Surround: The Current Technological Situation

David Griesinger
Lexicon
3 Oak Park
Bedford, MA 01730
www.world.std.com/~griesngr

Introduction

The Audio Engineering Society named its annual conference in Los Angeles "Surrounded by Sound". Here in Paris we are holding another conference on surround. Clearly "Surround" is becoming the "next big thing." In spite of all the activity, the answers to many questions are unclear. What is "surround sound"? How do you record it? How do you play it back? And where is it heading in the market? In this paper I offer some personal answers to these questions.

What is surround sound?

The answers to the questions above are personal in part because there are two major approaches to the idea of surround - and I want to talk about only one of them. The divide comes over the issue of how we listen to sound, and with whom. If we want to listen alone, and we are willing to restrict our listening position to a single point, we will design systems very differently than if we want to move around or to listen with friends.

For many people the ideal music listening experience is solo - complete immersion in a performance or sonic experience. I do not have statistics, but it seems likely that the most serious music listening is done this way. One may go to a concert with friends, but one concentrates completely on the music during the performance. If you count listening to music while driving, it is likely that recorded music is also primarily heard by a single individual. If this individual is willing to listen only at a single position, systems that work optimally only at one point in the room may make sense. The problem is - I am not one of these people. I like company and I like to move around - rather fast sometimes.

Yet the concept of listing solo at a precise point is so pervasive that we have turned it into a standard. Figure one shows the familiar speaker arrangement for 3/2 multichannel sound systems. The listening area is not drawn. There is no listening area, there is only a listening point.

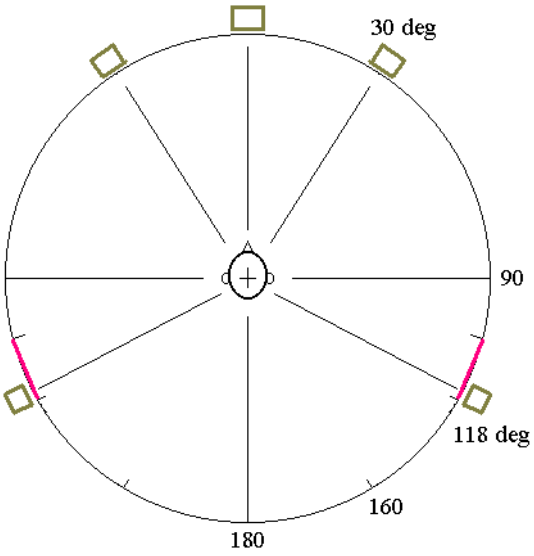

Figure 1*: The standard 3/2 speaker arrangement. Note that there is a single listener, precisely positioned at a point equidistant from all the speakers.*

Yet almost all of the surround systems on the market are sold as home cinema systems, and almost all the available surround recordings come from the film industry. Cinema is oriented toward couples and groups. Indeed, the whole idea behind home cinema is to entertain your family and friends. For cinema we need sound recordings and systems that work well over as large a listening area as possible - and the standard speaker layout seems a bit silly for this task.

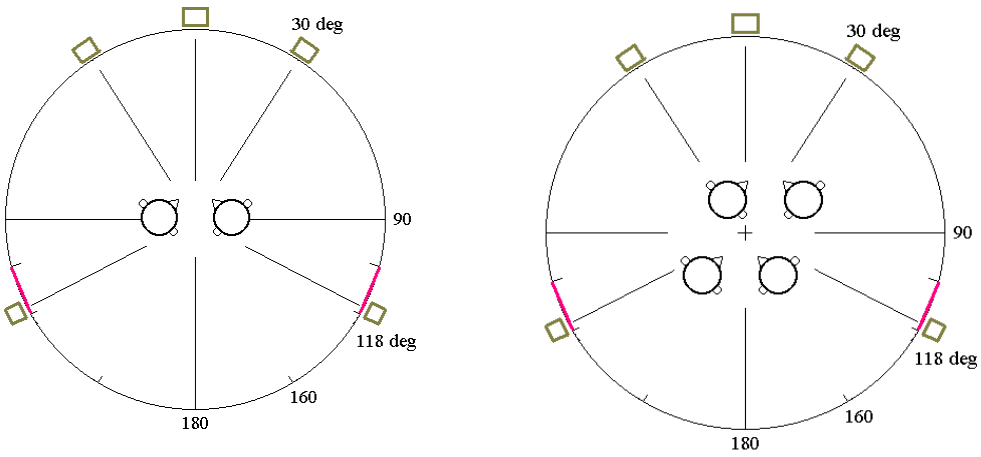

Figure 2: *What happens to our careful layout when we listen with a friend - or even with friends? Will someone - or everyone - be disappointed?*

