera shot of an orchestra shows little detail where a close-up may show the strings of the violins.

With HDTV, and its higher (albeit far from perfect) resolution, there is less need for close-ups, and they can use less drastic changes of camera perspective and angles. While, on artistic or conventional grounds, many programmes will retain current approaches to changing camera angle (and for several years to come, producers will still have to take into account viewers of low-definition TV), there will be an increasing area of programming using relatively stable camera angles. For such programmes, being able to match sound image positions to the picture is a realistic and useful option.

Good phantom imaging causes a higher audio fidelity and quality such as is required for music programmes and also ones requiring a very natural ambience or subtly creative sound effects (including adverts). In many such cases, matching the visual image is either totally unimportant, *eg* when a creative sound effect is used, or sound fidelity takes artistic precedence over image matching. In all such cases, being able to use the three channels to improve phantom images is important – it is certainly not adequate merely to convey 2-channel stereo via the outer loudspeakers, especially if these are widely spaced apart.

If the technical means for improving phantom imagery via three channels are implemented, one might be in the strange situation where TV audio will be capable of better overall fidelity than conventional audio-only programmes based on two channels. If this should happen, there would be strong pressure to incorporate the improvements from three channels into audio-only media.

Any technology for improved 3-channel stereo is therefore also potentially important for audio-only media in the future. There are various possible technical means of adding a third channel compatibly to conventional audio media, *ie* with digital media the third channel can be smuggled in as a data-compressed signal using up the two least significant bits of the two existing stereo channels, with special precautions taken to psychoacoustically mask this altered information for existing listeners. There is room for a third channel in FM radio broadcasting (the quadrature modulation of the 38kHz subcarrier) and cassette also has the possibility of a third channel within the space currently occupied by two. All these cases can be done in a manner compatible with existing mono and 2-channel stereo uses.

The only media where it could be difficult adding a third channel are the vinyl records, which are unlikely to survive long into the 3-channel era in any case because of CD, and AM stereo broadcasting, which is a low-quality medium in any case.

One of the main uses of 3-channel stereo is much less obvious. It turns out to be an extraordinarily useful production medium for mastering for mono and 2-channel release. If one mixes down to three channels as an intermediate stage, one has many options for improving the quality of the results obtained in the later final remix to 2-channel stereo. For example, one can derive 2-channel stereo with a wider image width or with better phantom-image psychoacoustics than with conventional 2-channel panpots by going through the intermediate 3-channel stage.

Also, a 3-channel mix can be remixed to other formats (mono, stereo, 3-channel Ambisonic surround-sound, psychoacoustically-improved stereo and even binaural) and allows special AM airplay, video mixes or rebalancing to be done simply without having to go back to the original multitrack. Thus, providing 3-channel technology can be got right, there is a strong reason to start mixing down to a 3-channel mastering format even when current release formats are still 2-channel.

Whatever means are used to convey a third channel, we see that the optimisation of 3-channel stereo is an important issue, affecting the whole future and well-being of both the audio and video industries. It seems wise for audio professionals, studios and the manufacturers of studio equipment to start facing up to the issue of 3-channel stereo at this stage, and not to leave all the important decisions to backroom industry committees who might get things wrong from the end-user point of view if they do not take on board the user's needs. It is also important to ensure that, as soon as is practicable, any programmes produced will be maximally compatible with future 3-channel use, so as to prevent premature technical obsolescence of programmes.

Psychoacoustics

All the above is contingent on getting good phantom imaging from 3-speaker stereo. Yet it is a fact that, as currently implemented, 3-channel stereo does not give particularly good phantom images even for central listeners. Far from being an improvement, the quality for central listeners of phantom images away from the centre of the stereo stage can actually be worse via three speakers and three channels.

To understand this and try to remedy it, we have to examine the psychoacoustics of image localisation. This is a complex topic but the theoretical methods were presented in relatively simple form in *Wireless World*[1] some time ago. Using these methods of psychoacoustic analysis, plus additional empirical know-how, one can optimise the speaker feeds for three speakers to give as good a localisation quality as possible. One certainly can't get everything right (and surround-sound Ambisonics can get certain things right that a 3-speaker system can't even in the frontal stage sector) but the results can be markedly improved and rendered subjectively much more convincing.

There are two basic theories used in analysing localisation of sounds, although there are other significant methods used by the ears and brain. One theory applies to frequencies below about 700Hz, and the other between about 700Hz and 6kHz. There are two different theories because the ears and brain use different methods of localisation below and above about 700Hz, which is the frequency at which the wavelength of sound becomes comparable to the size of the head. In practice, the transition between the low and high frequency theories is not sudden but there is a rather fuzzy band of frequencies over which both theories have some application.

Two-speaker stereo is capable of quite good low frequency localisation – the theory of which was understood by Alan Blumlein in 1931. However, this localisation is somewhat unstable when the listener moves or rotates his/her head. As the speaker separation is widened beyond the usual 60° angle subtended at the listener, this poor stability of phantom images markedly worsens, leaving the famous 'hole in the middle' and also in images that can be perceived as being elevated as discovered by de Boer at Philips in the 1940s.

This instability of low frequency phantom images cannot be cured even by improved 2-channel stereo panning methods (such as the transaural stereo of Duane Cooper and Jerry Bauck) or by special loudspeaker types (*eg* those recently designed by Canon) designed to improve image centring. This is precisely the area where 3-channel stereo can give marked improvements, as can Ambisonics.

An improved image stability at low frequencies can be achieved by using a 3-channel panpot law as shown in Fig 1. This shows, for each intended panned image position, the amplitude gain of the sound in the three channels. It will be noted that in this panpot law, the gain becomes negative, *ie* with a polarity inversion in one extreme speaker when the sound position is panned between the other two. The optimisation of this LF panpot law requires careful theoretical mathematics but its general form shown in Fig 1 is enough for general descriptive purposes. It is possible to design simple analogue panpot circuits and simple digital algorithms to implement such a panpot law well without any great complexity.

By way of contrast, Fig 2 shows the panpot law devised by Snow and others at Bell Telephone Labs in 1934. This obviously very different law has poor LF localisation properties but it was designed for use with reproduction in large auditoria, in which the full requirements of LF localisation cannot be met anyway.

Where the Bell panpot law wins over the law of Fig 1 is in its high frequency phantom image localisation. Fig 3 shows the computed HF theory localisation of the panpot law of Fig 1 via three speakers. This displays a pronounced and severe 'detent' effect at the centre speaker, whereby sounds panned (by the panpot law of Fig 1 optimised for low frequencies) fairly near to the centre are pulled right into the centre speaker at high frequencies. There is also a (much less pronounced) detent effect near the two outer speakers.

