Proceedings of the Institute of Acoustics

SOUND SYSTEMS FOR HDTV

D. J. Meares.

BBC Research Department, Kingswood Warren, Tadworth, Surrey.

1. INTRODUCTION

Within the foreseeable future High Definition Television (HDTV) will be available in the UK, either as an extended definition format such as High Definition Pal or as a fully fledged HDTV system with approximately double the definition in all three dimensions. That this has long been recognised is evidenced by the large number of research institutions around the world involved in related developments. What is not terribly surprising is that the vast majority of the effort has so far been expended on the vision side of the HDTV problem: very little effort, by comparison, has been applied to the question of what form of sound should accompany this new format of picture.

That is not to say that no effort has been applied to the sound studies. Indeed it is the purpose of this paper to draw together an outline of the work that has been carried out and to indicate those areas where most work is still needed.

2. THE VISION FORMAT

Though there are competing formats around the world, based on the local existing television systems and specifically the picture repetition or frame rate, the scene can be adequately set by referring to the HDTV system being developed in Europe by the Eureka 95 Consortium.

Remembering that the existing UK TV system is 625/50/2:1 (i.e. 625 lines/50 fields per second/2:1 interlaced line structure), the new HDTV proposal is for an ultimate studio standard of 1250/50/1:1. This means that there will be approximately twice the definition of picture both horizontally and vertically. Also the move from interlaced (2:1) to progressive (1:1) scanning will give twice the temporal definition, i.e. motion portrayal will be improved as well. This is shown schematically in fig 1, where to be more precise, the figures for the active picture area are shown. (Thus the overall line rate of 1250 lines per picture is shown as 1152 lines per active picture.)

Also shown is one of the most immediately noticeable features of HDTV, namely the change in aspect ratio from 4 x 3 to 16 x 9. Subjectively this new aspect ratio is much more pleasant to the viewer, perhaps reflecting the normally horizontal nature of nature, or the fact that our eyes are disposed that way and thus give a wider lateral perspective than vertical perspective. Obviously there is nothing fundamentally new about this, the film industry has always

worked with wider aspect ratios, but it does make an impact in the context of home viewing of television.

One should also bear in mind that these developments are accompanied by the proposition that larger pictures will be the norm for HDTV. Again this may be partially following the example of the film industry and/or real life, where the image presented to the eye covers a much greater part of the retina than would the image of a normal television picture at normal viewing distances. But how large a picture will be considered acceptable in the domestic environment as opposed to the cinema?

Tests recently carried out as part of the Eureka 95 study [1,2] looked specifically at this point asking not only how large a picture could be fitted into peoples' existing home setups, but also asking how large it could be even taking into account a rearrangement of the TV room or lounge. Results of this study confirmed suspicions that there was indeed an upper limit to the size of picture that would be acceptable. For existing arrangements the preferred size was 1m, whilst for rearranged conditions the preferred size was 1.21m.

It would appear therefore that, with the advent of HDTV, viewers will be subjected to pictures of much greater clarity and size. Surely then the accompanying sound will also have to be that much more impressive as well.

3. SOUND SYSTEMS: THE FUNDAMENTAL REQUIREMENTS

One way in which the sound system for HDTV could be more impressive, would be to provide surround sound instead of mono or stereo. In this context some of the pioneering work of the seventies should be borne in mind, even though that work was for sound-only systems.

The particular study of relevance [3] examined the accuity of the ear to directional cues. (The term ear is used here to cover the ear/brain combination.) Fig 2 shows the accuracy with which a group of subjects were able to detect the direction of a sound, whose source they could not see, from all directions in an anechoic chamber. Though somewhat less accurate for those directions out of the frontal quadrant, nowhere could the results be said to be poor. Fig 3 shows the results for the co-positioning of two sound sources under otherwise similar test conditions. These results are even more accurate.

However if one looks at the generation of phantom images between pairs of loudspeakers arranged around a listener in a square format, very variable and uneven results are produced. Whilst front and back quadrant results are rather similar, the side quadrant images show a strong tendency to be drawn to the front of the listener and are very unstable with head movement. There is also a noticeably diffuse quality to the side image making it less certain just where they are located. Thus the use of surround sound is not conclusively the right way forward.

SOUND SYSTEMS FOR HDTV

In the domestic environment there is seldom a household where people listen/view only one at a time. It is much more common and probably the norm for families to wish to view collectively, at least some of the time. The BBC's experimental Television Stereo work, now amounting to the production of many hundreds of programmes, has shown that whilst aural and visual directional cues can be made to match for a single viewer/listener, it is much more difficult to achieve such a result even for just two people sitting side by side. If therefore those directional cues are to be considered a valid and useable part of the producer's armory in creating an element of a programme, he has to be sure of where the sounds will be reproduced for his audience. It could be argued in the context of HDTV that if the picture is both larger and more detailed, then the importance of this factor is increased.