The solution - or the only solution that interests me - is to make a recording that works well for both single listeners and groups. We have to eliminate the "listening line" of

standard two channel stereo, and the "listening point" of the standard surround layout. We must have a "listening area" and a large one at that.

To make a large listening area we must abandon the notion that the listener is in a specific spot, but this notion is deeply ingrained in our methods of sound recording. Many of these methods rely on time delay to provide horizontal localization, and time delay does not work unless the listener is centered between the speakers.

Using the center speaker correctly requires that the phantom image of the center is at least 6dB softer than it is with no center speaker. There IS no stereo technique that has this property. It would not work for stereo. Thus ANY "main microphone" technique we have learned to use is now useless.

So… we will talk about how to make a terrific recording without assuming the listener is centered between the speakers. Perhaps, when we are through, our recording will be even better when the listener is centered. But this is not our goal.

Why bother to use surround sound - isn't stereo good enough?

Mono recordings can be wonderful, and many car radio systems have so little stereo separation they might as well be mono. The music still comes across. Why do we bother to use two channels, let alone five or more?

Basically, there are two reasons to use two or more channels:
1. Horizontal localization
2. Enlarging the apparent size of the playback room – “being there”

The recording and reproduction of horizontal localization has been well studied - although nearly all of this work assumes the listener is centered between the loudspeakers. For most sound engineers enlarging the apparent size of the playback room is pretty mysterious. Yet at least for the author, it is the sense of "being there" that is the most important.

Those of us raised on mono remember when stereo was introduced. There was something wonderful about a stereo recording that you could hear anywhere in the room - and sometimes even in the next room. The stereo recording changed the acoustics of the listening space. The room became larger and more enveloping.

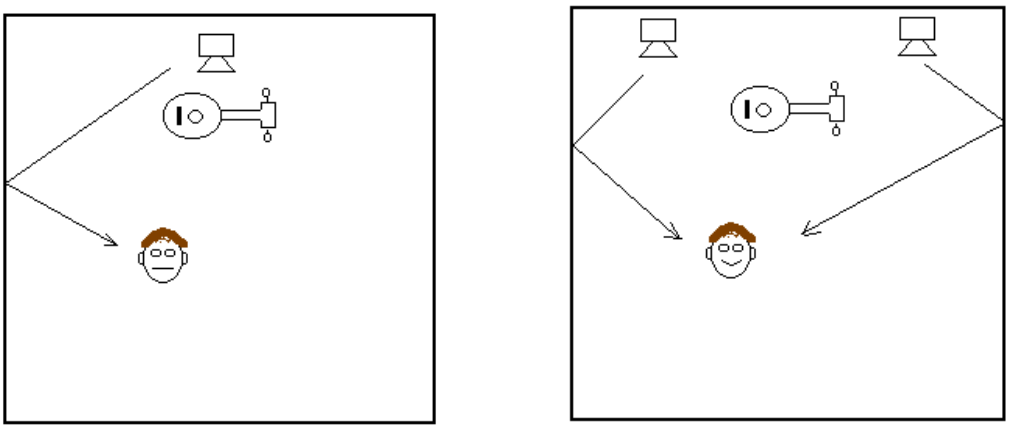

Figure 3: *With a single loudspeaker there is no possibility of reproducing the spatial properties of the original space. With two – plus a recording with decorrelated reverberation – a few of these properties will come through.*

Why does the room sound larger? This question is not trivial. (It has taken me 30 years to answer it to my own satisfaction.) To answer it you must know why a large room sounds large, and why a small room sounds small. Then you must understand how the perception of a large room can be transmitted to a listener within the confines of a small room. But the basics of the answer are simple: if the reverberation in a two channel recording is not correlated between the two channels, and if this reverberation is reproduced in a small room through two separated drivers, then some of the directional fluctuations in the original reverberation will be heard by the listener.