Now such detent effects are familiar with all directional sound reproduction systems with a 'discrete' panning law (*ie* one which positions some directions at individual loudspeakers without any crosstalk) and was identified in conventional 2-speaker stereo by Harwood in 1968, and by Kohsaka and others at Nippon Columbia for discrete quadraphonic systems. However, the HF detent effect at the central speaker of the 3-channel panpot law of Fig 1, shown in Fig 3, is very extreme – even sounds supposedly panned significantly far from the centre are pulled right to the middle.

What this means is that the LF panpot law – ideal for LF localisation – is just about the worst possible at high frequencies. The localisation of the Bell panpot law of Fig 2 is shown in Fig 4. While this has much better HF localisation, the localisation stability under listener movement is quite poor – including for central images. Also, the apparent HF localisation is about 1½x as wide at high frequencies as at low frequencies. (There is a somewhat similar discrepancy between LF and HF localisation for conventional 2-speaker stereo.)

The question thus arises of finding an optimised HF panpot law and of devising a reproduction method that conforms to the optimum LF law at low frequencies, and to an optimum HF law at high frequencies.

This question is complicated by several practical operational constraints. A frequency-dependent panpot circuit would be quite complex to implement, and the resulting 3-channel stereo would not be compatible for mixdown to conventional 2-channel stereo or mono – both of which will remain important for TV and other use. Moreover, ideally, one wishes to use the same mix both for auditorium film and home TV use, since remixing can be extremely expensive. However, the LF panning law is inappropriate for auditorium reproduction, where the large time delays from the loudspeakers make optimal LF localisation academic. Rather, for auditorium reproduction, one wishes to optimise according to HF localisation laws even at lower audio frequencies. It can be shown that such HF law optimisation gives the best obtainable LF phantom images under auditorium conditions.

This need for two different optimisations of 3-speaker feeds for auditorium and home use means that it is wise to use a single basic frequency-independent 3-channel panpot law, as shown in Fig 1, but that for playback or monitoring, the resulting 3-channel signal should be subjected to an additional 3-input 3-output processing (termed 'decoding') to produce the three speaker feeds with optimum psychoacoustics for a given environment. For film release prints, carefully designed crosstalk can optimise phantom image localisation in large auditoria, whereas for home use, such crosstalk will only be implemented at highish frequencies in the consumer decoder. The home decoder should be in the home and not at the recording or transmitter end, since one requires that the recorded or transmitted signal be compatible also with mono or 2-speaker stereo reproduction, or even with reproduction via alternative decoders via 4-speaker stereo systems or Ambisonic surround-sound systems.

Should the home user place the highest priority on locking central images to the middle of the screen, the raw left-centre-right signals from the panpot can be fed direct to the three loudspeakers. Those requiring more subtle phantom images will use a simple decoder network to achieve this. Fig 5 shows the HF localisation that can be achieved from the panpot law of Fig 1 via a simple decoding network. Comparing this with Fig 3 shows that the detent effect has been almost completely eliminated and that the low and high frequencies now match in position to within about 4° over a 90°-wide reproduction stage. (Similar matching of low and high frequencies can be obtained for other reproduced stereo stage widths.) The particular decoder involved still gives excellent stability of central images – far better than conventional 2-speaker stereo or the original Bell 3-channel panpot.

System considerations

The many possible reproduction modes of a 3-channel stereo signal require careful system design to ensure that all modes work well. We have already seen that it is possible to use the LF panpot law of Fig 1 with a decoder to provide improved phantom image localisation quality via three speakers, while giving a more rudimentary effect over three speakers without a decoder.

We have already mentioned the practical importance of having a compatible mix in mono and 2-speaker stereo and that the use of a frequency-independent panpot law, such as that of Fig 1, is vital for this if frequency-dependent mono and 2-channel results are to be avoided. However, 3-channel stereo mixdown proves to be a powerful production tool for a wide range of other important uses than mono, 2-speaker stereo and 3-speaker stereo.

Fig 6 provides an overview of how many of the different uses of the 3-channel material can be derived by subsequent signal processing. The basic idea is that, whatever reproduction mode is chosen, the left, centre and right channels are panned to their respective associated positions by panpots satisfying the appropriate law for that reproduction mode.

Other possible reproduction modes include 'wide 2-speaker stereo' conveying image positions beyond the two loudspeakers, 'psychoacoustic 2-speaker stereo' giving an improved phantom image illusion for half-left and half-right positions, Ambisonic 3-channel surround-sound for reproduction via a surround-sound speaker array, binaural reproduction for headphone listening, and transaural reproduction, *ie* a form of stereo reproduction via two loudspeakers aiming to recreate via loudspeakers binaural signals at the ears after the sound has mixed in the air.

For example, for conventional stereo, the left, centre and right channels are conventionally panned (with constant power gains) or a mixing circuit having the same effect is employed. This results in a conventional 2-channel stereo mix in which intermediate positions are panned conventionally with an almost perfectly constant power gain (within about 0.2dB).

A mono mix having substantially constant power gain at all positions in the original 3-channel stereo stage can be obtained simply by summing the three channels at equal level.

In Ambisonic surround-sound mixdown, the centre channel is panned to front centre, and the left and right channels to a given angle θ (up to say 80°) to the left and right of centre at the same level. This results in an Ambisonic mixdown covering a stage width 2θ wide, which can extend up to 160° wide while conforming to Ambisonic encoding specifications even for intermediate phantom positions.

A wide 2-channel stereo mix, with stereo images beyond the left and right speakers, can be obtained by mixing down to 2-channel stereo with the left and right channels panned to those beyond-the-speaker positions, and the centre channel panned to centre position, all at equal levels. This results in a 'wide' mix with all phantom positions at almost the same level. This is unlike widened conventional 2-channel stereo, which has excessive sound levels at the edges of the stereo stage.

Further refinements of 2-channel stereo are possible to give improved subjective results. Two examples of this are as follows. It has long been proposed to 'shuffle' conventional stereo, with wider width at low frequencies than at high frequencies, but such shuffling conventionally introduces an undesirable position-dependent frequency response. If one uses a separate low and high frequency panning matrix from three channels, via a (phase compensated?) crossover network, with the low frequency stereo mix being wide, then the resulting shuffled 2-channel stereo will have a substantially flat frequency response for all image positions. Such optimally shuffled stereo is ideal for widening reproduction from two closely-spaced speakers, such as those to the sides of the screen in an all-in-one TV set or in an all-in-one portable 'ghetto-blaster' unit.

