Proceedings of the Institute of Acoustics

A NEW STANDARD FOR AUDIO VISUAL MIXING THEATRES

A.S. Munro

Munro Associates, 28 Park St. London SE1 9EQ

1. Introduction

For the last 10 years or more Dolby Labs have licensed and supplied surround sound technology to the audio and video industry and at the last count 32 million homes are equipped with some form of multi channel decoder. In addition over 600,000 5.1 decoders have been shipped to customers in apparent anticipation of the 2^{nd} generation digital home entertainment revolution. In the next few months the specification for DVD audio encryption will (hopefully) be agreed. In any event there will be new demands placed on recording and post production studios to provide six track masters which have been mixed and monitored using something approaching a standard format for the electro acoustic interface.

2. DVD Mastering Audio in 5.1 Format

With the growth of interest in surround sound encoding there has been a great deal of confusion regarding the best way to set up a mastering facility to monitor accurately the audio channels in a given format.

It is useful to return to the roots of surround sound in order to define exactly what is expected of the new media, in terms of the creation of a three dimensional sound picture.

The modern form of surround sound was created in mid 70s by Dolby Labs in San Francisco. The trick was to use the optical sound track space on 35mm film to fit an encoded, analogue, stereo signal which could then be processed to produce a centre front channel, left and right stereo, plus a discrete sub bass channel and a 'surround' channel.

Several benefits were derived from the new format which apply equally well to current multi channel modes;

- the centre channel contained all the dialogue and those on screen effects which need maximum spatial stability. This is the equivalent of the original mono optical track and the Dolby Stereo track is mono compatible to this day. The single sound source ensures that there is no phasing or combing of the dialogue, off axis of the centre line of the theatre, which would be detrimental to intelligibility and naturalness of reproduction. The centre channel is the absolute reference source and all other channels must be phase and amplitude aligned to it.

- The left and right channels reproduce the music and sound effects in such a way as to create a sense of space and distance which matches the on screen action as closely as possible. Some effects are also fed to the centre channel for maximum impact.

  Occasionally dialogue will be panned into the stereo 'picture', to create movement.

- The surround channel was, and is, used for off screen sound effects and atmosphere, such as wind and rain, spooky echoes, etc. etc.. The bandwidth of the surround channel was limited to 7 kHz for improved noise and distortion performance but with 5.1 formatting this is not the case.

- The sub channel was used to good effect in movies such as Earthquake but it was always considered as an option as far as film processing was concerned and the use of a sub speaker in domestic formats is mostly to allow reduction in size of the main speakers by filtering all signals below 120 Hz and recombining them in a single channel. With large cinema systems the sub channel has come into its own and the advent of Dolby SR and SRD has allowed much more use of dynamic range and high energy at low frequencies.

With the advent of digital, discrete multi channel recording all of the limitations of Dolby Stereo and Surround were removed at a stroke and yet there was an important need to maintain backwards compatibility between the thousands of existing theatres with Dolby equipment and a new generation with digital playback systems.

This was achieved quite neatly at the projection end of the chain by squeezing the digital code into the space between the film perforations, leaving the optical track in its original form. The total flexibility and compatibility of this system means that one print can be used in any modern theatre. There is the added bonus that the analogue track automatically cuts in if the digital track drops out. Although DVD will be used for other formats it will be heavily influenced by the precedent for Dolby Digital (AC3) so I make no apologies for harping on about the film business.

That said I do recognise the potential in DVD for more linear audio formats which do have potential for much higher fidelity that AC3 but that is another story all together.

3. Monitoring Format

The essential question at the moment is how do the smaller control room environments used for DVD mixing compare with the traditional film dubbing theatres which have been mixing 5.1 for several years?

To judge this it is important to recognise the relationship between the direct and reverberant sound in each room.

The person judging the sound balance and quality must hear the correct blend of direct energy from each speaker and a pre determined amount of room reflection and reverberant energy. The ratio between the two can be controlled by the following factors;

- speaker directivity over the required frequency range
- distance and subtended angle from the speaker to the listening position
- room volume and geometry
- room acoustic and absorption

It is by manipulating these factors that it possible to obtain a balance in a smaller room of say 50 sq. metres which is similar to a much larger theatre with a highly directional horn, at least in the speech frequency band which is the most important from an intelligibility point of view.