This point has been addressed in a recent CCIR submission from the Federal Republic of Germany [4]. Fig 5 shows the relevant results demonstrating clearly that an increase in the number of loudspeakers (and channels) used to reproduce a frontal sound stage will increase the useable listening area for a given accuracy of sound reproduction. The move from two channel to three increases the width of the listening area by a factor of about three, whilst the use of four channels increases it by a further factor of four.

Another factor that should be borne in mind is that the eye is more powerful than the ear: if there is doubt as to where a sound is coming from the brain places more emphasis on visual cues than on aural cues. Recent work by NHK in Japan [5] has examined this in some detail. Fig 6 shows both the experimental setup and the results for this work. Specifically they found that for a picture of a talking person if the sound was mislocated by 10 deg listeners were easily able to detect it, even if they were not disturbed by the mislocation. However an error of 15 deg or more started to cause some annoyance to the audience. Such an error is easily created by off-centre listening.

What then is the optimum choice of sound system for HDTV? Specifically how many channels are needed, and how should they be reproduced? Though the Japanese have again addressed this question [6], it is one that is still being explored by many workers in the field. The Japanese results can be summarised by fig 7. This shows seven different loudspeaker arrangements and the results of evaluations on them, both for realism and for stability and accuracy of sound location for an off-centre listener. Interestingly, it is the two surround systems with a central loudspeaker between the stereo pair that score well on both counts. As far as directional cues are concerned the three channel system also scores highly. This work however only used one musical item for its test material and thus much more work is required before such results could be claimed to apply to the whole gamut of television programmes. It is however a useful pointer for other workers.

It is just this question of how many channels and how they should be disposed that the author is studying at the moment. With programme colleagues at the BBC, several recordings of multichannel sound with HDTV have been made and subsequent mix down sessions in a Sypher channel have now started. The first

sessions covering both the 1989 Wimbledon and FA Cup Final have been completed and additional tests on a Promenade Concert recording and Top of the Pops recording are being planned. The Sypher sessions took the multichannel source tapes and attempted to produce mix downs that could be reproduced compatibly in 5 channel, 4 channel, 3 channel, stereo or mono. Similar work will take place for other types of material and then the success or otherwise of the different forms of sound presentation will be judged by way of subjective tests.

4. SOUND SYSTEMS: ADDITIONAL FACTORS

The significance of the question of compatibility should not be underestimated. Whilst a system may be developed that can provide, say, five channel surround sound for HDTV, just how many of the audience will want surround sound? When that question was raised in the seventies for Radio, the answer was "Not many". Admittedly there were other factors involved, such as how could it be transmitted, but public resistance to the complexity of surround sound should be taken into account. In essence a specific room would have to be given over to the activity of HDTV viewing so that the correct loudspeaker arrangement could be set up. Even given an arrangement that allowed more than one listener at a time, would an optimum arrangement require centrally located seats or could the system be modified to accommodate the conventional domestic arrangement of seating, i.e. seats around the walls? In this context it is worth quoting the final paragraph of reference 1.

> "The results help to confirm the view that the absolute size of domestic HDTV displays will be determined by practical limits such as room size and the willingness or otherwise of the householder to rearrange family viewing conditions to suit the HDTV display."

Though written specifically on the question of screen size, with a few choice word changes and the passage of a short period of time it may well be shown to apply equally to the sound side of HDTV.

And what of the non-surround listener? Would the transmitted sound signals be optimised with him sufficiently in mind? In the seventies it was clearly apparent that the reduction from surround sound to stereo in a multichannel system was not a simple matter. Some programmes, such as music from any concert hall, required less reverberation in the stereo than in the surround: thus rear channels should perhaps be attenuated in the compatible signals. For drama, does the same compatibility matrix provide the right answer? Insufficient work has yet been carried out to answer this reliably.

There is also the thorny question of multilingual broadcasts. If the HDTV signal is to be provided by a satelite service then it is likely to cross international borders with inordinate ease. Indeed, the cost of HDTV programmes may mean they are viable only if made with an international audience in mind. How many channels are needed to provide a multilingual service? For programmes such as sport or music with commentary, then a clean feed of surround sound

effects could be provided on, say, four or five channels (depending on the answer to the earlier question) and each language could be provided merely by adding one commentary channel per language. This cannot however be assumed to be true for other types of programme. As soon as the human voice becomes part of the programme, as in drama, and is intended to be in the environment being portrayed by the picture and the surround sound presentation, then each language requires a separate surround sound mix. One can only portray a person in a reverberant environment by ensuring that the reverberation is triggered by that person's voice.

5. TRANSMISSION FACTORS

As already mentioned the European thrust into HDTV is being coordinated by the Eureka 95 Consortium. This covers all aspects from the studio through transmission to the receiver. The transmission of HDTV is being designed to be an evolutionary step from the MAC form of transmission. Thus as far as the sound signals are concerned, they will be conveyed via the digital "packet system" already part of the MAC/packet specification [7].