A second psychoacoustic 2-channel mixdown from three channels uses the extra degree of freedom in the 3-channel law to optimise the phantom images halfway between centre and left or right, improving on conventional amplitude panning. Such psychoacoustic 2-channel stereo mixdown can achieve sharper phantom images away from the central panned position without a marked deviation from flatness of frequency response in either mono or stereo. Thus, by initially mixing to three channels and then going through a psychoacoustic mixdown processor, the phantom image results from ordinary 2-speaker stereo can be significantly improved without the complexity of mixdown (or the mono incompatibility) of transaural stereo.

Three-channel stereo material can be converted into a binaural or transaural mix by separately encoding and mixing the left, centre and right channels via binaural or transaural panpots. This gives good imagery of those three positions, although some intermediate positions will not be encoded correctly at higher frequencies. It is possible to modify the mixdown from three channels into binaural or transaural formats to spread these errors more uniformly across the sound stage, and thereby reduce their magnitude. Conversion from 3-channel stereo to binaural or transaural formats, however, can never be perfect but is a much better compromise than is possible from ordinary 2-channel stereo.

Finally, the raw 3-channel signals can be reprocessed to give sharper phantom image reproduction via three (or more!) loudspeakers via suitable decoding networks. The optimum decoding network depends on the circumstances. A normal domestic environment requires a different decoding circuit to a large auditorium environment (*eg* film, SR or A/V applications). A properly-designed psychoacoustic 3-speaker decoder can give quite well-defined phantom image positions without excessive image movement with listener position. It is also possible to design decoders that are to be used with four or more loudspeakers.

Besides handling all these different modes of reproduction, it is worth noting an important subsidiary advantage of 3-channel stereo mixdown. This is in applications where it is necessary to alter the level-balance of the mix. At the expense of some narrowing (or widening) of the stereo image near central image positions, the centre-channel material can be faded up (or down) by increasing (or decreasing) the centre-channel gain before reprocessing into the final reproduction format. This allows listeners/viewers of 3-channel stereo broadcasts or recordings to alter balance to taste and solves the problems of those with non-standard hearing who find it difficult to hear dialogue in the presence of background music or sound effects. Such listeners can fade up the centre-channel feed relative to the outer channels, whether they are actually listening in mono, 2-speaker stereo or 3-speaker stereo modes.

Another use of such level-balance alterations is with sound effects libraries or with library music. If recorded in 3-channel stereo format, the level-balance of the mix can be changed as it is mixed into the final programme in order to meet that particular programme's requirements. Other uses for preparing AM airplay mixes or dance remixes are also evident.

Operational aspects

Fortunately, there are various simple technologies for implementing optimised 3-channel panpot laws without great complexity. The simplest 3-channel analogue panpot can be realised by a minor modification of existing 2-channel panpot designs, involving two modifications to existing 2-channel stereo mixers: the addition of a few extra components in each mixer channel strip and the addition of another mixing bus (or the reallocation of one of the post-fade mixing buses); the addition of a moderate amount of extra signal processing after the three mixing buses. This will mean that most existing 2-channel stereo mixers can easily be redesigned for 3-channel use and most existing designs should be retrofittable for 3-channel use at low cost.

A 3-channel mixing desk can still be used for conventional mono and 2-channel stereo use by incorporating the conversion circuits for mono and stereo into the mixer – indeed the mono, stereo and 3-channel mixes can be achieved simultaneously for different release formats. Thus, at relatively little extra cost, studio and PA desks can be provided with the facilities for mono, 2-channel stereo and 3-channel stereo (perhaps also incorporating a 3-speaker decoder for monitoring). Thus, for example, a PA desk normally used for mono or stereo would be '3-channel ready' for those venues where a central speaker cluster is practical, and could give a 3-channel output for recording purposes even when the actual live sound is reproduced in mono or 2-speaker stereo.

Importantly, studio or SR desks of the 3-channel kind described will be operationally identical to present-day

2-channel stereo desks, apart from the user having to select the output mode to be used. Thus operationally, there is no relearning involved in using a 3-channel mixing desk.

Further refinements of the panpot law can be achieved by designing desks around purpose-designed optimised 3-channel panpots. There are very simple designs of such optimised panpots using all three major panpot technologies: digital mixing, VCA technology and using ganged pairs of linear potentiometers. Such optimised 3-channel panpots differ from the 'modified 2-channel panpots' described above in having slightly better psychoacoustics and noise performance – but in most situations, the differences are marginal. The optimised 3-channel panpots are, however, the preferred option when new mixer designs are developed mainly oriented towards the 3-channel market.

As noted earlier, 3-channel mixers will be of advantage even to those users only requiring 2-channel stereo, since the psychoacoustics of the resulting 2-channel stereo can be improved over standard amplitude-panned 2-channel stereo by using a psychoacoustic 3-to-2-channel conversion network. Also, by mastering in 3-channel format, the master can be re-released in future audio formats, *eg* 3-speaker stereo, Ambisonic, without remixing from the original multitrack, thereby protecting the investment in mixdown time.

There is still the problem of the lack of a standard 3-channel tape format. The only 'standard' 3-channel tape format is the 1950s ½ inch 3-channel analogue format, which still has much to commend it, especially at 30in/s. One of the available 4-channel digital recording formats could also be used, although it should be possible to modify existing digital 2-channel reel-to-reel formats to handle three channels. Users of digital 48-track machines can lay down a 3-channel master mix on three spare channels, and users of videotape formats having three or four audio channels should also have no problems in 3-channel mastering.

This being said, the industry needs to look carefully at mastering media for multichannel stereo.

Three-channel media

There is also the question of how we get three (or more) channels to the consumer. Digital media offer an 'easy' way to incorporate additional channels. Essentially, the least significant bits of the existing stereo channels can be 'stolen' and re-allocated to additional data-compressed channels. By appropriate dithering and noise-shaping, most of the subjective loss in the original channels of these stolen bits can be compensated. For example, the loss of the two least significant bits in CD can still give a psychoacoustically weighted S/N ratio for existing 2-channel listeners of around 94dB – about 3dB better than currently achieved with the full 16 bits. This subjective improvement is due to the use of optimally noise-shaped subtractive dither, based on work by Peter Craven and myself on optimal dither and noise shaping.

In digital satellite broadcasting, a compatible digital 'sub-band' method of using the least significant audio bits for additional channels has been proposed by Philips. Such systems can be fully compatible with 'existing' listeners, *ie* those listening to just the basic channels, since the altered least significant bits can be well masked by the basic audio signal by a judicious use of noise-shaping and dither. To minimise loss of quality, this must be done with great care, using proven results. Decoded 3-channel results can be further improved by using a carefully-designed subjectively compatible companding system, and I have devised appropriate algorithms for this.

