Proceedings of the Institute of Acoustics

EXPERIMENTS IN MULTICHANNEL SOUND

D. J. Meares,

BBC Research Department, Kingswood Warren, Tadworth, Surrey, KT20 6NP.

1. INTRODUCTION

The international debate on HDTV continues, particularly on the question of transmission formats for the picture. Whether or not, in the long run, a single system will become the standard, the picture will need to be supported by a matching sound system, i.e. one which complements the visual presentation.

In a previous paper [1], the author presented the background information and the then current studies on HDTV sound systems. This paper is in the form of an up-date on that and presents the current consensus on the likely service improvements beyond simple two channel stereo for an HDTV sound system. The international discussions and standards making process are discussed, as are the practical experiments both completed and still continuing.

2. IMPROVED SERVICES

As stated before, one of the most visually obvious changes from conventional television to the proposed HDTV format is the aspect ratio. This will change from the current 4:3 to the wider 16:9 format and will almost certainly be accompanied by larger screen presentation. On the other hand limited tests to date [2] have indicated that there is an upper limit to the size of picture that will be acceptable in the domestic environment and it is likely that this will be set, by consumer choice, to somewhere around the 1 metre diagonal. Perhaps less immediately obvious to many consumers will be the improvement in picture portrayal. The HDTV standards have the
ability to portray a doubling of information density both horizontally and vertically, and with upconversion of the pictures an improvement also in motion portrayal.

If the pictorial presentation is that much improved it is necessary to consider how the sound can be optimised to support this form of presentation. Indeed this question is currently being considered

by a number of international committees all of whom are coming to much the same conclusion as far as a list of service improvements is concerned. (The techniques by which these service improvements may be achieved is however, not yet universally agreed). First and foremost there is a need to improve the presentation of sound images in front of the listener. The reasoning behind this being that any mismatch between visual and aural users is likely to be that much more disturbing with a higher definition picture. There is a need therefore to stabilise the sound presentation irrespective of the listener's position relative to the picture and loudspeakers. The second most important feature for improving sound is agreed to be the provision of surround sound, mainly as a result of its introduction in the cinema industry, where surround sound has been available for a number of years. Additional sound features include the provision of multiple language options, particularly for programmes such as sports, and new services to support better the people with hearing or sight impairments. The latter services are commonly seen to be the provision of clean dialogue for the hearing impaired and commentary for the blind. The aim in both cases being to enable the hearing and sight impaired consumer to share the same programme selection as the rest of the family unit.

3. SOUND PRESENTATION

As already stated the methods of achieving these new services is currently still open to much debate, although the techniques for providing a larger listening area giving the common aural experience are subject to a great deal of consensus, the EBU [3] in discussing this matter have defined the listening area according to figure 1. Essentially they have defined the need for representative listening conditions for programme making and subjective appraisal in such a way that a minimum of three people can gain near identical reproduction of the sound in a domestic size environment. However their recommendation goes further to the extent that a significantly larger group of people with the same listening arrangement can still gain a substantially improved sound presentation compared to that for simple two channel stereophony. To this end the EBU are recommending (see fig. 2) the use of three loudspeakers to support the presentation of in-vision sounds plus a pair of loudspeakers to reproduce the additional surround information. The specific positions of the frontal loudspeakers is already defined, whilst the positions of the side/rear loudspeakers is still open to debate and experimentation.

EXPERIMENTS IN MULTICHANNEL SOUND

Whilst the five loudspeaker arrangement, or more specifically the 3/2 arrangement, just described is now universally accepted as the means of providing surround sound reproduction in the home there is still debate over the appropriate sound system for the cinema. Proposals for the cinema extend from the 3/2 proposal up to a 4/4 proposal which is claimed to give significantly better sound presentations. However, in this context it is necessary to look to current trends in the cinema industry and apart from a few special exceptions this author believes that extension beyond a 3/2 proposal is unlikely. This conclusion is in part supported by the current developments of digital sound systems for the cinema [4,5] which, apart from a sub-woofer channel, provide only five separate (discrete) sound channels.