The data capacity available to the sound signals depends firstly on which form of satellite service is being considered. MAC signals destined for onwards conveyance to the ultimate customer via a cable distribution signal will be of the D2 format, whilst those intended for direct reception will,in the main, be of the D format. In rough terms the D MAC format provides for the equivalent of eight high quality audio channels, whilst the D2 format provides for only four high quality audio channels. There is also already pressure to consider the release of some of this capacity for other data services, such as teletext. Under such circumstances the use of all four channels of a D2 system just to provide surround sound to the few who want it could be considered a very extravagant use of the available capacity. Furthermore such an option leaves no capacity whatsoever for the provision of multilingual broadcasting. The hunt is on therefore to try to find ways of releasing data capacity by reducing the requirements of a surround sound service.

One such development, being studied by Philips in Eindhoven, uses a variation on the theme of subband encoding to reduce the data rate, see fig 8. Essentially the incoming signals are divided into 'main' and 'hidden' channels and pairs (one of each type) are processed together. For each pair, both signals are processed via the subband codec to determine how many bits of each audio word are required to convey the necessary audio impression. The 'main' channel is then allocated the approriate number of MSBs in the digital word whilst the remaining LSBs are used to convey the 'hidden' channel. If the subband rules have been correctly applied, then the combined signal will just sound like the main signal, the other one being 'hidden' by the rules of auditory masking. If this system can be shown to work on a wide range of material, and at the time of writing the real-time hardware is only just coming on-stream, then it will mean that two signals can be conveyed via the capacity of one channel. Additionally the combined signal will have to be compatibly

received by MAC receivers without the subband decoder.

Yet another strand of the Eureka study is accepting conventional digital encoding of the first two audio channels to provide the necessary MAC compatible reception of the signals and is proposing to use the full strength of either subband or transform coding to reduce the data rate of the remaining audio signals by factors of 4:1 or more. Thus surround sound would require no more than two and a half channels of the four channel D2 capacity.

But why not use one of the 4-2-4 phase/amplitude matrix systems proposed in the seventies investigations? Specifically, why not use the Dolby Surround system now accepted as a de facto standard in the film industry. The main problem with these systems is the very limited interchannel separation that they provide unless one resorts to the use of non- linear circuit, or logic, enhancement. Such logic enhanced decoders subjectively overcome the limited separation, but impose severe limitations on the producer's freedom of placement of sound images. Figure 9 shows just how extensively the BBC producers use the stereo sound stage in TV Stereo productions. Such complexity of sound stage leads to very interesting sound productions, as was clearly demonstrated at a previous Reproduced Sound Conference. The use of Dolby Surround would restrict a producer's work and eliminate such complex presentations.

Clearly then the way in which the sound signals of an HDTV service will be transmitted has not yet been determined. It is obvious, however, that whatever system is to be developed, it should not at this stage predetermine what sound signals are envisaged or could be conveyed. Obviously an upper limit of channels may have to be set, but it should not override the necessary programming studies and conclusions of whether surround sound or whether multilingual broadcasting is appropriate.

6. CONCLUSIONS

New HDTV video formats are being studied to bring to the home viewer greater definition and reality to their programmes. On a similar basis new forms of sound system are being examined.

Much work has already been carried out to study relevant properties of the human hearing system, but much more still needs to be done. Experimental programmes have been made by the BBC and others in order to decide whether surround sound or some lesser form of sound presentation is required to support the new pictures. Similarly the possibility of multilingual programming is being studied fully to exploit the possibily expensive HDTV productions.

On the transmission front, various groups are studying ways of combining the audio signals in a compatible way, such that the scarce resource of available capacity is not squandered.

7. ACKNOWLEDGEMENTS

The author would like to acknowledge the valuable discussions and programme making assistance that has been afforded to him by his colleagues in BBC Television. The author would also like to thank the Director of Engineering of the BBC for permission to publish this paper.

The results shown in figs 6 and 7 are reproduced by kind permission of NHK.

8. REFERENCES

[1] N.E.Tanton, M.A.Stone. HDTV displays: subjective effects of scanning standards and domestic picture sizes. Proceedings of 1988 International Broadcasting Convention, pp 204-211.

[2] N.E.Tanton, M.A.Stone. HDTV displays: subjective effects of scanning standards and domestic picture sizes. BBC Research Department Report RD 1989/9

[3] P.A.Ratliff. Properties of hearing related to quadraphonic reproduction. BBC Research Department Report RD 1974/38.

[4] Federal Republic of Germany. Suitable number of sound channels to accompany wideband HDTV. CCIR Document Number 10/267-E

[5] S.Komiyama. Subjective evaluation of angular displacement between picture and sound directions for HDTV sound systems. J.Audio.Eng.Soc., Vol 37, No. 4, April 1989, pp 210-214.

[6] K.Ohgushi, et al. Subjective evaluation of multi-channel stereophony for HDTV. Proceedings of the 81st Audio Engineering Society Convention, November 1986. Preprint No. 2363.