Figure 1 shows a spread sheet which is useful for calculating the final sound level produced by all the different combinations of energy, time and space which go together to create a 'sound field'.

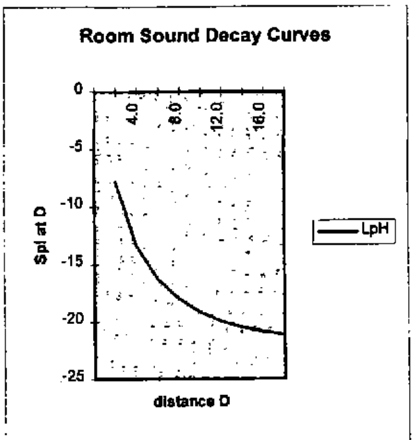
A NEW STANDARD FOR AUDIO VISUAL MIXING THEATRES

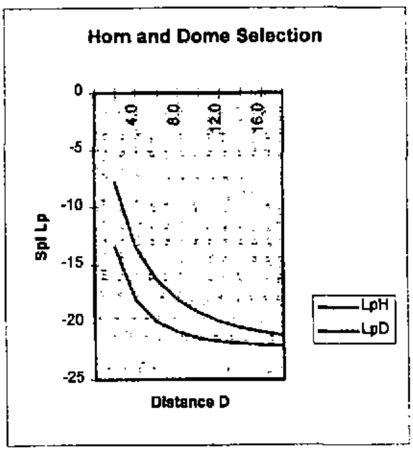
Freq.	L	W	H	Lw(1W)	Lw(Rqd)	EPR	Q	D	S	V	a	T60	Lp c	Dc
63	20	10	5	100	130	1000	2	15.0	700	1000	0.3	0.64	111	2.89
125	20	10	5	100	130	1000	2	15.0	700	1000	0.4	0.45	110	3.34
250	20	10	5	100	130	1000	4	15.0	700	1000	0.5	0.33	109	5.28
500	20	10	5	100	130	1000	8	15.0	700	1000	0.5	0.33	109	7.47
1000	20	10	5	100	130	1000	10	15.0	700	1000	0.5	0.33	110	8.35
2000	20	10	5	100	130	1000	15	15.0	700	1000	0.6	0.25	110	11.20
4000	20	10	5	100	130	1000	25	15.0	700	1000	0.6	0.25	111	14.46

L	W	H	Lw	Q horn	Q dome	D	S	V	a	T60	LpH	LpD	Dc
20	10	5	0	8	2	2.0	700	1000	0.5	0.33	-8	-13	7.47
20	10	5	0	8	2	4.0	700	1000	0.5	0.33	-13	-18	7.47
20	10	5	0	8	2	6.0	700	1000	0.5	0.33	-16	-20	7.47
20	10	5	0	8	2	8.0	700	1000	0.5	0.33	-18	-21	7.47
20	10	5	0	8	2	10.0	700	1000	0.5	0.33	-19	-21	7.47
20	10	5	0	8	2	12.0	700	1000	0.5	0.33	-20	-22	7.47
20	10	5	0	8	2	14.0	700	1000	0.5	0.33	-20	-22	7.47
20	10	5	0	8	2	16.0	700	1000	0.5	0.33	-21	-22	7.47
20	10	5	0	8	2	18.0	700	1000	0.5	0.33	-21	-22	7.47

KEY

L	length
W	width
H	height
Lw	sound power of source speaker
Q	directivity index
D	distance to point of measurement
EPR	electronic power required (amplifier)
S	room surface area
V	room volume
a	average absorption
Sa	total absorption
T60	reverb decay time
Lp	sound pressure at D
Dc	critical distance Ld=Lr





The graphs show the sound level produced by a horn loaded speaker and a direct radiating driver in the same theatre at varying distance from the screen. It can be seen that although the horn produces a much higher sound level over the first 10 metres, when we measure at a distance of 16 metres the levels are almost the same. This is because we are well into the reverberant sound field and the direct sound no longer has much influence, except that it arrives at ear first and this why we can still perceive some kind of stereo image, even when the sound is dominated by the room itself. The distance from the speaker at which the direct (Ld) and reverberant (Lr) sound levels are equal is called the critical distance, Dc and in a large theatre it is important to design the acoustics and the sound system so that all of the audience is beyond the critical distance and yet still within the coverage angle of the speakers. Ignorance of this fact accounts for the large number of complaints received by cinema operators of the 'too loud', 'couldn't understand what he was saying' variety.

