Synthesised Multichannel Signal Levels versus the M-S Rations of 2-Channel Programme Items 4653 (P6-1)

Francis Rumsey
University of Surrey Guildford UK

Presented at the 104th Convention 1998 May 16-19 Amsterdam

AN AUDIO ENGINEERING SOCIETY PREPRINT

Synthesised multichannel signal levels versus the M-S ratios of two-channel programme items

Francis Rumsey, Music Department,
University of Surrey, Guildford, UK

Abstract

Multichannel surround synthesis algorithms derive a centre channel and rear channels from a two-channel source, so that two-channel programmes can be replayed over five loudspeakers. Typical two-channel programmes were analysed to determine their sum-and-difference (MS) ratios, and a test signal with controllable MS ratio was used to examine the static behaviour of selected 2–5 processing algorithms.

0 Introduction

The work described in this paper was carried out as part of EUREKA Project 1653, called MEDUSA, which aims to study methods by which consumer sound reproduction may be enhanced by the use of multichannel techniques. The project considers the whole signal chain from programme origination to consumer reproduction, and is studying ways in which systems can be optimised so as to deliver the best subjective results under a range of different conditions. This work was undertaken at the BBC Research and Development Department in the UK between July and December 1997.

The experiments described here relate to what will be termed 'upmix processors' – that is algorithms which process a two-channel signal source for reproduction over multiple channels. Normally this amounts to either 2–5 or 2–4 channel conversion, designed for extracting a centre loudspeaker signal and rear surround loudspeaker signals from the information present in the two channel signal, although some devices are capable of driving more than five loudspeakers. It is argued that there may be applications in which discretely recorded or mixed multichannel surround material is either not available or not practical, and in such cases upmix processors might be employed to deliver an enhanced spatial impression compared with that available from two-channel stereo. Such processors could be used either at the consumer reproduction end of the signal chain, in which case programme material would be distributed or broadcasted in a two-channel form and processed by the consumer, or employed during the post-production of programmes in order to be able to deliver already-processed material to the consumer in a multichannel form. The main difference is clearly related to where control over the sound reproduction quality lies.

A parallel paper by this author [1] describes a series of subjective tests that were carried out on a limited selection of algorithms. Those tests were intended to assess the effect of upmixing on the reproduced attributes of 'front image' and 'spatial impression', using a range of representative programme material, and to determine listener preference.

In this paper, a series of measurements is described which simply attempts to provide information about the steady-state behaviour of a number of currently available algorithms in their default states, demonstrating that the signal levels fed to centre and surround channels when upmixing two-channel material varies very considerably between algorithms, potentially leading to widely differing sonic effects. The result of measurements on a range of different styles of two-channel programme material gives information about their sum-to-difference signal ratios (MS ratios) in octave bands across the audio spectrum – a factor which was found to relate closely to the centre and surround signal levels after upmixing.

A key reason for analysing the MS differences of representative two-channel programme items was to enable the author to devise an artificial test signal that had an MS ratio similar to that of the programme items to be assessed in subjective tests, in order that SPL differences between two- and five-channel versions of the same programme could be minimised. More programme items and algorithms were measured than were included in the subsequent subjective tests, but the complete set of measurement data is included in this paper. The algorithms concerned are identified directly in this paper.

1 MS ratios of programme extracts

Using an octave band real-time spectrum analyser, measurements were made of the average (L_{eq}) difference between the sum (M) and difference (S) components of selected two-channel programme items, using extract lengths of between 2'30" and 3'00". Although naturally the instantaneous differences between M and S components varied, the difference averaged over the extract duration was considered to offer a useful indication of the nature of the programme material.

The extracts were selected to represent a range of typical styles of recorded and broadcast programme material, such as classical music recorded using basic coincident and spaced microphone techniques, multitrack pop music with artificial effects, electronic music, television sport, 'natural' and 'sci-fi' radio drama. It is acknowledged that these items only represent a sample of current recording and mixing techniques, and that every recording would have unique characteristics, but they provide a starting point for comparison and illustration. More importantly they contained the extracts which were to be used in later subjective tests, as described in [1].

Figure 1 (a)–(f) shows the average differences between M and S level for each of the programme extracts (except the sport), both in octave bands and overall (linear weighted L_{eq}). An M component considerably higher than S is indicative of a signal with a strong monophonic or centre component, whereas a small difference between M and S is indicative of a 'wide' stereo balance or one with considerable out-of-phase components, such as reverberation elements and other spatial effects. A single source present in only the left or right channel of a stereo pair results in equal M and S components. Rare cases in which the S component is seen to be higher than the M component are usually indicative of excessive stereo width or strong out-of-phase components in the stereo pair. In general, the higher the level of the S component the greater the amount of spatial information present in the signal, although excessive levels of S result in severe phasiness. Centrally panned mono material, on the other hand, consists only of an M component.