But having decided on the number of loudspeakers in a reproduction system one has to ask what the aim of that sound system should be in terms of the sound experience for the listener. For instance, is reality a key aim or not? In real life we are exposed to sounds coming from all directions and some suggest that this should be the aim of the HDTV sound system. On the other hand in real life one has an option, when hearing a sound from a particular direction, to turn and face it. This option is removed with HDTV systems, even with large screens, simply because a significant turn of the head will direct the eyes outside the limits of the picture giving no visual support to the sound image. It is necessary to tone down acoustic reality, ensuring that the consumers attention is not drawn, at least for significant periods of time, away from the limits of the video screen. It has therefore been accepted by the international working parties that the frontal sound stage will be used to generate sounds with a specific directionality, and that the side and rear loudspeakers will be restricted to ambience information plus sound effects where specific directionality is not particularly important.

In considering the sound system, it is also important to bear in mind the way the television medium is already being used in ways that conflict with the spatial presentation of sound. The specific problem here is that of reverse angle shots, i.e. where significantly different camera angles are used within the same element of a programme giving rise to a dramatically different visual perspective on the screen. It is the author's opinion that this causes no psychological problems simply because one can look at reflected presentations of the real world, in a mirror or in water. On the other hand in the case of the aural senses, the environment only changes at the rate at which the individual moves through that environment: there are no sudden reversals of the acoustic perspective without other tactile senses supporting such

a change. In the case of the television medium, if one was to reverse the acoustic perspective in synchronism with the visual perspective, the sudden change of sound presentation with no tactile support is found to be most disturbing. It is therefore necessary when mixing the sound presentation, either for existing stereophony or for multichannel sound, to take a stable aural perspective and maintain that throughout a particular scene. This problem is heightened in the case of high definition presentation and it has been found necessary to constrain significant changes in the visual presentation.

4. PROGRAMME EXPERIMENTS

In order to explore the possibilities of multichannel sound the BBC and other broadcasters have been carrying out experimental programme productions with different types of material. Experiments have included concerts, ceremonial events, drama, sports, pop music and documentaries, and a few examples will now be explored in order to illustrate some of the findings of that work [6].

One of the events that the BBC has exploited on a number of occasions now, has been the Wimbledon Lawn Tennis Championships in London. In this particular case the aim of the multichannel sound presentation was to attempt to reproduce the reality of a fixed seat in the audience at one end of the Centre Court. The cameras themselves were placed looking along the centre line of Centre Court from two different heights, which provided an important supportive role in the context of a fixed aural perspective. The main sound pick-up (see fig. 3) was from a sound field microphone placed above the end wall of the tennis court. This was supported by a gun microphone on the umpire's chair picking up sound effects from the far half of the tennis court, a pair of microphones high in the roof of the stand picking up the more distant elements of the crowd sounds and spot microphones on the umpire and commentator. The studio mix of these sources used the front channels of the sound field microphone to define the stereo presentation of the court action, with the gun microphone on the umpire's chair filling in the far court effects in the centre of that stereo stage. The rear channels of the sound field microphone and the stand roof microphones were fed to the surround channels of the mix with the spot microphones on umpire and commentator being mixed half right and half left in the stereo stage. The subjective effect achieved was very pleasant giving a real sensation of being present at the event. The supportive role of the stable visual perspective was found to be of real benefit.

Not everything however, goes according to plan and in the case of an early drama experiment the location sound was the subject of very serious difficulties. The basic concept of one of the scenes in the drama was of a boy-meets-girl type of situation in a park near to the BBC's old transmitter at Alexandra Palace. In principle the sound pick-up arrangement need only have been a simple combination of sound field microphone for the ambience and spot microphones on the actors. However when it came to the recording date two serious problems were encountered. In the first instance the date of the recording coincided with severe storm weather, and though the recording went ahead in between showers there was a constant gale blowing, giving rise to continuous howl effects from the wind's passage through the transmitter tower. A further practical difficulty with the surround sound recording on location, was that half way through one of the takes the Public Relations Manager of one of the actors crept quietly up behind the sound field microphone and fired off a number of shots on his single lens reflex camera. Each time he took a photograph there was an inevitable punctuation on the sound field recording and the take was therefore unusable. The message from this is clear: there is no "off-mic" location with surround sound recordings and thus much more care is needed to avoid either many more retakes or the need for dialogue replacement.