[7] Specification of the systems of the MAC/Packet family. European Broadcasting Union, Technical Report No. 3258-E, October 1986.

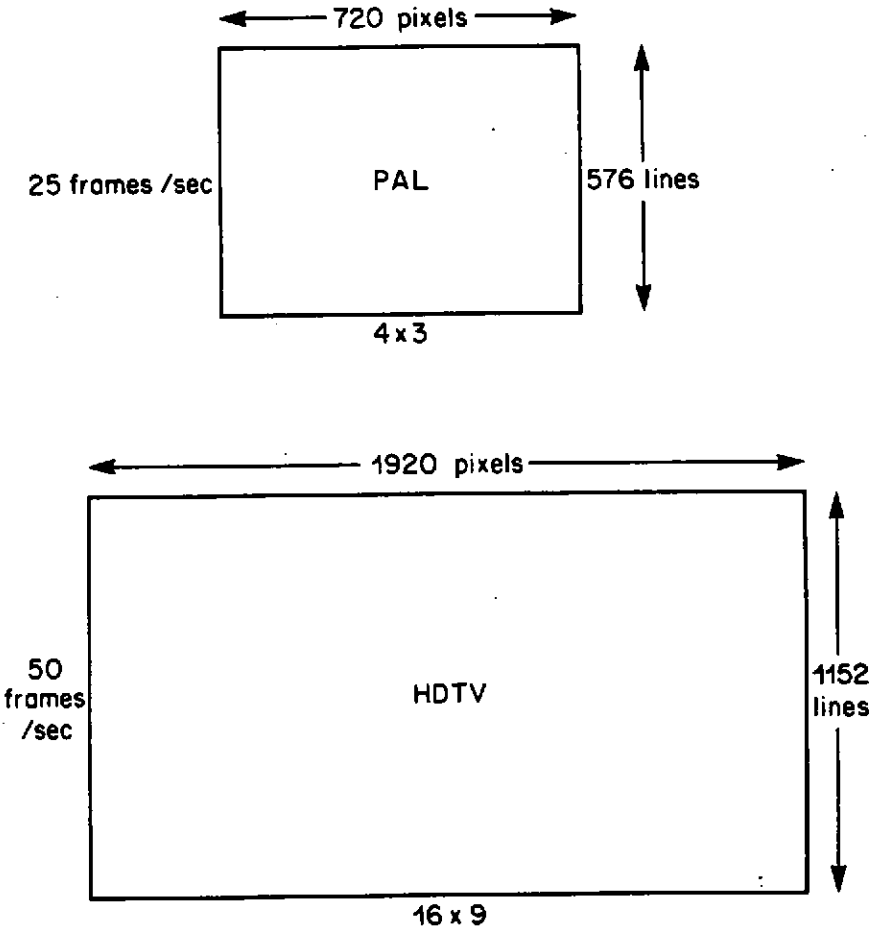


Fig 1 The shape of things to come.

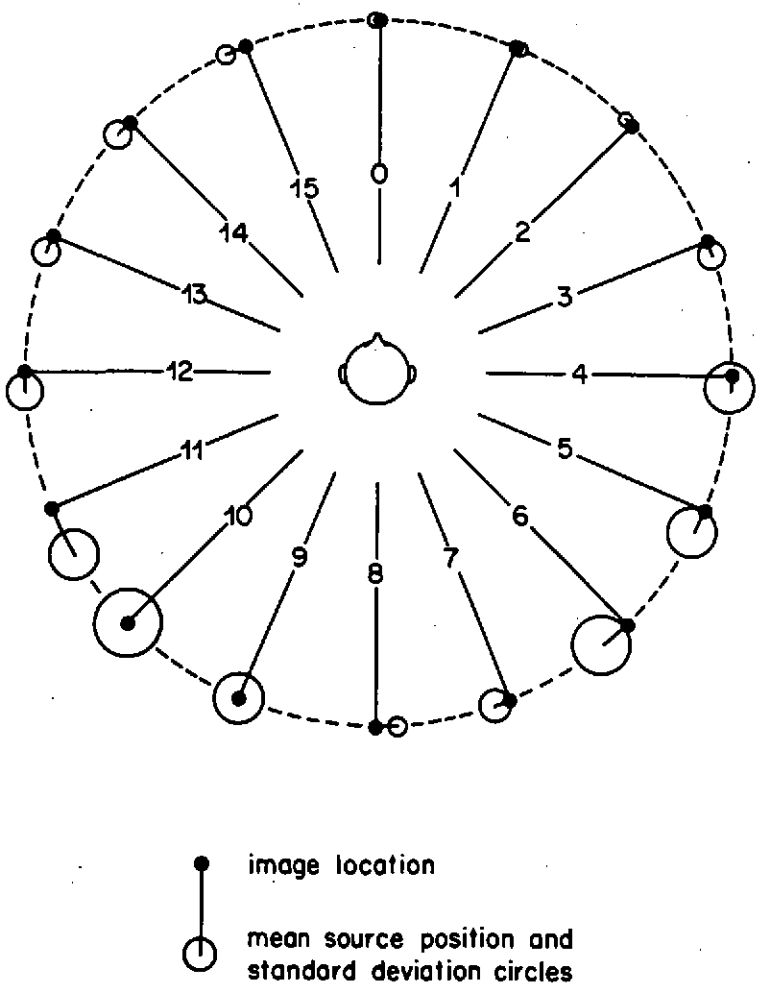


Fig 2 Absolute sound localisation in the free-field room.