Intercompatibility

There is an issue we have not yet really dealt with – the intercompatibility of different multichannel audio formats. For example, supposing that we have a 2-channel stereo programme, *eg* sound effects, library music, commercial music recordings, how can this best be conveyed via a 3-channel stereo medium? Again, suppose we have a 3-channel Ambisonic surround-sound mix. How can this best be reproduced over 3-speaker stereo? With a variety of different signal formats co-existing, all conversion options need to be considered if we are to avoid chaos. The fact is that several audio formats do exist and one will often need to use material from one format in another. This problem is not a new one in that binaural and 2-channel stereo have never really been compatible with one another – both binaural over speakers and stereo over headphones sound wrong. So far, we have merely lived with this problem without solving it but the addition of further formats make it important to think out solutions before the problems get worse.

The problem of reproducing 2-channel stereo over three loudspeakers turns out to have some reasonable solutions – although the results are obviously not as good as full 3-channel stereo. For narrow stage widths, 2-channel material can simply have its left and right channels panned (using 3-channel panpots) to the desired positions but this solution works poorly for wide stage widths, especially when one wishes to fill the whole stereo stage.

We have discovered some remarkably effective linear 2-channel 3-speaker decoding matrices capable of improved image stability for non-central listeners and improved image sharpness for central listeners, as compared to ordinary 2-speaker reproduction, and to prior proposals, *eg* the Bell/Klipsch 'bridged centre channel' method, for 3-speaker reproduction of two channels. Obviously, such 3-speaker decoding of two channels cannot be as good as true 3-channel decoding but it is quite effective on a wide range of material. We hope to be able to publish both the theory and methods of such 3-speaker decoding of two channels in the near future but it does seem to offer a good second-best option to true 3-channel stereo.

We are rather sceptical about the use of 'logic', 'gain riding' or 'variable matrix' 3-speaker decoders for 2-channel stereo, due to their signal-dependent 'pumping' side-effects, which can cause both dynamic wandering and instability of subsidiary images and increased listening fatigue. This negative comment is not based only on experience of commercially available designs but also on detailed development work on advanced experimental multiband logic decoders based on more sophisticated psychoacoustic design than those on the market. Most commercial logic decoders have paid very little attention to the proper localisation of image directions between the loudspeakers.

A 3-channel receiver needs to know which reception mode is in operation to reproduce each mode optimally over the chosen loudspeaker layout. I suggest that the reception modes involve at least the following options: mono, 2-channel stereo, 3-channel stereo with the panpot law of Fig 1, 3-channel 3-speaker feed-signal mode and 3-channel Ambisonic surround-sound mode. For digital broadcasting, suitable flags in the data stream could indicate the mode being transmitted.

For each of these transmitted modes, a different optimum psychoacoustic matrix is required to feed the three loudspeakers of a 3-speaker stereo receiver. For example, 2-channel stereo requires the use of an optimised 2×3 decoding matrix as described above, a mono signal will be fed to just the centre speaker, a 'figure 1 law' 3-channel signal will be fed to via a 3-channel decoder, speaker-feed 3-channel signals will be fed straight to the loudspeakers, and Ambisonic signals will be fed to the speakers via another matrix.

These transmitted signals can be designed to minimise the need for mode switching if only a basic 3-channel stereo effect is required but a receiver wishing to get optimum results from each mode will require to know the transmitted mode so decoding can be optimised for that mode.

This complication arises because the optimum decoding for each reception mode is frequency-dependent due to the frequency-dependence of human directional hearing, and the frequency-dependent speaker feeds for optimum reproduction are not compatible with a frequency-independent mono and stereo fold-down for mono and stereo listeners.

Also, the transmission of a mode flag is important because future technological developments may reveal future improved decoders, and set designers should have the option of incorporating these improvements into receivers, which should detect the mode being received. However, system standards should be such that a basic 3-channel stereo receiver without mode switching will receive an acceptable, if not psychoacoustically ideal, result.

Besides the modes described above, TV transmitting systems may also wish to include other modes, such as 4-speaker stereo, different varieties of horizontal surround sound, and even full-sphere surround-sound. Any system design for HDTV sound should find a way of making all these modes as intercompatible as possible. Such a system design is feasible but requires very careful thought. Certainly, the one approach to avoid is one that assumes a once-and-for-all rigid loudspeaker layout since this will limit any future improvements in the art.

Programme origination

Besides 3-channel panpots, one also needs means of producing 3-channel stereo from non-monophonic sources, including live soundfields. For live soundfields, one possibility is to use spaced mono microphones (as did the experiments at Bell in 1933) panned to positions across the 3-channel stereo stage, *eg* four microphones might be panned to nominal azimuths at ±45° and ±15° for the panpot law of Fig 1. An ideal 3-channel [signal] cannot easily be derived from available coincident microphone arrays but a quite reasonable non-ideal 3-speaker feed signal can be obtained via a suitable 3×3 matrix circuit from a soundfield microphone. Where such a very approximate feed is not adequate, a soundfield microphone can be matrixed to give an accurate match to the panpot law of Fig 1 for sounds arriving from a frontal stage but at the expense of an excessive pick-up of sounds from the rear. If it is possible to place a large acoustic absorber behind the soundfield microphone, *eg* below or above the field of view of an HDTV camera, or disguised behind scenery, then such an accurate matrix might be workable otherwise rear sound pickup is a serious problem.

Two-channel stereo material, such as from sound effects recordings, library music, stereo microphones or commercial music recordings, can be mixed into a 3-channel programme either by restricting it (by means of 3-channel panpots) to a small part of the stereo stage, or by using suitable 2×3 matrix decoders as described earlier. The latter option does tend to give less good image quality than true 3-channel material, and has the problem that it must be made compatible with 3-channel decoding from the 3-channel mixed programme.

An alternative microphone technique can use spaced stereo-pairs of microphones each, panned across a relatively small part of the 3-channel stereo stage. This gives more convincing phantom images than do spaced mono microphones. The use of two or more stereo pairs placed at different locations and mixed into the 3-channel stage might often prove to be a practical means of live stereo pickup for HDTV applications, with 3-channel panpots being used to control the imaging from each pair within the 3-channel stage.

Given the fact that new microphone techniques are still being developed for 2-channel stereo almost 60 years after its initial development, we expect much innovation to occur with 3-channel microphone technique in the future – varying from the development of proper all-in-one 3-channel stereo microphones to quite sophisticated 'matrix' techniques developing the 2-channel MS technique. However, it is not to be expected that the empirical rules-of-thumb for 2-channel stereo mic technique will always work with 3-channel stereo.