Most of the recordings are nowhere near as problematic as the drama recording just described and a number of general rules, particularly on the acceptable distribution of sounds have been devised. Figure 4 shows the BBC's current guide lines on how voiced sounds should be distributed within the frontal sound stage. It is inevitable that some of the in-vision sounds are so intimately connected with the picture that they have to be placed precisely in the centre of the sound stage; the obvious example of this is the in-vision news reader. However, it is clearly inappropriate to conclude that dialogue should be restricted entirely to this central location and experience has shown that voiced sounds can be usefully spread across the front sound stage. The benefit of this is that it makes for a much more interesting sound presentation, but it brings with it a need to constrain the use of reverse camera angles as already discussed.

Sound sources other than the voiced components need not be restricted to the front sound stage, figure 5 shows the distribution to all locations of ambience and transient sound effects. In the case of ambience, as in real life, this should be distributed evenly amongst the loudspeakers even though the sound

system itself may not reproduce all directions with the same ease. Sound effects, as long as they are of short duration, can similarly be located outside the front sound stage.

One of the additional problems, which is introduced by the provision of a multichannel service, is the need to ensure during the sound mixing that the signals make sense to all groups of listeners. With stereo, there was an obvious need to monitor the sound balancing in both stereo and mono. This will continue into the HDTV mixing of, say, a five channel programme, with provision being made for checking the reproduction in four channel and three channel formats as well as stereo and mono. At present this is very difficult to provide within conventional stereo sound mixing desks and thus external monitor switching needs to be provided. However, it is essential that such monitor switching obeys the same laws as those implemented in the domestic receivers. At the present time therefore, whilst the mechanics of such a monitoring arrangement can and have been designed, the precise interchannel relationships of the signals have not yet been agreed. This is just one of the aspects of compatibility, which will be discussed later.

5. NUMBER OF SOUND CHANNELS

Though it has been agreed that the maximum number of loudspeakers for the domestic surround sound system will be five, the number of transmission channels required to convey the signals through the broadcasting chain is still open to debate. Whilst it is universally agreed that there is no point in having more than five channels, there are strong arguments in favour of providing fewer than five channels.

In the earlier period of this debate there had been suggestions of up to four channels as well as loudspeakers at the front of the listening area, on the basis that laboratory subjective tests had indicated more accurate sound location from a four channel presentation than from a three channel presentation. (No-one, in fact, had disputed the need for more than two frontal loudspeakers.) However, when it came to subjective appraisal of real multi-element programmes rather than single sound images, it was found [7] that a properly mixed three channel presentation at the front was virtually as good as a four channel presentation. A similar analysis of the number of surround channels indicated that two channels for the surround information were almost as good as four and thus a 3/2 system was derived.

Further tests have also been carried out to compare two channel surround against single channel surround and of particular interest are those reported by the IRT [8]. Figure 6 shows these results. In this particular test the number of loudspeakers was always four but the number of transmission channels was selected from 1, 2 or 4. It can be seen that, on both ambience information and crowd effects, both two and four channel transmission were judged to be significantly better than single channel transmission and, as already stated, the difference between two channel and four channel was minimal. It is interesting to speculate why such clarity of difference was not shown by the music itself, but it is believed that as the main and loudest sounds during the music were coming from the front of the reproduction system, the significance of the surround channels was severely diluted.