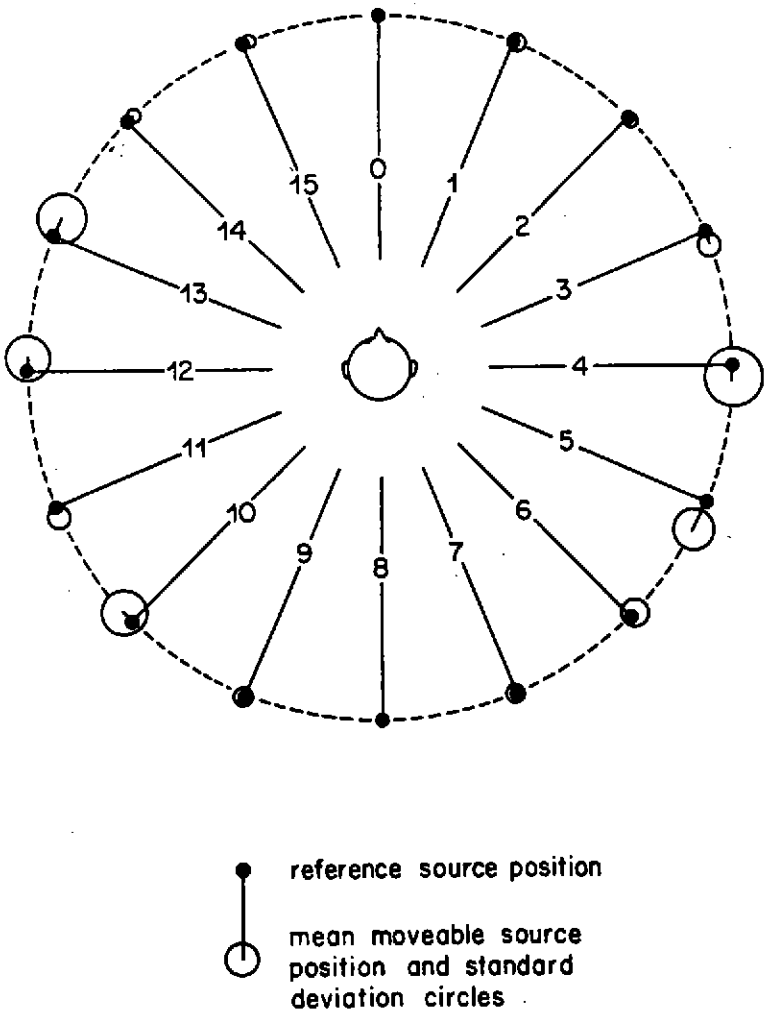
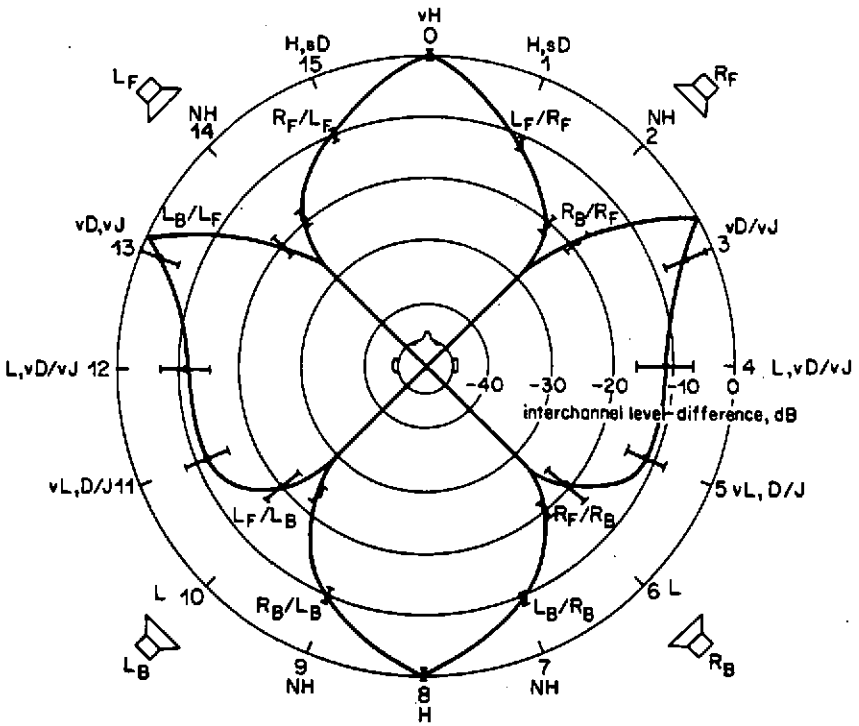