Table 1 shows a summary of the average MS differences for each programme item.

Table 1 Average MS differences (dB) by programme item

Item	*M–S (dB)*
Renaissance church music (Decca tree of spaced omnis)	4
Orchestra (coincident pair of figure eights)	2
Synthesised music (Jean-Michel Jarre)	1
Sci-fi radio drama (Hitch-hikers Guide to the Galaxy)	6.5
Rock – multitrack panned with effects (Alan Parsons)	3.5
Traditional radio drama (BBC afternoon play)	10.5
Television sport (BBC Wimbledon final)	4

The item with the smallest MS difference was the synthesised electronic music (1 dB). The music items in general have higher S levels than the items involving a lot of dialogue, such as drama, which is to be expected since dialogue is typically constrained to the centre area of a stereo image. The sci-fi radio drama has a higher average S level than the traditional radio drama as it involves a number of quite wide spatial effects interspersed with dialogue, whereas the traditional drama (10.5 dB) only makes use of subtle stereo ambience and modestly panned dialogue.

We see here, then, a range of MS ratios running from almost unity to just over 10 dB, which was found to be representative of much other programme material investigated. In none of the cases investigated did the overall (linear weighted) S level exceed the M level on average over the duration of the item, although there were many cases in which it did so on an instantaneous basis, and a number of cases in which some frequency bands showed an average S level higher than average M level. It may be seen that S level often exceeds M level at the extreme low frequency end of the spectrum, and occasionally at the HF end. A possible reason for this may lie in the nature of reverberant information or ambient noise signals at the extremes of the spectrum where little primary signal content is present. In the majority of the middle band M exceeds S by a considerable degree.

2 Creation of an artificial test signal

In order to measure the steady-state upmixing behaviour of the selected surround algorithms it was necessary to create an artificial test signal whose characteristics were known and which closely paralleled the nature of a typical programme signal. A signal with controllable MS ratio was desirable, in order that measurements could be made at a number of points, to determine the relationship between MS ratio and centre/surround signal derivation. It was anticipated that the level derived for the centre channel would be closely related to the magnitude of the M component, and that derived for the rear surround channels would be closely related to the S component.

It was acknowledged that some of the upmixing algorithms in fact used steering logic with a variety of time constants, in order to distribute signal power appropriately in the short term according to psychoacoustic requirements. Steering also takes into account dominant signal components in different panned positions. The measurements given here are therefore only a guide to the long-term average multichannel signal levels that might be derived from two-channel signals.

The test signal was based on uncorrelated pink noise generators. Using three channels on an ordinary stereo mixer, a single noise source was panned centrally (to enable control of the M component) whist two uncorrelated noise sources were panned equally to left and right mix buses (to enable control of the S component). By varying the ratio between centre-correlated and left-right-uncorrelated components it was possible to obtain a two-channel stereo signal with virtually any MS ratio, generating a stereo image that varied from central mono (no S component) to highly decorrelated stereo. The left and right channels of the resulting signal were adjusted in order that the overall signal level did not change as the MS ratio was varied.

Test signals with MS ratios of 6, 8, 10 and 12 dB were subsequently recorded onto a DAT tape and used to measure the multichannel signals that were synthesised by each of the algorithms under test.

3 Synthesised centre and surround signal levels

3.1 Equipment configuration and calibration

The equipment configuration used when making the measurements is shown in Figure 2.

The algorithms/surround processors surveyed were as follows:

Meridian 565 (a high-end consumer digital surround sound processor)

Modes

- Music
- Trifield
- Super stereo
- Prologic (Dolby matrix decoding for comparison, although not really intended for upmixing of ordinary stereo material)

Circle Surround (an analogue 5-2-5 matrix process, the decoder of which is claimed to be suitable for 2–5 upmixing)

Modes

- Music
- Video

Lexicon DC-1 (a high-end consumer digital surround sound processor)

Modes

- Music Surround
- Music Logic

Quintium (a digital surround algorithm designed by Peter Craven and Michael Gerzon, currently licensed for use in some consumer TV products)

All of these were configured in their default/preset states, as set by the manufacturer, since it was considered that this provided a reasonable starting point for a comparison and ought to represent a 'best compromise' setting. Some algorithms also allowed for some variation of levels, equalisation and delays by the user.

The loudspeakers used in the listening room, which meets the acoustic requirements of the ITU-R BS.1116 standard [2] except that its volume is slightly smaller, were matched BBC/Rogers LS5/8 monitors. The gains of their power amplifiers were adjusted according to the standard so that a pink noise signal with an rms level equal to digital programme alignment level at –18 dBFS produced an SPL of 78 dBA at the listening position when sounded from each loudspeaker individually.