As mentioned earlier, discussions continue on where the side/rear loudspeakers should be placed for optimum reproduction in the domestic environment. A recent experiment conducted by the BBC specifically addressed this question. The layout of the test room is shown in figure 7: three frontal loudspeakers were supplemented by three pairs of side/rear loudspeakers, any one of which could be auditioned at a time from one of two seats. The test used a wide range of programme material, mixed for 3/2 presentation. The subjects were asked to decide which of the presentations of surround sound was preferable and why. The general consensus was that the side loudspeakers (+/- 90 deg.) were not liked because small head movements could cause substantial changes in the sound impression, particularly in the side seat: there was also no sense of envelopment in the sound field. The rear loudspeakers (+/- 135 deg.) were somewhat disliked because of the apparent dissociation between the front and back parts of the sound event (a hole at the sides was one description given). The preferred loudspeaker placement was therefore the middle one, i.e. +/- 120 deg..

These same tests also briefly examined the comparison between single channel and two channel surround (i.e. 3/1 versus 3/2) using wo surround loudspeakers on some of the programme items. Even using decorrelating delays between the rear loudspeaker feeds for the 3/1 presentation, the subjects preferred the 3/2 presentation.

Thus, all evidence currently points to the 3/2 presentation utilising five subjectively separate channels in a balance between subjective benefit and system complexity.

6. COMPATIBILITY

In exploiting any multichannel sound system it is important that a broadcaster serves not only the requirements of the new multichannel listeners but that the broadcaster also continues to serve the existing audience of monophonic and stereophonic listeners. This is seen to be even more important when one considers that even in the longer term this latter section of the audience may remain in the majority. Clearly important also to the broadcaster is that much of the programme material in the early stages of a new service will come from the archives which, in the case of the postulated multichannel sound service, may be of mono or stereo presentation. Any multichannel sound system has therefore to include significant elements of intercompatibility in the way it can be used both to ensure that multiple audiences can receive a useful service from each broadcast and to ensure that the archives of existing programme material can continue to be exploited.

There are in fact two ways of providing such intercompatibility in a multichannel sound system. The first is an extension of the A/B - M/S transmission of stereo FM radio and constitutes a group of reversible encoding equations. These reformat the source loudspeaker signals into a group of sum and difference derived transmission signals with inverse equations being applied in the receivers: this technique is called compatibility matrixing. The essence of such transmission coding is to ensure that the first two signals provide a service for the stereo listener, the first three signals provide a service for the three channel listener etc. There are in fact many bases on which such transmission coding equations can be derived but the interested reader is directed towards information in references 9 and 10.

In the context of HDTV the format for transmitting multichannel sound is likely to be digital, and, as with all digital systems, there is a need to reduce the data capacity as much as possible to provide space for other services. There is unfortunately a potential conflict between the needs for bit-rate reduction for transmission [11] and the needs of compatibility coding for transmission. Many of the bit-rate reduced signals are not capable of accommodating significant post coding signal processes such as that required with compatibility matrixing. In essence, the practice of subtracting signals in the receiver (in order to provide loudspeaker feeds) could expose some of the otherwise inaudible artifacts of the bit-rate reduction process. Thus a second group of proposals has been tabled which avoid the need for generating difference signals. In these proposals the loudspeaker

feeds for a five channel presentation would be transmitted and, for instance, a stereo receiver would need to pick up all five signals in order to generate the necessary two loudspeaker signals by simple addition.

The final aspect of compatibility that should be addressed when considering a multichannel sound system is that of spatial compatibility within the monitoring environment. Inevitably the desire to reproduce surround sound in a monitoring area is going to require larger spaces for these activities, if only to accommodate the extra loudspeakers and the necessary distances between each of them and the Sound Supervisor. Such additional space is unlikely to be available in the short term in most sound monitoring areas, particulary on outside broadcast. Precisely how this problem will be resolved is still open to debate.

The solution in the case of one of the BBC's monitoring areas is shown in figure 8, which represents one of the BBC's post-production (Sypher) areas. Though this was itself a relatively large room, a considerable amount of technical equipment had to be moved around to accommodate the additional loudspeakers and surround sound monitoring. Even with such "furniture removal" there was still only space for one person at a time to listen to the reproduced sound correctly, although it should be clearly noted that better off-centre sound impressions were created with the multichannel system than with conventional two channel stereophony.