To give a modern example, let's start with a multitrack recording, made with a main microphone and an array of accent microphones. To this we add hall sound from a widely spaced omni pair. Now make two different two channel mixes, one as you would regularly do it, and one with all the microphones - except ones for the hall - panned to the middle.

While these recordings will sound different, the almost mono one will still sound pretty good to a listener who is not centered between the playback loudspeakers. In fact I have heard many commercial recordings that sound very similar to this mono mix!

Now switch the hall microphones to mono. Instantly the sound in the room collapses. We are back to the 1950s. The music may be wonderful, but it is crammed into a small space, and much of our involvement with it disappears. I contend that stereo recordings sell better than mono ones because they reproduce a little of the sound of the hall, not because they put the violins on the left and the cellos on the right.

For reproducing the sound of a hall, five speakers are better than two - and seven are better than five. Professor Boone in Delft has shown that at least 8 plane waves from

different directions are needed to reproduce the diffuse field of a concert hall - and this is in just two dimensions.

Recordings of music in surround show the importance of hall sound in another way. Most music mixers (particularly of classical music) use the surround speakers only for reverberation. They do not use them for the direct sound at all. “Of course”, they say, “in a concert the music is up front, so the speakers behind you should only reproduce the hall”. I don't want to argue with this opinion at this time. (Although it seems pretty silly.) I mention it only to demonstrate that some people think it is worthwhile to add two extra channels just for the hall sound.

But the major reason to use more than two channels is to create the "listening area" instead of the "listening line". A five channel system - when the mix is properly made - can sound quite good off-axis. Even a 5 channel matrix system, working from a conventional two channel CD, can enlarge the listening area enormously.

How do we hear sound images, and how do we hear the hall?

Human hearing is capable of creating an auditory image in which at least the horizontal direction of various sound sources is localized. However this aspect of sound imaging is dominated by our visual perception. Almost always we hear a sound source in the direction that we see it. When we hear a recording we tend to localize sound sources where we expect to hear them. We hear the violins on the left and the basses on the right because that is where we have come to expect them.

With very careful listening to a recording that has been made with a technique capable of good horizontal imaging, we can sometimes detect where the sound sources actually were. But this process is enormously aided by experience - and possibly a photograph on the album cover.

Sonic images and streaming

But there is another dimension to sound perception that is separate from horizontal localization. That dimension is sound streaming. We are capable of separating sound sources from each other - even in the absence of localization cues. For example, we can usually easily separate an oboe from a flute, and a flute from a violin, even though they play in the same register. The melody from the oboe will be heard separately from the melody of the flute. Both instrumental lines form a sound stream - just as the words of a particular person in a party form a stream. These streams are examples of foreground streams. They carry specific and often different content, or meaning, and we can choose to listen to one while excluding the others. Listening to more than one stream at the same time can be difficult but worthwhile - such as following the various voices in a Bach fugue. We can think of sound streams as similar to objects in the visual field, except that their space includes time as a dimension.

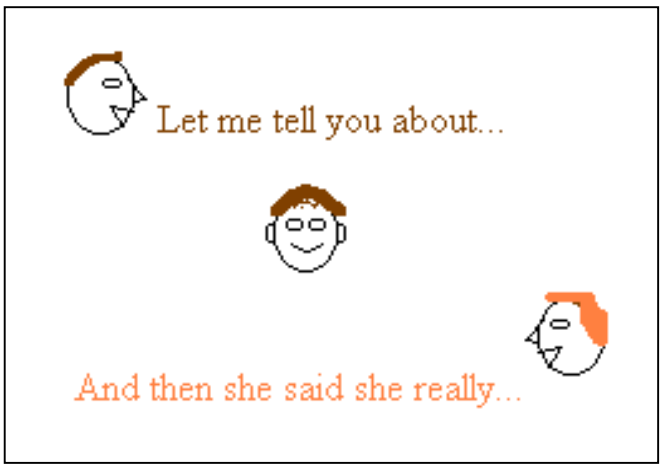

Figure 4: *We can separate the words from one person from the words of another – even without spatial cues – if the two people have different timbre. The words form separate sound streams.*

We can separate foreground streams because sound elements (phones) in one stream do not overlap sound elements in another stream. In other words, we hear the syllables (phones) from one speaker in the gaps between the syllables of the other speaker. In music it is also helpful if the streams seldom overlap in frequency. Where there is sufficient overlap, in either time or frequency, separation becomes impossible. Note that foreground streams are separated from the total sound field by detecting the starts and stops of sound events. Sound events (phones) are small bursts of sound from a particular source. Thus foreground streams are composed of little pieces of sounds. Although the stream may be perceived as continuous – particularly when some people I know talk - one can easily tell that the events in the stream are definitely not continuous.