The artificial reverberation of 3-channel stereo ideally requires the use of a reverberation unit with three or more appropriately related independent outputs panned across the 3-channel stereo stage. There may be suitable units on the market, *eg* the Yamaha *DSP* processor series, and other surround reverb units. Two-channel output reverb units can be used if they are fed into the three channels by an appropriate 2x3 decoding matrix as described earlier but this will in general not give as good results.

Monitoring and domestic playback

The ideal loudspeaker layout for 3-speaker stereo is of the general form shown in Fig 7, with all three loudspeakers lying on a circle centred at the nominal ideal stereo seat. The equal distance of all speakers from the ideal stereo seat gives maximum phase coherence for phantom imaging, and helps optimise performance away from the stereo seat.

There is no obvious optimum subtended total angle of the loudspeaker layout at the ideally-positioned listener – figures between 60° and 180° have been suggested. Our panpot law of Fig 1 is optimised for good results for any subtended angle up to 160° although image stability degrades as the subtended angle increases, being poor beyond a 120° angle.

For audio applications, there is no need to adopt a rigid standardisation of angle as long as the reproduction method is designed to give good phantom images for the angle used by the listener. For TV applications, however, it is important that sounds from on-screen images should substantially match the position of the visual image. However, for a given loudspeaker layout, it will be possible to incorporate a 3-channel 'width control' adjustment that will allow audible and visible image positions to be matched.

While decoding to three loudspeakers is the nominally correct way of reproducing the 3-channel sound, the use of three loudspeakers may not always be practical or desirable. This is because a middle speaker will be either in the middle of the TV screen, or in the middle of a control room window.

Ways round this are either to accept a speaker below or above a picture, with an associated height error, or to use four (or more) loudspeakers. These speakers can involve either a narrow or 'inner' stereo pair and a wide 'outer' one, or can split the central loudspeaker into a 'below-picture' and 'above-picture' pair. In either case, the speakers must be provided with psychoacoustically optimised feeds adapted to the specific layout in use in order to get an optimised image illusion. Naively-chosen speaker feeds will not work well. A number of possible 4-speaker decoders have been devised for use with 3-channel stereo signals.

Although there are several different options for monitoring a 3-channel signal, different monitoring arrangements do sound slightly different, so thought needs to be given either to devising a standardised monitoring arrangement or to understanding the differences between different arrangements, so their effects can be allowed for.

The basic design theory for decoding 3-channel stereo assumes that all speakers are at the same distance from a central listener. If this is not the case, *eg* if the loudspeakers all lie in a straight line, then the speakers closer to the listener can be fed via a compensating time delay (and also a slight gain reduction) to restore the correct phase coherence of the sounds reaching the stereo seat. Such delay compensation is difficult in analogue systems but is quite easy to implement in systems with digital recording, transmission and signal processing.

In order to design different monitoring and decoding arrangements to meet the widest range of needs, it is important that there be a basic reference method of panning sounds into three channels, such as that of Fig 1. This acts as a reference for evaluating the quality of imaging of different designs. One expects future innovations to discover improved or refined decoders for different speaker layouts but the optimisation of such decoders requires knowing what is to be decoded.

Surround sound

This article has been primarily about stereo over a frontal stage since the instability of phantom images of 2-speaker stereo is an important defect of existing technology, especially with an associated visual image. However, the extension to the 360° of horizontal surround-sound, to height portrayal, and even to the 4π steradians of full-sphere surround-sound, is also an important issue, which we cannot fully deal with here. The most reliable existing surround-sound technology is that of Ambisonics, which requires the use of three transmission channels for horizontal surround sound and four transmission channels for full-sphere surround-sound[2].

Such surround-sound is capable of reproducing sounds from every direction while satisfying a variety of psychoacoustic requirements for directional localisation. There is empirical evidence that supplementing such systems with an additional front-centre channel and loudspeaker for large-screen and auditorium applications can be a useful enhancement. Such an enhancement can be done in a way compatible with 3-channel stereo for the frontal sector of directions.

Conclusions

Three-channel stereo is not simply two sets of stereo pairs (left/centre and centre/right) but properly designed technology using all three speakers and channels together and capable of subjectively enhanced realism as well as the improved stability of central images.

To the writer, one of the big hidden gains of 3-speaker stereo is its lower listening fatigue and artificiality as compared with 2-speaker systems. If properly-designed studio technology is used, the results will not only provide a better match to widescreen TV but offer a superior sound and a more social listening experience for several listeners in a room for audio-only applications.

It is essential that the industry makes the right decisions both about the systems aspects of 3-channel stereo (including mono and 2-channel compatibility) and about the right production technology (notably mixer design and monitoring methods, but also 3-channel mastering formats) to fully realise the potential gains. This would include the potential benefits of using 3-channel mastering as a production format even for 2-channel releases, and the adoption of the same formats for audio-only and TV applications.

I am preparing a detailed technical report aimed at professional equipment manufacturers and major users that will flesh out the above with much detail, both in general theory and detailed designs and methods, particularly as regards mixers and decoders. However, anyone seriously interested in keeping up with the future of stereo would do well to familiarise themselves with the literature covering work already done over the decades. The history and basic theory of 2- and 3-channel stereo is well covered in a useful compendium of important technical papers[3]. An informed knowledge of the technical foundations of stereophony among audio professionals will help them contribute to the important technical decisions that will determine the future of stereo technology.

The author is preparing a detailed technical report on 3-channel stereo, which will be made available to professional audio equipment manufacturers in the audio and video industries.

References

1) Gerzon M, 'Surround Sound Psychoacoustics', *Wireless World,* December 1974
2) Gerzon M, 'Ambisonics in Multichannel Broadcasting and Video', *JAES,* November 1985
3) *Stereophonic Techniques,* Audio Engineering Society