Each upmixing processor was configured and calibrated according to the manufacturer's instructions for a five full-range loudspeaker set involving direct rear radiators and no subwoofers, according to the ITU-R BS.1116 standard configuration for five channel reproduction. This configuration was chosen to represent a typical professional five-channel surround replay system, according to the majority of recommendations present in international standards, although it is acknowledged that systems used in domestic environments may well differ considerably and some may use dipole rear radiators. The channels were labelled L (left), C (Centre), R (Right), LS (Left Surround) and RS (Right Surround) according to convention. It was considered inappropriate to use subwoofers for the application concerned. This was the configuration that would be used in subsequent subjective tests.

Two of the algorithms measured (Prologic and Quintium) were 2–4 channel processes rather than 2–5 channel processes. In those cases feeds of the mono surround signal to the two rear channels were arranged, either automatically by the processor concerned or manually using the mixer. Appropriate gain compensation for the correlated rear channel signal sum at the listening position was introduced either automatically or manually as appropriate.

All the processors provided an internal noise test signal that could be used to calibrate the output levels of each channel, and output gains were trimmed where necessary to ensure that the levels to each loudspeaker were set as recommended by the manufacturer, using an NTP multichannel level meter and/or SPL meter.

3.2 Measurement results and discussion

Measurements were made of the differences between the two-channel and synthesised multichannel signal levels, for each algorithm, having first set the gains of each of the processors so that their left and right outputs produced the same level as that of the original two channel source material when switched into bypass mode (in other words, with a signal going through the processor and its gain controls but with the algorithm switched out). The processors therefore displayed unity gain in bypass mode. 0 dB on the graphs in Figure 5 indicates no difference in level compared with the original two-channel signal/bypass mode on L, R channels.

Signal levels were measured using an NTP multichannel level meter with a PPM-like integration time.

The graphs in Figure 3 indicate the algorithm to which they relate. It will be seen that the LS and RS signal levels typically tend to increase in a fairly linear relationship with the decreasing MS ratio, indicating that surround level is closely dependent on the S level of the signal. Most algorithms clearly base their extraction of rear channel information on the two-channel difference signal, which is a process that goes back many years in concept. It is the S component which contains the majority of reverberant/ambient information. It is not an entirely linear relationship between MS ratio and surround signal level in all cases, with Circle Surround showing a slightly expanded relationship and Quintium showing a more compressed relationship.

The treatment of left and right channels varies, some algorithms maintaining them at a similar level to the original (e.g. 565 Music mode), whilst many others drop the level by between 1 and 3 dB. Prologic (the Dolby process implemented on the Meridian 565 in this case) can be seen to bias the front signals strongly towards the centre, with L and R levels reduced considerably compared to those of the two-channel original. This is not entirely unexpected as Prologic is a heavily steered matrix, designed for decoding Dolby Surround encoded movies rather than for processing unmatrixed conventional stereo material. The effect on conventional material is to pull the image strongly towards the centre. It is suggested that ordinary stereo material would have to be mixed with a very wide front image to compensate for this centre tendency if it is intended to be replayed through a Prologic decoder, but this could result in an unsuitable effect for two-channel listeners.

The behaviour of centre channels is interesting. In the case of most algorithms it rises very slightly (but not linearly) as M:S ratio gets greater, while the L and R levels either stay constant or fall slightly. This depends on the degree of steering employed in the algorithm, since those which employ strong steering would tend to pull signals into the centre loudspeaker in cases when a dominant centre component is detected. The difference can be seen quite clearly with the Circle Surround processor in Music mode compared with Video mode. In Video mode the L and R channel levels drop considerably with increasing M–S because more front steering is involved in this mode (designed mainly for movie and television operation where centre dialogue is common). A similar but more subtle effect can be seen with the Lexicon DC-1 in Music Surround mode compared with Music Logic mode. The latter mode is slightly steered whereas the former feeds the original L and R channels directly to the new L and R without steering, just adding a small amount of centre to stabilise the middle of the image for off centre listeners.

It is important to note that time delay, equalisation and phase modification also play a large part in the action of the more sophisticated algorithms. Level is only one of the psychoacoustic factors that is used to generate an appropriate impression of stereo image and spaciousness. The graphs of levels alone, therefore, do not tell the whole story, particularly in the case of algorithms such as Trifield which introduce antiphase crossfeed information between the channels to maintain stereo width whilst adding a centre channel.

Quintium behaves unusually in that it displays very little variation in surround signal levels as MS ratio changes. This suggests that it is attempting to compensate for what it detects is a decreasing level of spatial information in the original. Notice that the surround level only drops by 1 dB while the MS ratio changes by 6 dB. Discussions with the designer suggest that this is intentional behaviour designed to add spaciousness to material which displays little or none, and currently takes place by automatic detection.