7. ADDITIONAL SOUND SERVICES

The provision of such services as multilanguage sound presentation is a non-trivial matter. Whilst a solution is relatively easy for sports programmes, the solution for other programme types is of necessity much more complex. The peculiar attribute of sports programmes is that the commentator is not only normally out of vision, but his voice would not be expected to fit in with the perspective of something like a sports stadium. Thus to provide surround sound for a sports programme, the broadcaster would need to mix a five channel presentation of the sound effects and ambience of the event and a supporting, single-channel, clean feed for each of the languages. In contrast for a programme like drama, the voice of necessity has to be part of the spatial presentation of the environment being simulated; the reverberation of a room containing action must respond to each of the voices that are supposed to be in the room. Thus for this type of programme each

language in surround sound would need a full set of sound channels. This requirement alone may rule out the provision of multiple language services for anything other than sports programmes.

Additional services for groups such as the hearing or sight impaired, must be considered at an early stage in the design of multichannel sound systems. Clearly it would be inappropriate to try to provide them with a special surround sound mix, but broadcasters would be unwise to ignore their specific needs if relatively simple steps can be taken to fulfil them. The proposal for the hard of hearing would be the provision of a single clean dialogue channel, which is certainly technically feasible as far as the transmission system is concerned. However, two aspects of such a service need to be borne in mind. Firstly, there may be recording venues where a clean feed of the dialogue just does not exist: in some sports stadia, even using lip microphones, there is a significant amount of crowd noise on the commentator's microphone, and thus a clean feed does not really exist. The other point to be considered is that in order to make such a service viable to the hard of hearing consumer, the system designers ought to avoid the need for special receiving equipment. Thus whilst the bit-rate requirements for a dedicated clean dialogue channel could be shown to be very small, it would make much more sense from the consumer's point of view to define the hard of hearing requirement is merely a special language option. Thus the standard hardware could be easily programmed by the hard of hearing to pick up the clean dialogue "language" rather than English, French, German etc.

8. CONCLUSIONS

It has been shown that a great deal of consensus already exists on the need for improved sound services with HDTV. Whilst many of the technical options have already been agreed there are still a large number of details, particularly relating to transmission coding, for which agreement is awaited.

The BBC's programme experiments have helped a great deal in deciding just how various events should be reproduced to create a worthwhile sound presentation. These have also helped to demonstrate some of the less essential features in some proposals. All broadcasters will need to continue to use existing programmes and to supply the needs of a number of audiences and thus a system demonstrating compatibility has been shown to be necessary. Two clearly different methods of achieving that compatibility have been described.

In debating multichannel sound systems the technology of the development should not be allowed to take overriding importance. Ultimately, it will be the programmes which will sell the system rather than the technology and even then the technology must not introduce difficulties for the consumer. A multichannel sound system will only be acceptable if it is both easy to use and non disruptive to the domestic environment.

9. ACKNOWLEDGEMENTS

The author would like to thank his BBC colleagues for their contributions to this work and the Director of Engineering for permission to publish this paper.

The results shown in Fig. 6 are reproduced with permission of the IRT.

10. REFERENCES

[1] D.J. MEARES, 'Sound Systems for High Definition Television. Proceedings of the joint IOA AES Conference 'Reproduced Sound 5'. Windermere, November 1989. Reproduced in Acoustics Bulletin, January 1990. pp. 6 to 11. (1989)

[2] 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 to 211 and BBC Research Department Report RD 1989/9. (1988)

[3] European Broadcasting Union, 1989. EBU Requirements for an HDTV Sound System. CCIR Document No. IWP 10/12-02-E.

[4] H.J. FLEMMING, 1991. 'Cinema Digital Sound: The Technology.' Proceedings of the BKSTS 12th International Conference. London. 9-12 July 1991.

[5] I. ALLEN, 1991. 'Matching the Sound to the Picture.' Proceedings of the AES 9th International Conference 'Television Sound, Today and Tomorrow'. Detroit, Mitchigan. 1-3 February 1991. pp. 117 to 186.