It is important to realise that acoustically 'dead' rooms are extremely bad news when it comes to even sound coverage as the front seats will be blasted and the people at back drowning all around them with the crunching of pop corn! That is why concert halls are designed to squeeze every last drop of energy out of the woodwork; it makes every seat worth sitting in.

4. The Grand Unified Theory of Sound

There is a general relationship for all these parameters which has been defined over the years in the form;

$$Lp=Lw+10\log\{(Q/4\Pi D^2)+4/R)\}$$

This formula is less daunting that it looks; Lp is the sound pressure at the listening distance D, Lw is the sound power radiated by the speaker (which is room independent).

Q is the directivity of the speaker, 1 being omni directional and 30 a very narrow mid range horn. R is the room absorption which can be measured or designed according to the required reverberation time.

Lets start by looking at a known quantity, a major league film dubbing theatre in downtown Hollywood or even downtown Shepperton!

We can say from prior knowledge and general consent that the reverb time should be around 0.5 seconds in a room of 1000 cu. metres at 1000 Hz. We also know that the distance to the screen and therefore the speakers is 15 metres. We can assume that our reference sound power output is 100dB for 1 watt for a typical off the shelf horn loaded speaker.

The formula tells us the sound pressure will be approximately 80dB. Now lets put a much smaller speaker much nearer to the listener and see how we can achieve the same level. No problem but what about our direct to reverb balance?

We can judge this by comparing the direct and reverberant parts of our formula. When both are equal, at a given point, this is called the critical distance, beyond which the sound level is dominated by the room energy and remains constant (apart from local effects caused by poor geometry and room resonance, of which more later.

To locate the critical distance Dc the main formula can be re written as;

$$Q/4\Pi D^2=4/R \text{ or more usefully } Dc = \sqrt{QR/16\Pi}$$

This shows that by reducing Q it is possible to arrive at the same direct to reverberant ratio with both horns and wide dispersion near field speakers as can be seen from table 1.

We can also compare our large dubbing theatre with its large horn speaker with a smaller room using a smaller pair of soft dome monitors, such as the Dynaudioacoustics M3F.

A NEW STANDARD FOR AUDIO VISUAL MIXING THEATRES

Fig 3

Freq.	L	W	H	Lw(1W)	Q	D	S	V	a	T60	Dc
63	20	10	5	100	2	15.0	700	1000	0.3	0.64	2.89
125	20	10	5	100	2	15.0	700	1000	0.4	0.45	3.34
250	20	10	5	100	4	15.0	700	1000	0.5	0.33	5.28
500	20	10	5	100	8	15.0	700	1000	0.5	0.33	7.47
1000	20	10	5	100	10	15.0	700	1000	0.5	0.33	8.35
2000	20	10	5	100	15	15.0	700	1000	0.6	0.25	11.20
4000	20	10	5	100	25	15.0	700	1000	0.6	0.25	14.46

L	W	H	Q	D	S	V	a	T60	Dc
20	10	5	1	15.0	700	1000	0.4	0.45	2.36
20	10	5	2	15.0	700	1000	0.4	0.45	3.34
20	10	5	4	15.0	700	1000	0.4	0.45	4.72
20	10	5	8	15.0	700	1000	0.4	0.45	6.68
20	10	5	16	15.0	700	1000	0.4	0.45	9.44
20	10	5	32	15.0	700	1000	0.4	0.45	13.35

KEY

L	length	S	room surface area
W	width	V	room volume
H	height	a	average absorption
Lw	sound power of source spe	Sa	total absorption
Q	directivity index	T60	reverb decay time
D	distance to point of measur	Lp	sound pressure at D
EPR	electronic power required (	Dc	critical distance Ld=Lr