Overall it may be seen that the algorithms tested synthesised a wide range of centre and surround signal levels from the artificial two-channel test signal, suggesting that there is little common agreement over what level of synthesised surround information is appropriate, even amongst those algorithms designed for music applications. The lowest level of surround signal extraction was observed with the Meridian 565 in Music and Trifield modes (–12 dB when M–S = 6 dB), while the highest was observed with Circle Surround (–5.5 dB under the same conditions).

Informal listening tests suggested that signal levels were far from the most important factor in determining the spatial quality of the reproduced sound. Two algorithms, for example, had a similar (high) surround level and yet sounded completely different. In one case the front image was made very indistinct because the rear channels seemed to pull the front information backwards, whereas in the other the front image was well preserved and a pleasing sense of ambience was added from the rear that did not distract in the least.

4 Level differences when comparing original two-channel and synthesised five-channel material

In the accompanying paper [1], controlled listening tests are described in which it was important to ensure that the original two-channel version of programme extracts sounded similarly loud to the synthesised five-channel equivalent, in order that AB comparisons could be made. Clearly, the comparison would be most valid if the sound pressure levels of the two versions were the same at the listening position, in order that one could eliminate any level dependent factors from the results (Bech [3] has shown that level calibration can be an important factor in the judgement of spatial impression). In fact this is impossible to achieve in practice because the signal level fed to the centre and rear loudspeakers by the algorithms varies depending on the instantaneous MS ratio of the programme material, amongst other factors such as steering. At best one could only achieve an approximation to level equality for the two versions.

In order to manage this issue as well as possible, the test signal described above was employed so as to enable SPL measurements of both processed and unprocessed versions of the reproduction under programme-like conditions. A test signal MS ratio of 6 dB was chosen because it corresponded to a reasonable mid-point of the average MS ratios of the programme items selected for the subjective tests. Measurements were then made of the differences in SPL between the two-channel and multichannel-processed reproductions, for each algorithm, both at the listening position and 10cm to the right of that (in order to obtain an indication of the position dependence of the level calibration), having first set the gains of each of the processors to be unity in bypass/2-channel mode, as described above. The two channel version was then aligned for an SPL of 85 dBA at the listening position.

The SPL was measured with the rear loudspeakers both on and off, in order to determine how much difference the rear loudspeaker contribution would make to the overall SPL. The results are shown in Table 2. (The algorithms are not identified here, as they were a subset of those described above, used for the controlled subjective tests.)

Table 2 Total SPLs (dBA) measured at listening position and 10cm to right for upmixed versions of test signal (M–S = 6 dB). 2 channel version aligned for 85 dBA at listening pos.

Algorithm	*Listening pos. 5 channels*	*Listening pos. LCR only*	*10cm right 5 channels*	*10cm right LCR only*
1	88.0	87.5	87.7	87.4
2	88.0	87.7	87.8	86.6
3	87.0	86.4	86.5	86.0
4	86.7	85.7	86.7	86.0

It will be noticed first that the upmixed SPLs are higher than the bypass/two-channel SPLs, due mainly to the addition of the centre channel. Second, the presence of the surround channels makes no more than 1 dB difference to the overall SPL at the listening position. This is an interesting result, and is mainly due to the fact that the surround signal levels are considerably lower than those of the front channels with all the algorithms concerned. The surround channels contributed more as the S level of the two-channel material rose, and less as it fell. Clearly the contribution from the rear channels would be slightly higher at MS differences of less than 6 dB. Given that this effect is programme dependent and cannot be controlled, it was decided to normalise the levels of the two versions using the front channels only, and to allow the surround level to vary according to the programme. This was not ideal, but the best compromise that could be found. Therefore gain corrections were applied to all the processed signals so that the SPL at the listening position was the same for the LR two-channel replay as it was for the LCR component of the five-channel replay using the test signal with an MS ratio of 6 dB. The rear channels were correspondingly attenuated to preserve the front-back relationship correctly.

5 Conclusion

The average MS ratios of typical recorded and broadcast two-channel stereo programme material was found to vary between about 1 dB and 11 dB, with most music material having small MS ratios and most dialogue-based material having higher ratios. Synthesised multichannel signals, derived by a variety of surround sound algorithms from a two-channel test signal with variable MS ratio, were found to have levels closely related to the MS ratio of the two-channel programme material, with rear channel levels depending strongly on the magnitude of the S component. The magnitude of centre and surround signal levels generated by the algorithms concerned varied widely.

The artificial stereo test signal described in this paper proved to be a valuable tool in calibrating the relative listening levels of two-channel programme items and their equivalent five-channel synthesised versions.