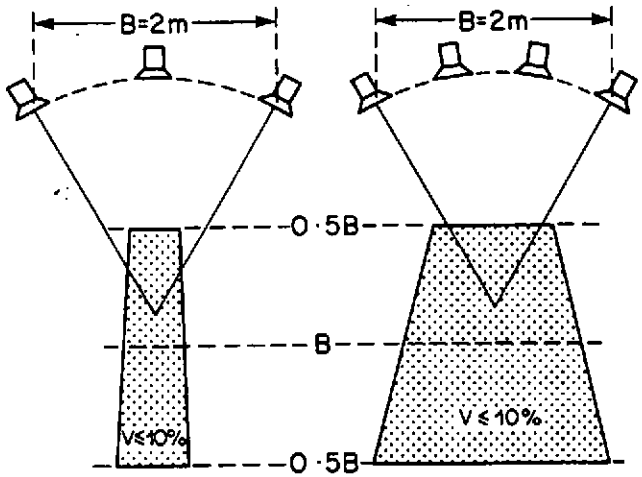
Fig 3 Relative sound localisation in the free-field room.



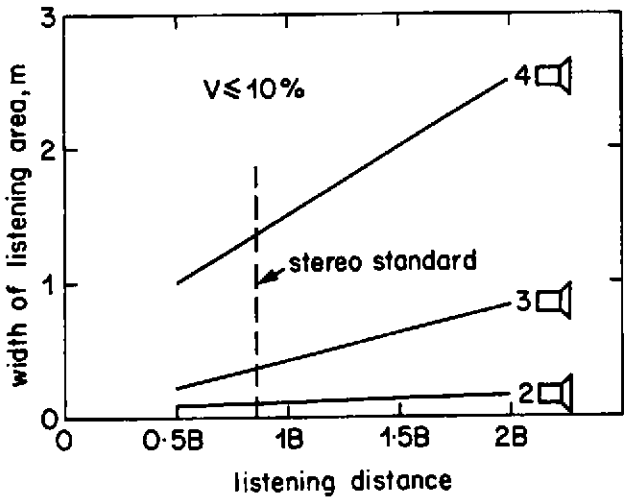
I experimental result showing
I standard deviation

D diffuse
H high
J jumpy
L low
NH normal height
s slightly
v very

Fig 4 Interchannel level-difference versus image location for adjacent pairs of loudspeakers, in a free-field room with a reflecting floor.