L	W	H	Horn	Q	D	S	V	a	T60	Dc	D/R ratio
10	7	3	soffit	4	3.0	242	210	0.4	0.27	2.78	1.5
10	7	3	soffit	8	3.0	242	210	0.4	0.27	3.93	4.6
10	7	3	dome	16	3.0	242	210	0.4	0.27	5.55	7.6
20	10	5	90*180	8	15.0	700	1000	0.4	0.45	6.68	-4.8
20	10	5	90*90	16	15.0	700	1000	0.4	0.45	9.44	-1.8
20	10	5	90*40	32	15.0	700	1000	0.4	0.45	13.35	1.2
20	10	5	40*40	64	15.0	700	1000	0.4	0.45	18.89	4.2

The direct to reverberant ratio [D/R] is given by;

$$[D/R]=10\log(4/\Pi D^2)/(4/R)$$

Fig. 3 shows that in a room of 200 cu. Metres, at a listening distance of 3 metres the direct to reverberant ration will be 4.6 for a Q of 8, which is typical of the mid frequency directivity of a dual driver, soffit mounted speaker system.

It can be seen that it would take a very directional horn (Q=64) to achieve the same ratio in our large theatre. In fact it can be seen that most horn configurations will produce a negative ratio, i.e. the room reverberation is the dominant factor) in most large dubbing theatres.

It is also the case that monitors in small rooms need to be as un-directional as possible if it is desired to emulate a theatrical sound field.

This is particularly true in the reproduction of sound in 5.1 format. The object of the exercise is to recreate a realistic sound field based on what we perceive as natural. That means discrete visually interactive sounds from the front channels and indirect or ambient sound from the rear, with the occasional directional effect to give interest to the mix.

This is a far cry from some of the audio-only 5.1 mixes I have heard which could cause severe neck ache, not to mention a low irritation threshold once the novelty of the lead guitarist hanging on the back wall begins to wear off. I am reminded of the early days of stereo LPs, when ping-ponging sounds and trains panning across the living room were the order of the day.

The most pleasing results will occur when the room (monitoring or listening) is not too dead and the speakers are very open and actually given some space to radiate before early reflections contrive to destroy the spatial images of the mix.
I have noted that multi channel digital sound has created a demand for mixing theatres which are if anything more live that the equivalent old analogue rooms.
I can only explain this terms of naturalness and adherence to audio fidelity which I think bodes very well for the future.

Many people are unsure about the best way to set out speakers for DVD mastering.
I can only give a clear answer in the context of mixing to picture and for this format I strongly recommend the layout shown by Dolby, in their excellent literature.

As DVD will use the Dolby Digital format for most applications it is logical to set up ones room and monitoring system as they recommend. The ITU-R specification has also been adopted by broadcast engineers and it is used for critical listening rooms. This internationally recognised document is full of obscure references to methods of appraising audio but is recommended for its depth and weight!

The front three speakers should be positioned equi distantly from the mix position with an angle of 60 degrees subtended by the left and right. All three should be identical models with a matched response within +/- 3dB in any 1/3 octave band between 250 Hz and 2 kHz, measured at the mix position. The limits can be slightly relaxed above and below this range but for mastering purposes this would not be advisable.

The surround speakers should be positioned at the same distance to the mix position as the main front speakers, at a angle of 110 degrees from the centre line. If this is not possible then a digital delay should be used to bring the speakers back into coincidence. Ideally the surround units will be the same as the front ones but if that is not possible they should be of the same manufacturer using identical driver technology. The polarity and phase summing of each speaker should be such that any two together will produce an increase in sound level of between 3 and 6 dB at any frequency up to 10 kHz. Above that frequency it is difficult to place the microphone so precisely that phase coincidence is perfect (but many purists will insist on trying).

Several differences to this set up must be considered in the context of film sound;

- The larger rooms used for film mixing dictate that several pairs of surround speakers are used, to ensure that everyone in the room hears the same delay and direct to reverb ratio. The use of time delay and separate power amplifiers is recommended to achieve consistent results. THX maintain strict criteria for surround speaker placement in order to achieve world wide compatibility between theatres and dubbing stages;

- The surround speakers must often be placed at high level and angled to cover a large area at constant level.