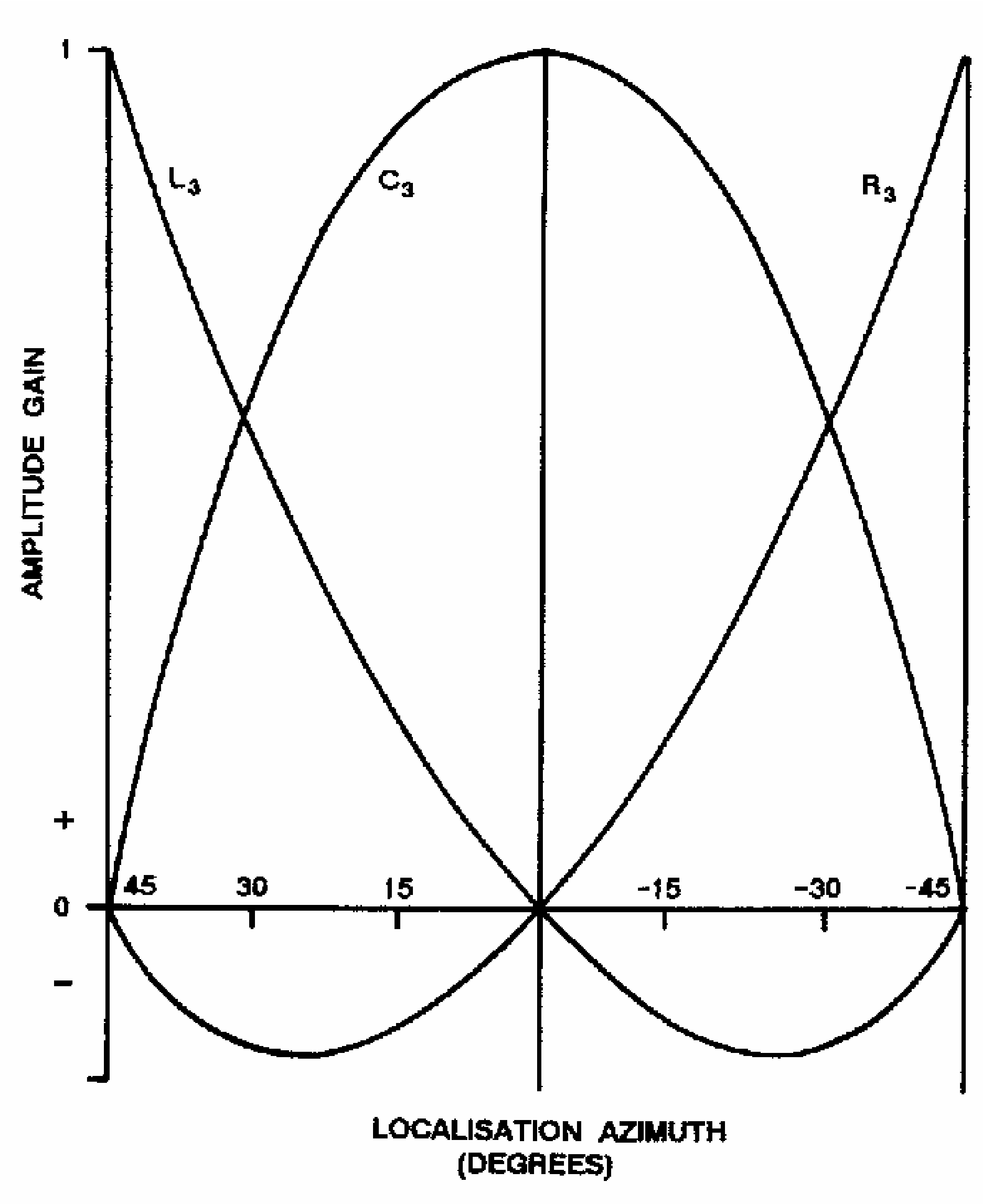

Fig 1: Typical 3-channel panpot law for optimum LF localisation

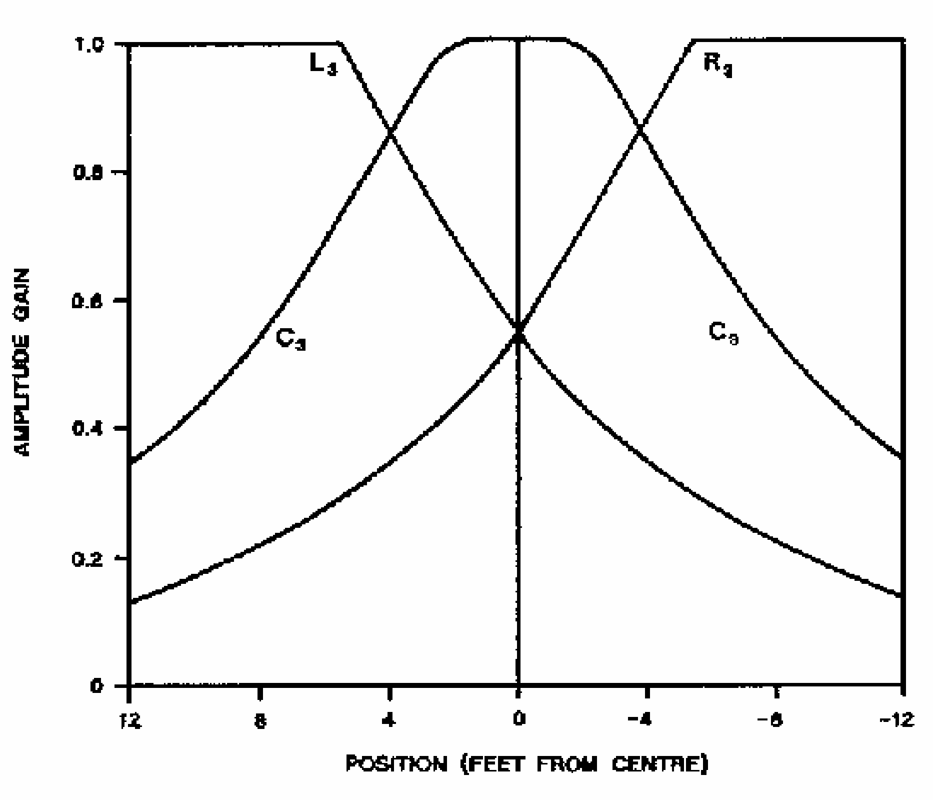

Fig 2: Three-channel panpot law used by Bell Telephone Labs in 1934

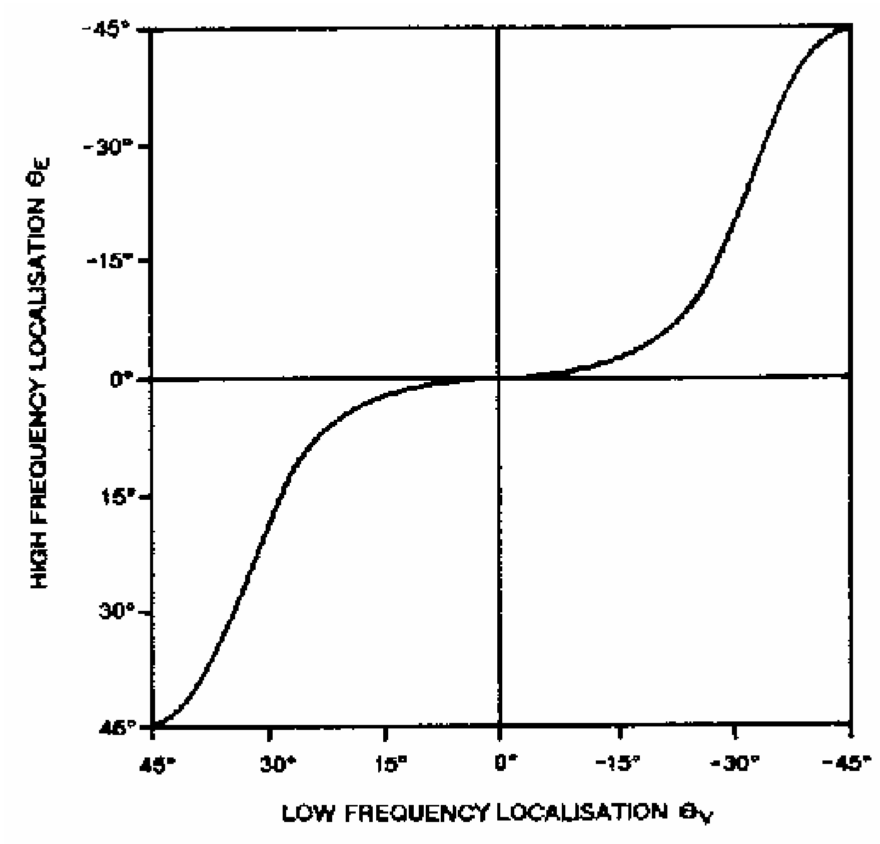

Fig 3: Apparent HF localisation angle θ_E plotted against LF localisation angle θ_V for the panpot law of Fig 1, via a 3-speaker layout subtending 90^O at the listening position