References

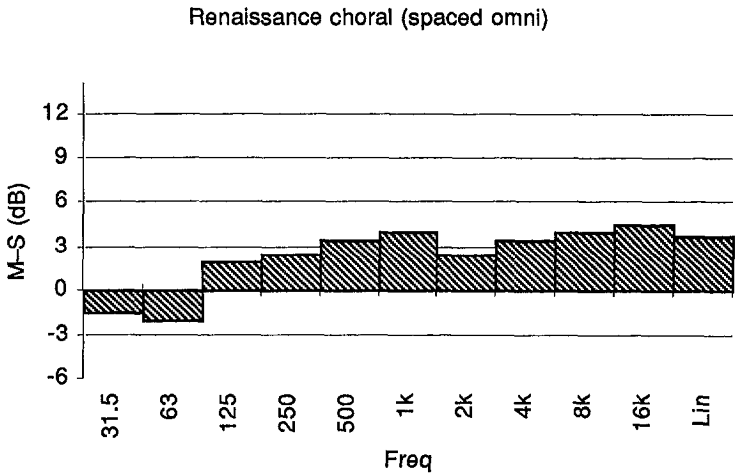
1 Rumsey, F. (1998) Controlled subjective assessments of 2–5 channel surround sound processing algorithms. Presented at *AES 104th Convention, Amsterdam, 15–18 May*

2 ITU-R (1994) Recommendation BS. 1116. Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. International Telecommunications Union

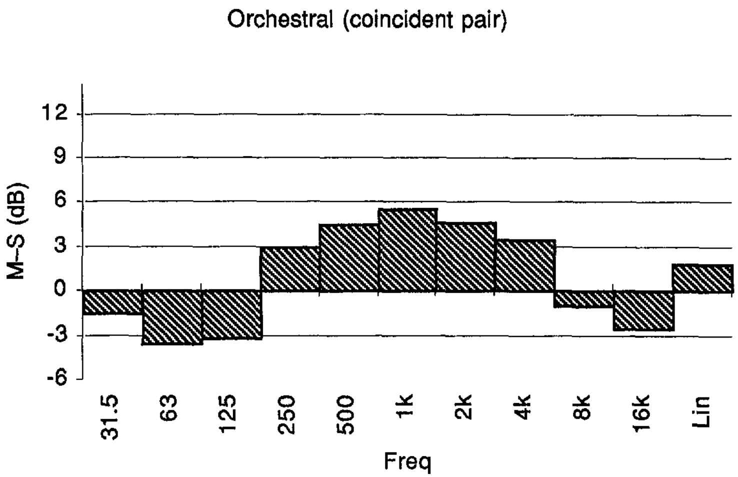
3 Bech, S. (1997) Calibration of relative level differences of a domestic multichannel sound reproduction system. Presented at *AES 102nd Convention, Amsterdam, 22–25 Mar.* Preprint 4433

Figure 1 Long-term average MS ratios of programme extracts

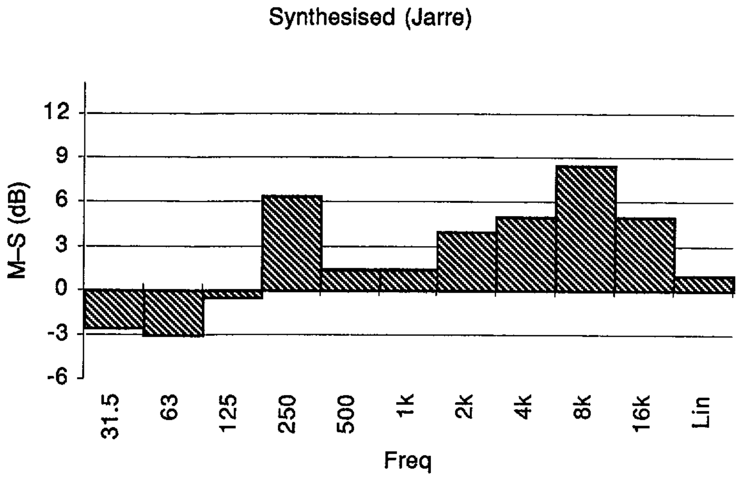
(a)



(b)

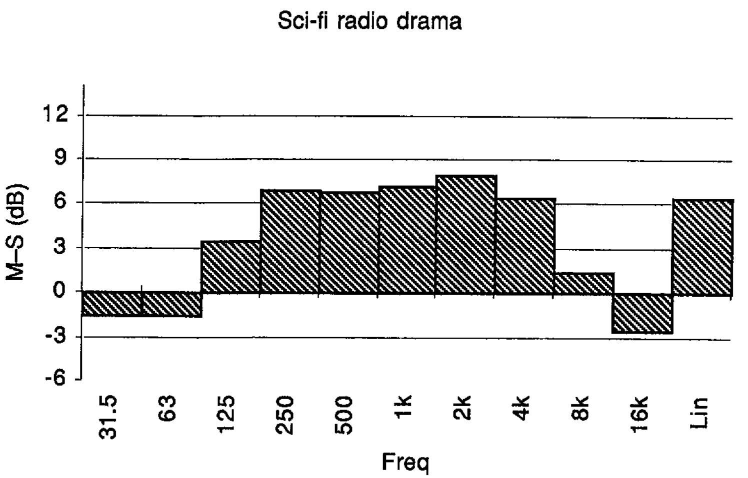


(c)