[6] D.J. MEARES, 1991. 'High Definition Sound for HDTV'. Proceedings of the AES/SMPTE 9th Int. Conf. 'Television sound today and tomorrow', Detroit, Mitchigan. pp. 187 to 215. 1-3 February 1991.

[7] D.J. MEARES, 1991. 'Developments in Multichannel Sound for HDTV.' Proceedings of the 4th Int. HDTV Workshop, Turin, 4 September 1991.

[8] G. THEILE, 1991. 'HDTV Sound Systems: How Many Channels?' Proceedings of the AES/SMPTE 9th Int. Conf. 'Television sound today and tomorrow', Detroit, Mitchigan. pp. 217 to 232. 1-3 February 1991.

[9] D.J. MEARES, 1990. 'HDTV Baseband Compatibility Matrixing.' EBU Document No V3/HTS-001. November 1990.

[10] M.A. GERZON, 1991. 'Optimal Reproduction Matrices for Multispeaker Stereo. To be published in the Proceedings of the 91st AES Convention. October 1991.

[11] G. THEILE, 1988. 'Low Bit Rate Coding of High-Quality Audio Signals - An Introduction to the MASCAM System.' EBU Review - Technical. No. 230, 1988. pp. 158 to 181.

EXPERIMENTS IN MULTICHANNEL SOUND

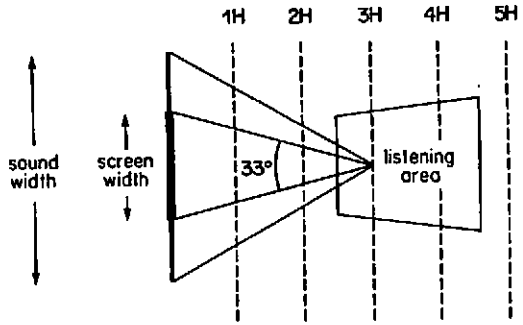


Fig 1. EBU recommendation on listening areas

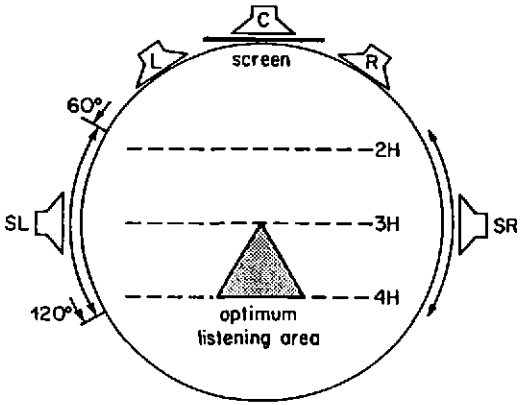
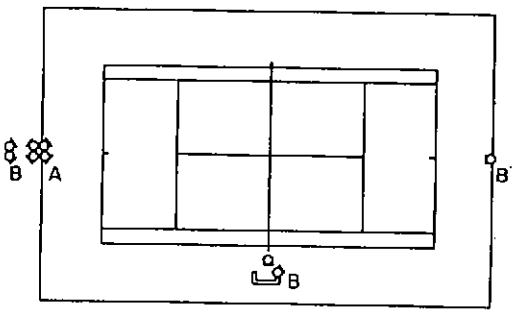