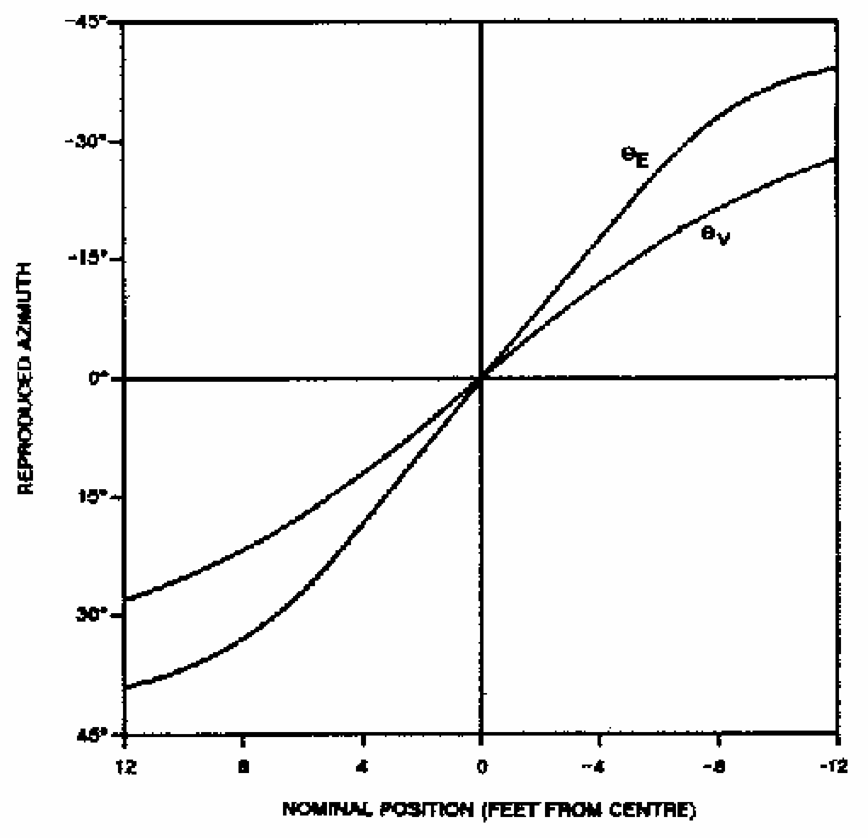

Fig 4: LF (θ_V) and HF (θ_E) localisation azimuths for 90^O speaker layout for the 1934 Bell 3-channel panpot law of Fig 2

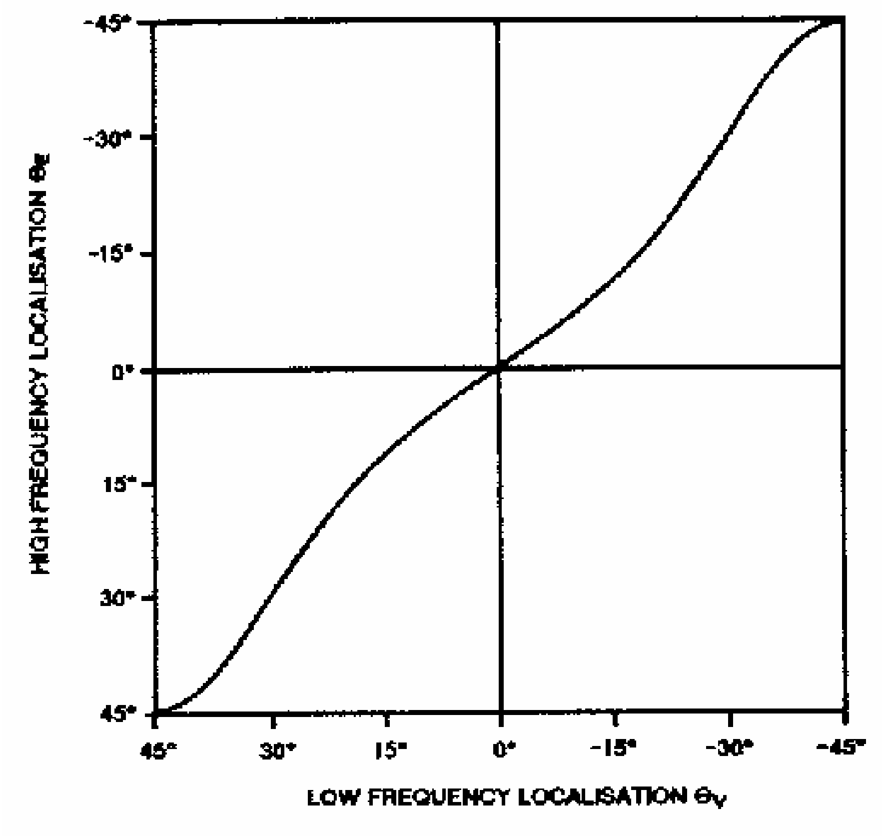

Fig 5: Apparent HF localisation angle θ_E plotted against LF localisation azimuth θ_V for one design of improved 3-channel decoding network