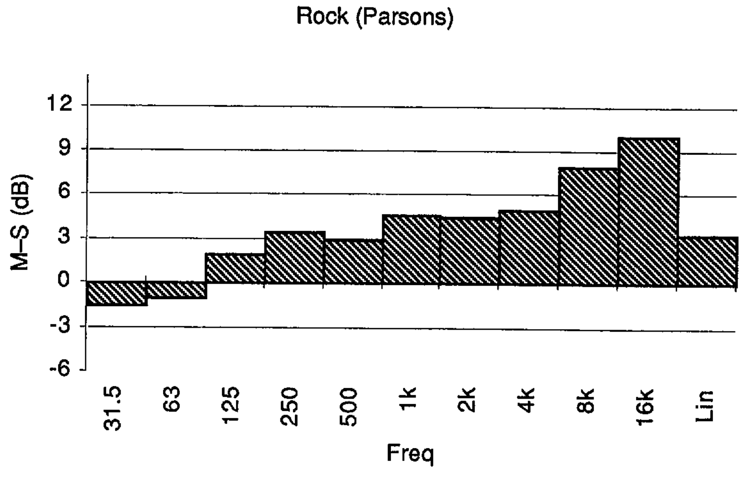
Synthesised (Jarre)
12
9
6
3
0
-3
-6
M–S (dB)
31.5
63
125
250
500
1k
2k
4k
8k
16k
Lin
Freq

(d)



Sci-fi radio drama
12
9
6
3
0
-3
-6
M–S (dB)
31.5
63
125
250
500
1k
2k
4k
8k
16k
Lin
Freq

(e)


Rock (Parsons)
M–S (dB)
12
9
6
3
0
-3
-6
31.5
63
125
250
500
1k
2k
4k
8k
16k
Lin
Freq

(f)

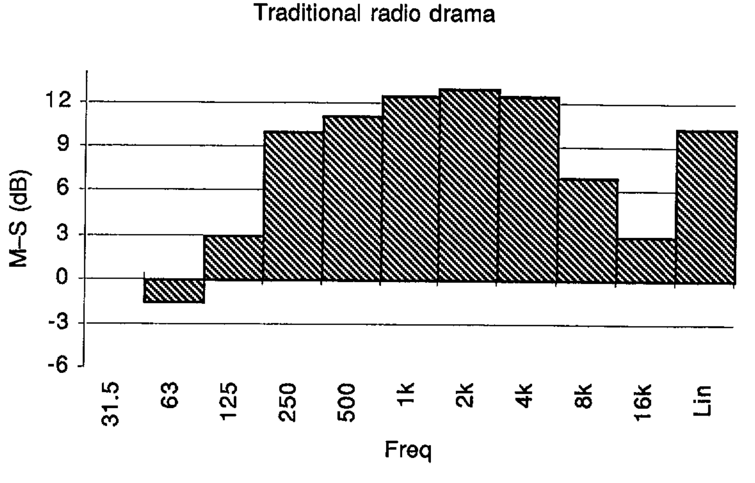

Traditional radio drama
M–S (dB)
12
9
6
3
0
-3
-6
31.5
63
125
250
500
1k
2k
4k
8k
16k
Lin
Freq