Fig 2. EBU recommendation on multiple loudspeaker layout

EXPERIMENTS IN MULTICHANNEL SOUND



A = sound field microphone
B = gun microphone, Sennheiser 416 and 816

Fig 3. Microphone arrangement for Wimbledon recording

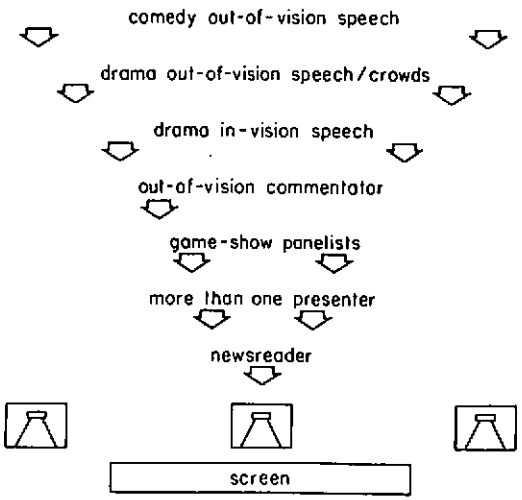


Fig 4. Distribution of frontal voiced sounds

EXPERIMENTS IN MULTICHANNEL SOUND

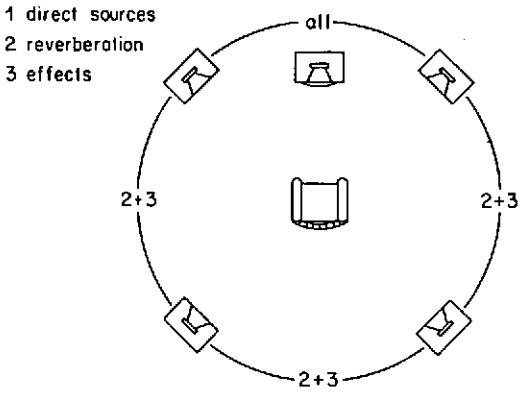


Fig 5. Preferred location of sound sources

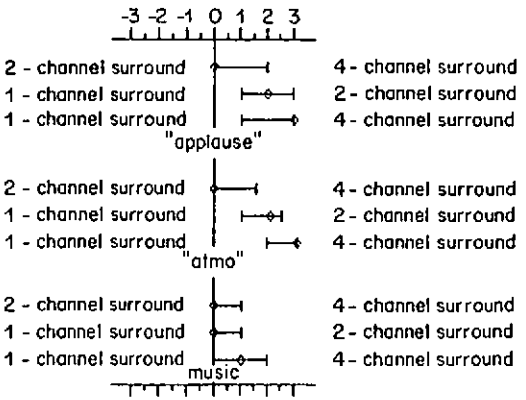


Fig 6. Preference for number of channels for surround sounds (IRT)

◇ median
⊢—⊣ interquartile range

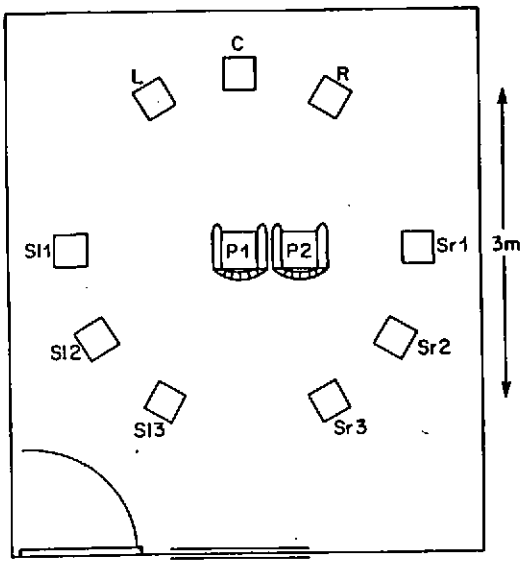


Fig 7. Arrangement for listening tests on rear/side loudspeaker placement

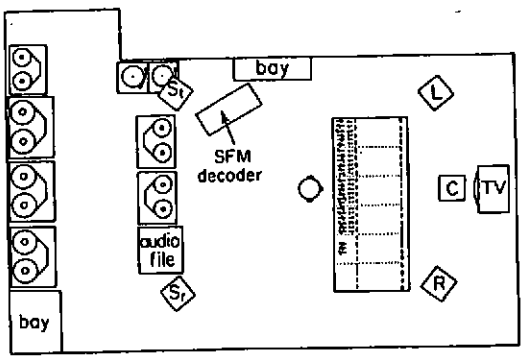


Fig 8. Sypher control room layout