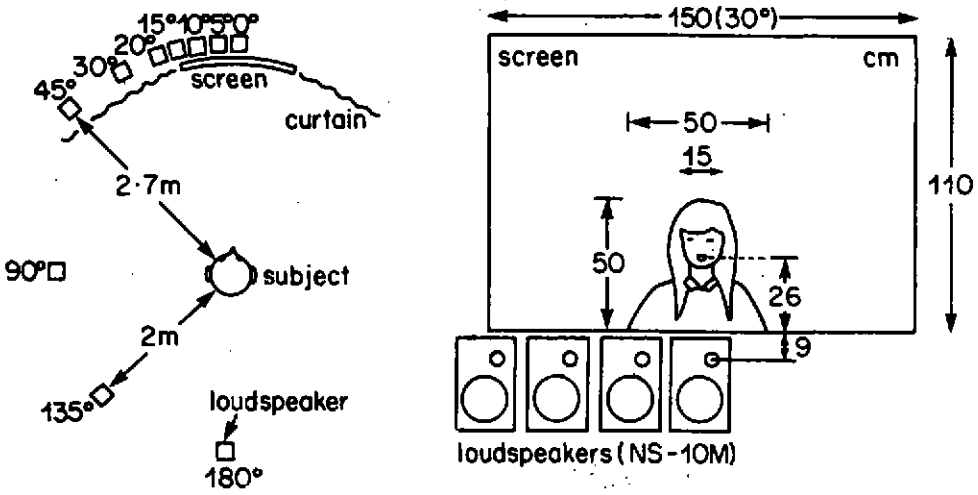


a. pictoral

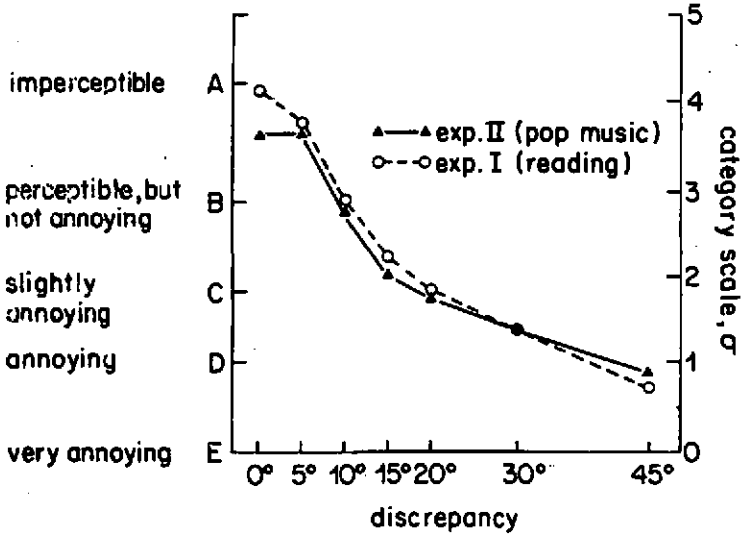


b. graphical

Fig 5 Listening areas for multichannel stereophony.

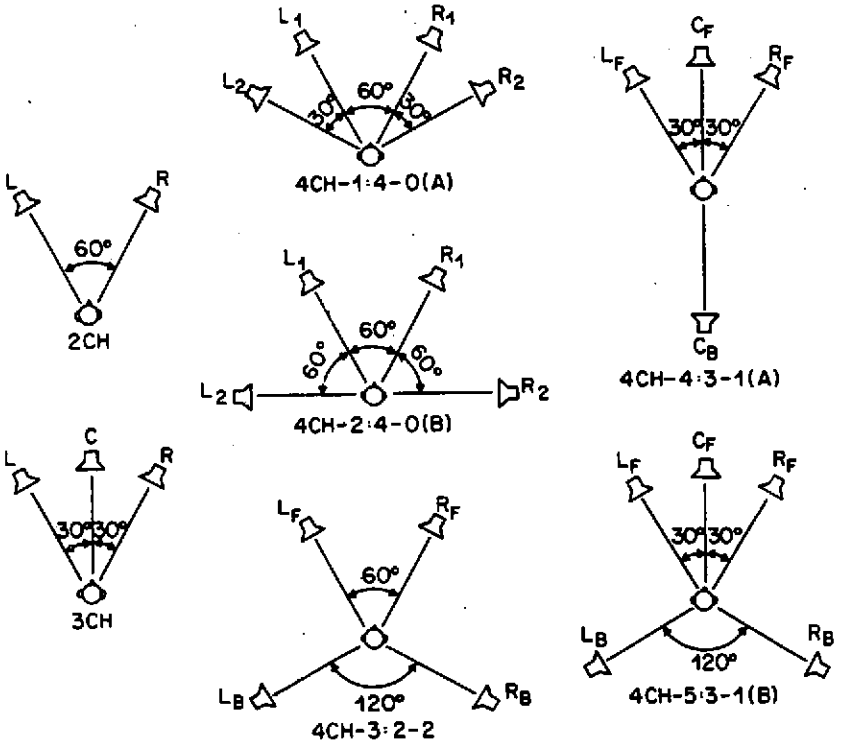


a. experimental setup

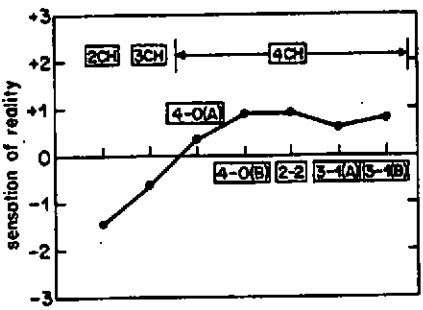


b. results

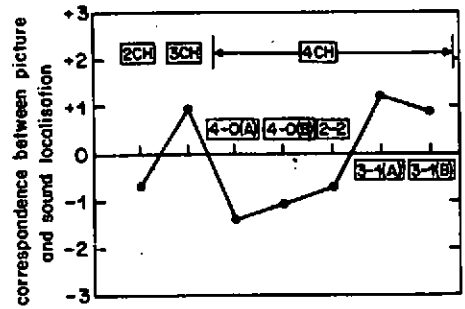
Fig 6 Response to errors in sound localisation.



a. loudspeaker arrangements

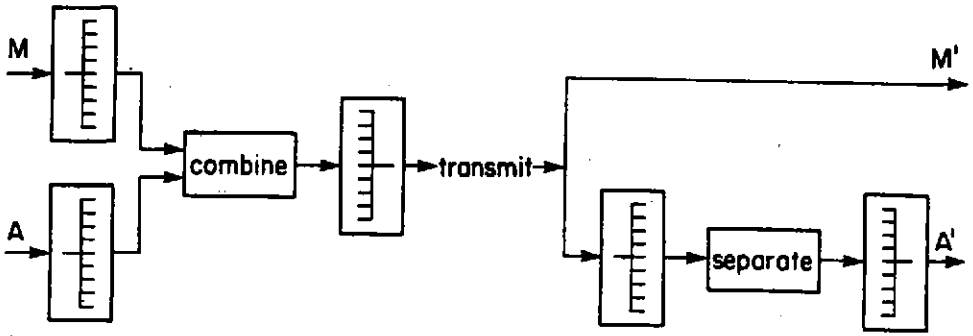


b. sensation of reality



c. correspondence between picture and sound localisation when an observer is seated 70cm off axis.