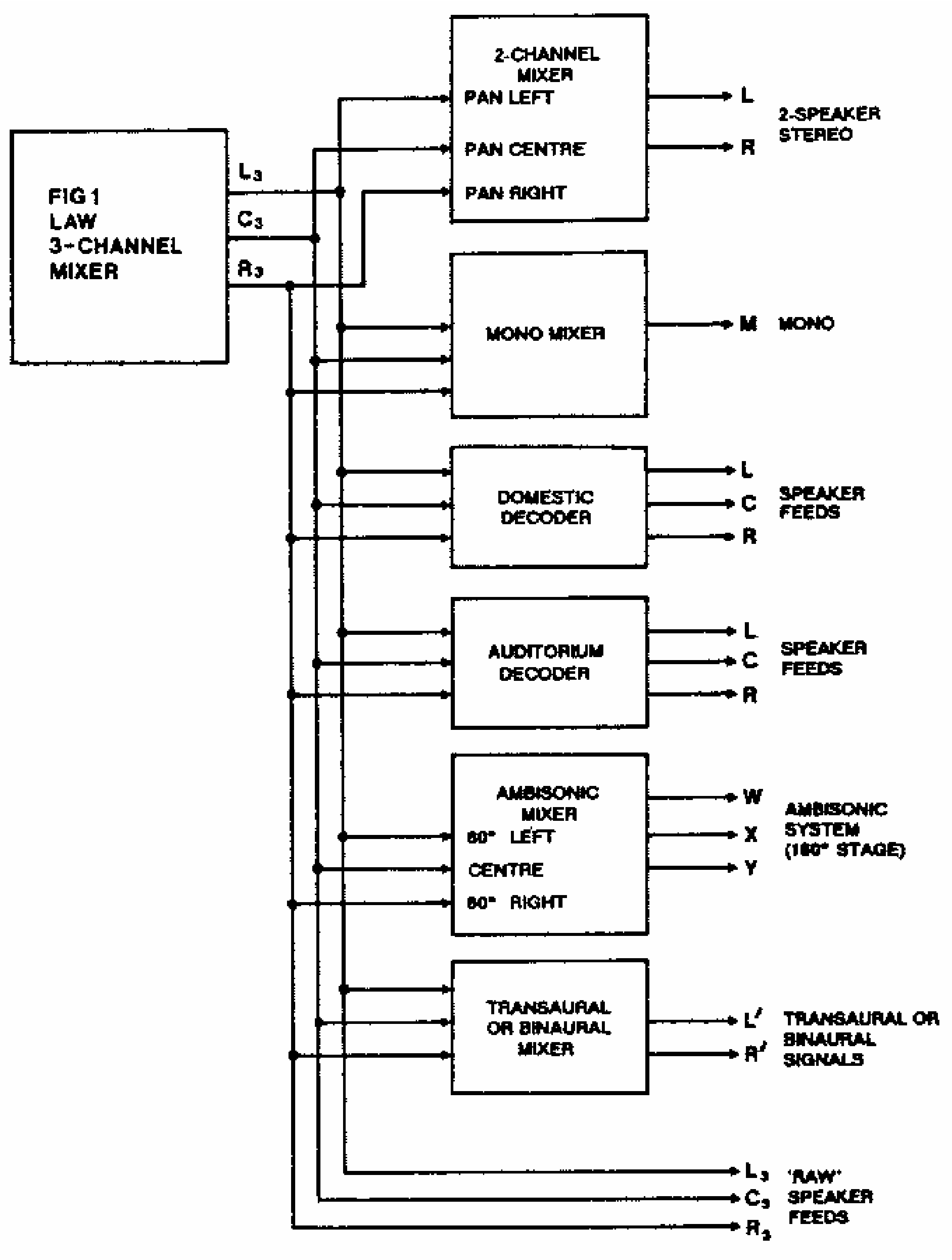

Fig 6: Different uses of basic 3-channel signals satisfying the LF frequency panning law of Fig 1

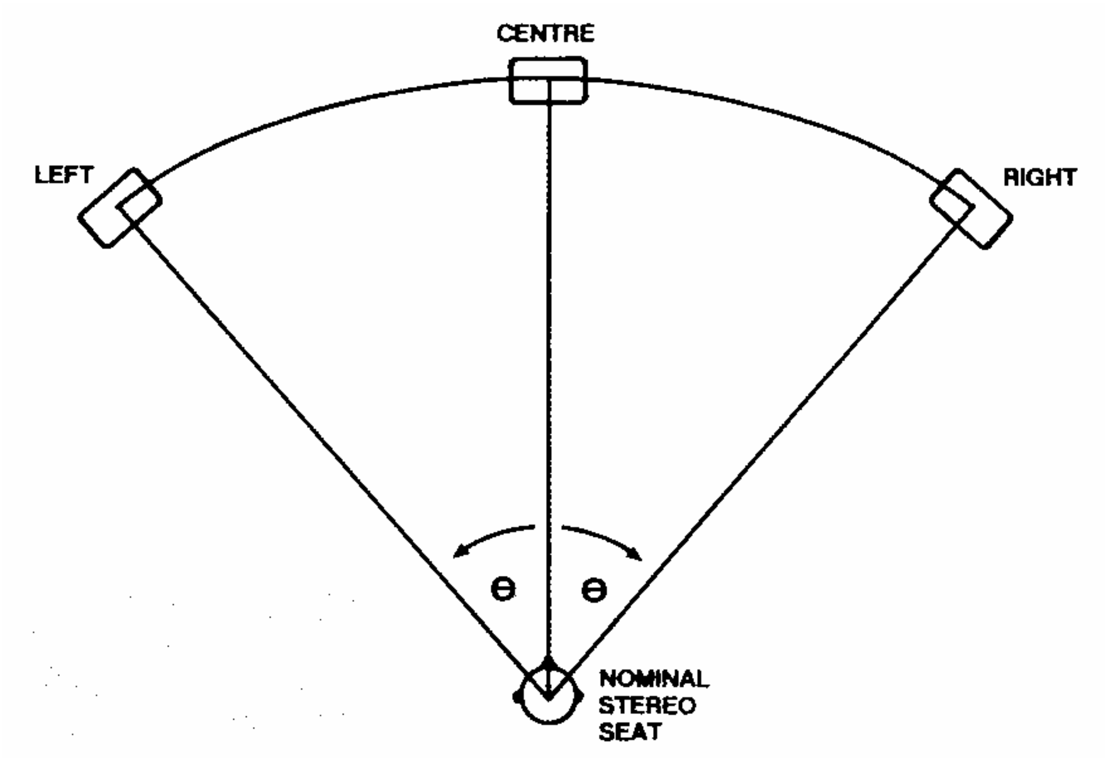

Fig 7: Three-speaker layout subtending an angle 2θ at a central listening position.
The three loudspeakers should be at the same distance from the stereo seat