Figure 2 Equipment arrangement for level measurements

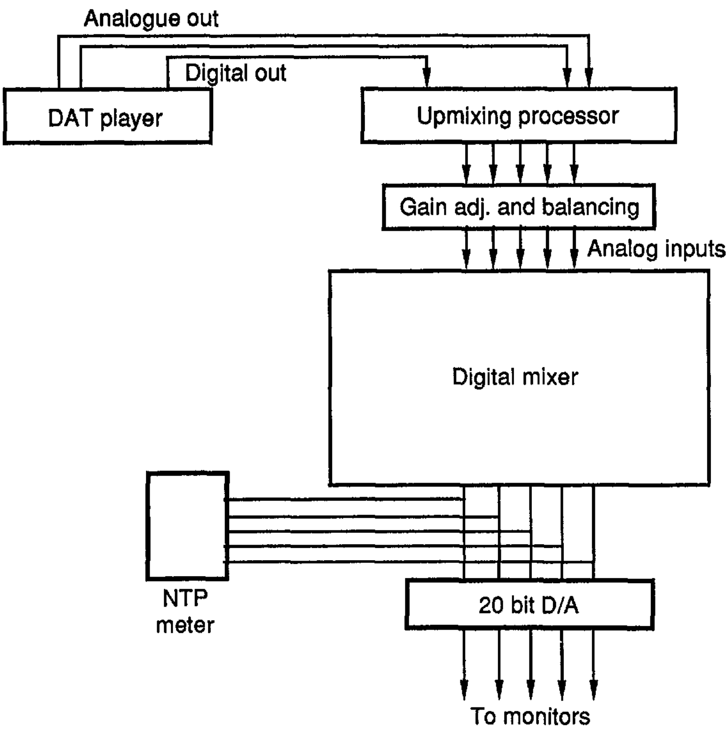
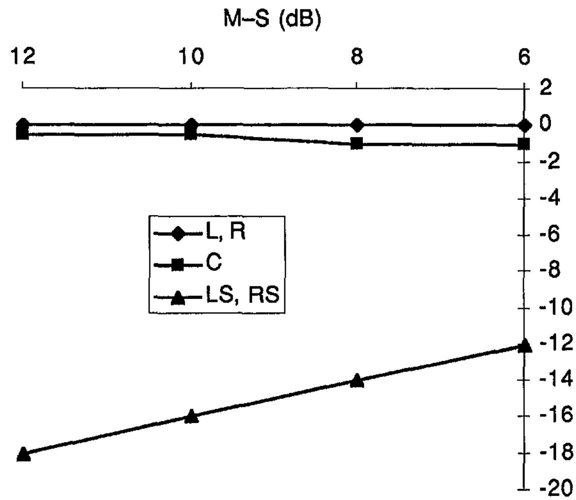
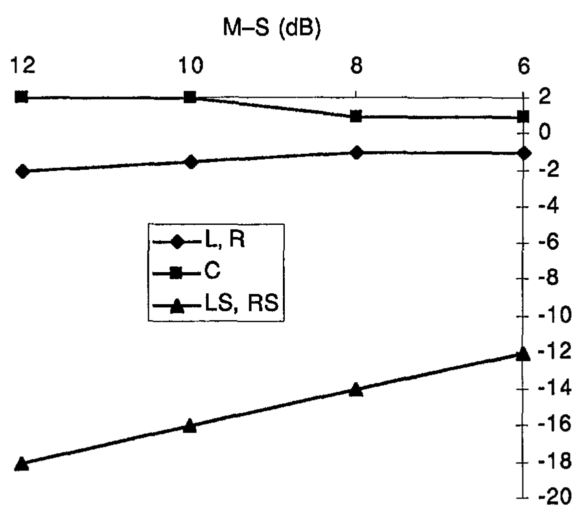


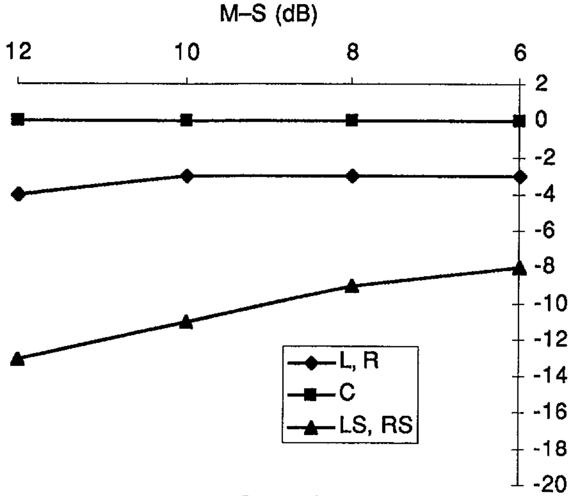
Figure 3 Synthesised multichannel signal levels versus MS ratio of test signal