- The surround speakers are rarely the same as the front ones which are invariably large, horn loaded systems of considerable bulk.

- The angle subtended by the main left and right speakers should be 45 degrees at the mixing console. This assumes the speakers are at the extremes of the screen which begs the question, 'how do we adjust for different film aspect ratios?'. One solution is to use 5 speakers at the front, the inner pair used for TV formats and the outer pair for wide screen ratios.

This apart it should be possible to maintain compatibility between the two formats so long as good use is made of low directivity speakers in small, dry rooms.

5. The LFE Channel

In a true 5.1 system the .1 is a discrete low frequency channel which is used by the mixing engineer to create special effects and to extend the normal frequency range of music and other mixed events. The channel can also be used as a sub woofer system to accept a feed of bass from the main speaker channels whose speakers may be too small or unsuitable to reproduce high levels of very low frequency material. This mode is created in the decoder of the playback system and is a user definable mode, not one which would normally be considered in the mixing or mastering room, except for comparative use or plain curiosity.

The mixing encoder channel contains a steep filter at 120 Hz which prevents higher frequencies leaking into the LFE track. Dolby recommend an 80 Hz filter in the console channels to make sure nothing is even heard (and therefore localised by the final audience).

The LFE channel speaker is set at a level 10 dB higher than the other channels in the mixing theatre so it is important to do the same in any replay situation. This should be done with band limited pink noise to smooth out localised pressure wave variations. Most music mixing engineers do not send material to the LFE channel as there is enough mutual coupling of the main speakers to produce more than enough low frequency energy; in fact a well matched system will produce up to 10dB more bass than a single speaker and care must be taken to balance the mix accordingly. The LFE channel would obviously be useful for extended effects such as cannons and organ pipes and music reproduced down to 20 Hz or less will no doubt become more interesting as skills and production values improve. We may even see an improvement in pop music mixes as engineers(?) finally realise that small, closed box woofers, rolling off at 70 Hz do the serious audiophile no favours at all.

6. More Acoustics

Several people have asked me how to ensure that all five speakers sound the same and the answer is basically down to a few key factors;

- the room must be symmetrical about the centre speaker axis, with windows and doors placed so as to steer sound reflections away from the mix position.
- early reflections (within 15ms of the direct sound) should have an amplitude of no more that 10dB below the direct sound, especially in any band between 250 Hz and 2 kHz.
- Each speaker should be placed at least 1 metre from any wall and not equidistant to two walls (or they should be completely flush with the wall for large soffit mounting systems).
- the room acoustic should be as diffused as possible with a reverberation time of (0.25(V/100)^1/3) +/- .05s. In other words the reference value is 0.25 seconds for a 100m^3 room, increasing with room size. My favourite spaces are often about this value so I would argue that the ISO/MPEG/SMPTE guys have got it about right.
- Room proportions should not be vastly different but never the same and cubic is a disaster. Room modes will always dominate the low frequency performance,

closely followed by first reflections from floor and ceiling. 7m by 5m by 3m is a good starting point for mastering rooms.

- Background noise should be judged according to material and the final medium. Film and pop music are often mixed in rooms with a noise floor of NR30 because a mix which is too wide in dynamic range will cause problems on playback (noisy theatres, car radios, Walkman in the street etc. A critical listening room needs to be very quiet so as to judge the limits of other systems' dynamics and NR10 to NR15 might be called for in some circumstances.

7. Final Take

DVD is a new and exciting medium with huge potential for extending audio quality beyond anything the average consumer has experienced before. The industry is suffering greatly at the moment because indecision and format wars are diluting the message. It is the task of the sound engineer to deliver a product which translates from the studio to the home and to achieve this it is essential to define the required direct to reverberant ratio, dynamic range and frequency response. The fact that film mixers work almost exclusively in the 'far field' and video dubbing engineers rarely have much choice but to use 'near field' monitors results in a conflict of practice which must be resolved if a consistently accurate DVD product is to be produced.