Fig 7 Assessment of multichannel sound systems.



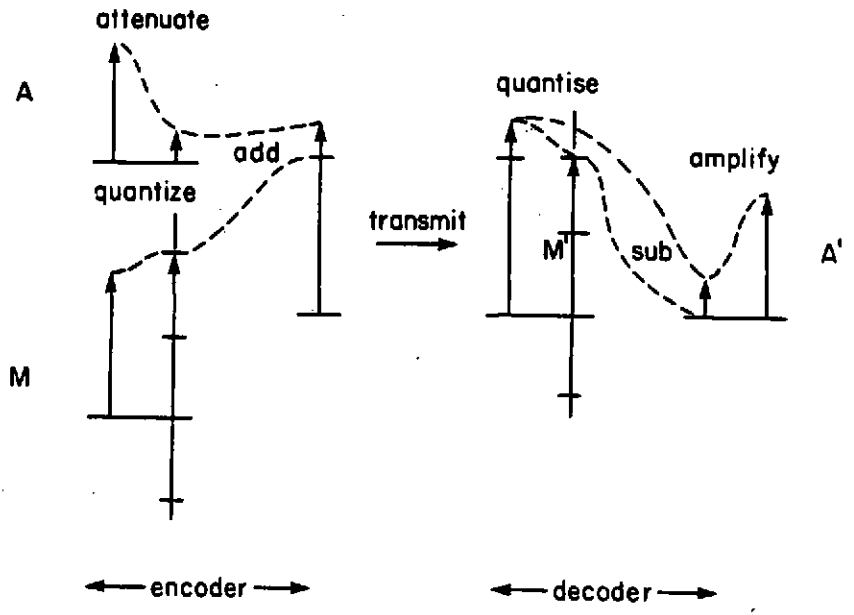


Fig 8 Surround sound encoding using 'Hidden Channel' techniques.

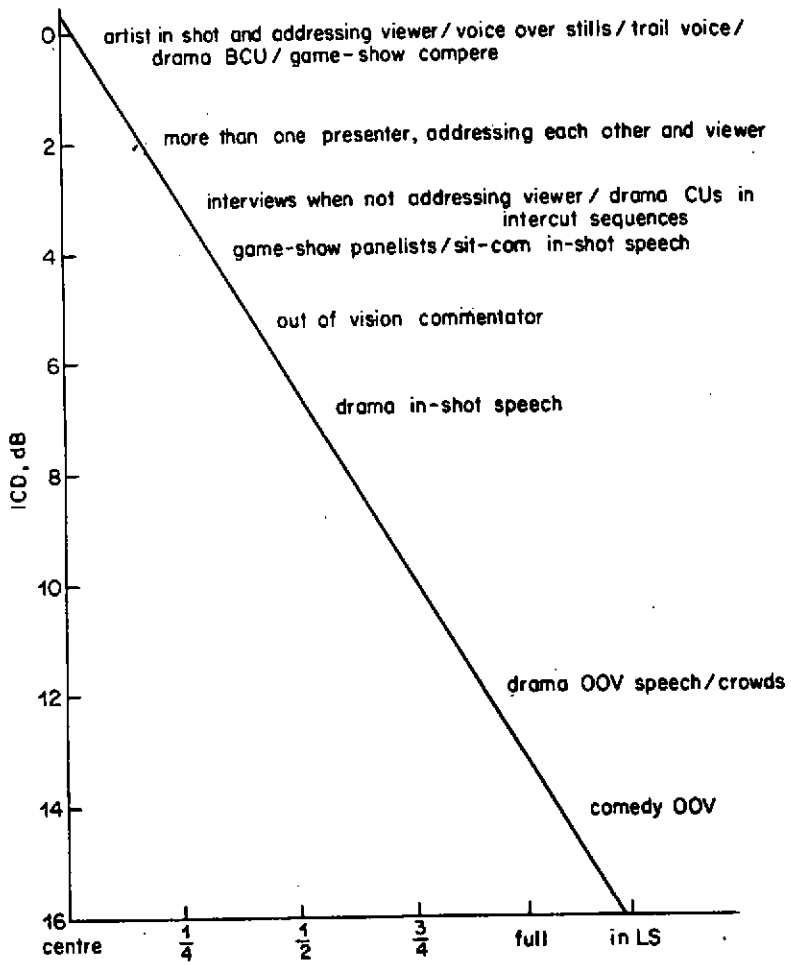


Fig 9 Distribution of images: suggested limits of interchannel difference.