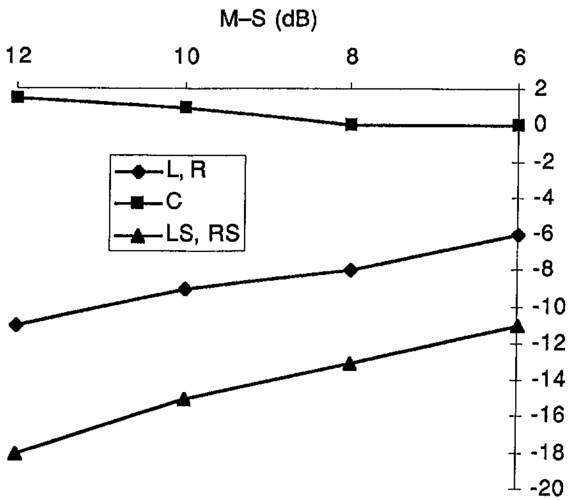
565 Music



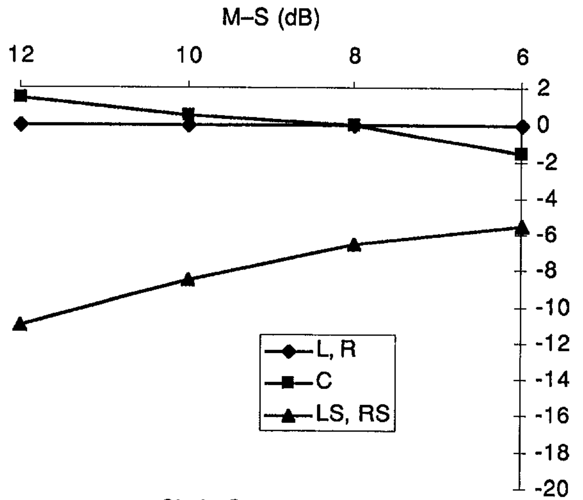
565 Trifield



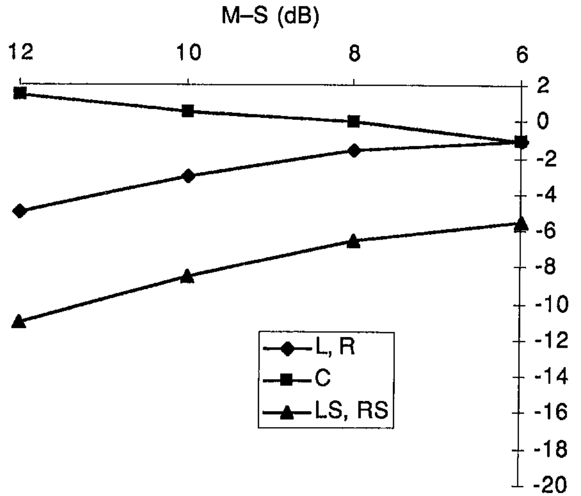
565 Super Stereo



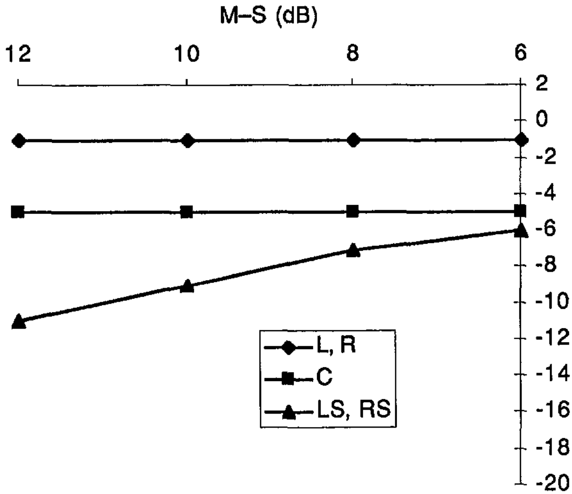
565 Prologic



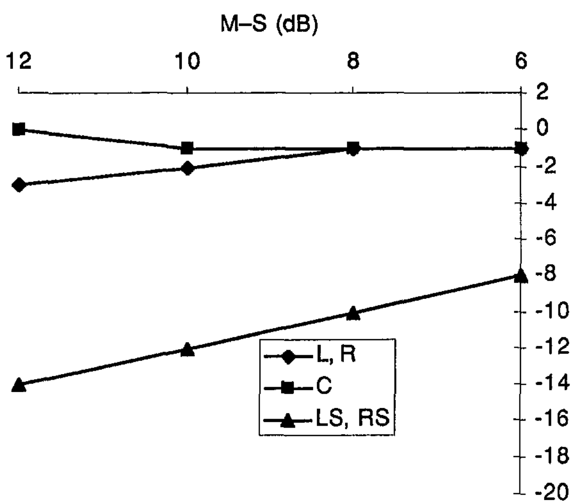
Circle Surround – Music



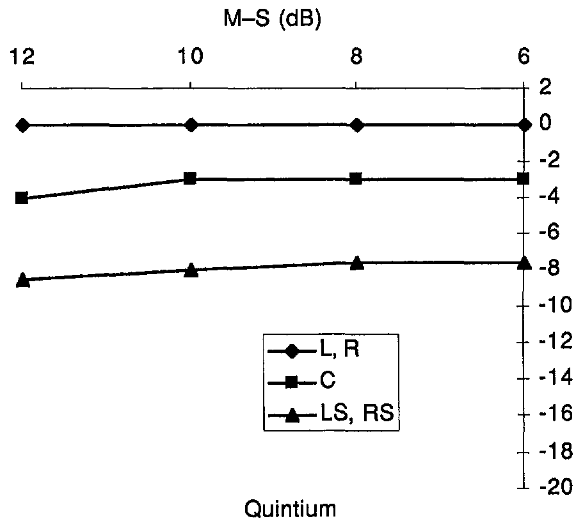
Circle Surround – Video



Lexicon DC-1 – Music Surround



Lexicon DC-1 – Music Logic


M–S (dB)
12
10
8
6
2
0
-2
-4
-6
-8
-10
-12
-14
-16
-18
-20
L, R
C
LS, RS
Quintium