But foreground streams are not the only type of auditory stream. Once the sound elements from each stream have been separated from the incoming sound, there remains a lot of sound that belongs to no particular stream. This is the sonic background - the sound between sonic events. The sonic background forms another type of stream – the background stream. Unlike the foreground streams the sound in the background stream sounds continuous, even when the elements in it are not.

For example, a person talking in a room generates reverberation. As long as the person is talking the reverberation sounds continuous, and has a constant level. However if we look at the sound signal on an oscilloscope it is clear that the reverberation is decaying rapidly between syllables. When the person stops talking the reverberation becomes a foreground stream, and is audible as a distinct sound event. Under these conditions it is easy to hear that it is decaying.

Separating foreground streams takes time. It is easy to recognize the beginning of sound events – but it is hard to find their ends. Human hearing generally waits 50ms after the apparent end of a sound event before definitely deciding it is over. Background sounds

that arrive during this 50ms waiting period are assigned to the foreground stream, not the background stream. This waiting period is the origin of the "Haas Effect".

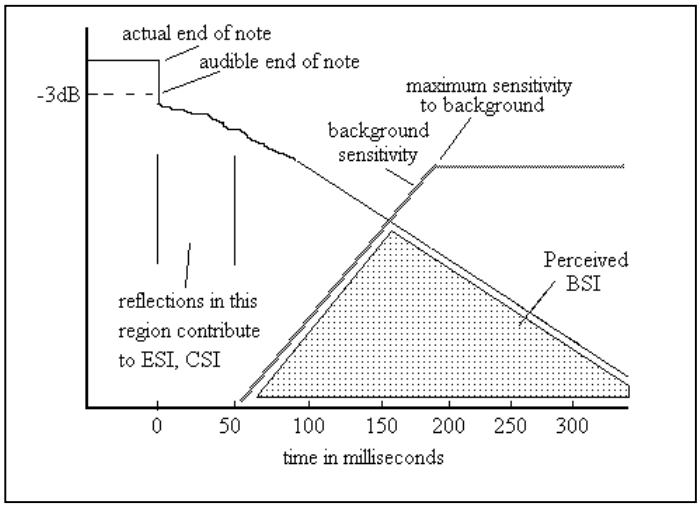

Figure 5: *Reflections that come within 50ms of the end of a sound event are combined with the foreground stream. Full sensitivity to the background stream occurs after 150ms.*

The background stream is vital to our perception of music, because it is the spatial properties of the background sound that give us envelopment. We detect the spatial properties of the background through fluctuations in the Interaural Time Delay (ITD) and the Interaural Intensity Difference (IID) between the two ears.

Fluctuations caused by reflections that arrive during the note and within 50ms of the end also produce a spatial effect, but since these reflections are combined with the foreground stream this spatial effect is largely one of apparent distance to the sound source.

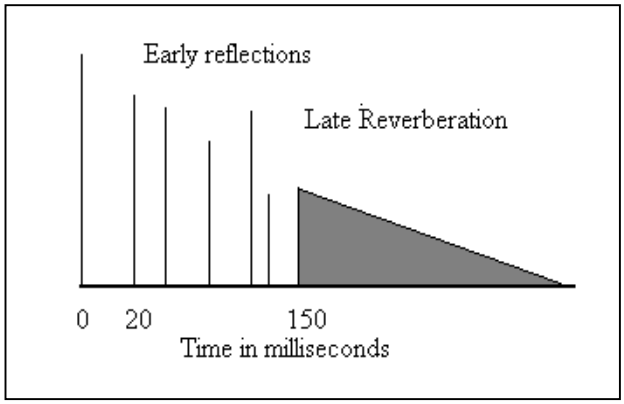

Figure 6: *The "typical" impulse response of a room seems simple, but this is not what we hear with music or speech.*

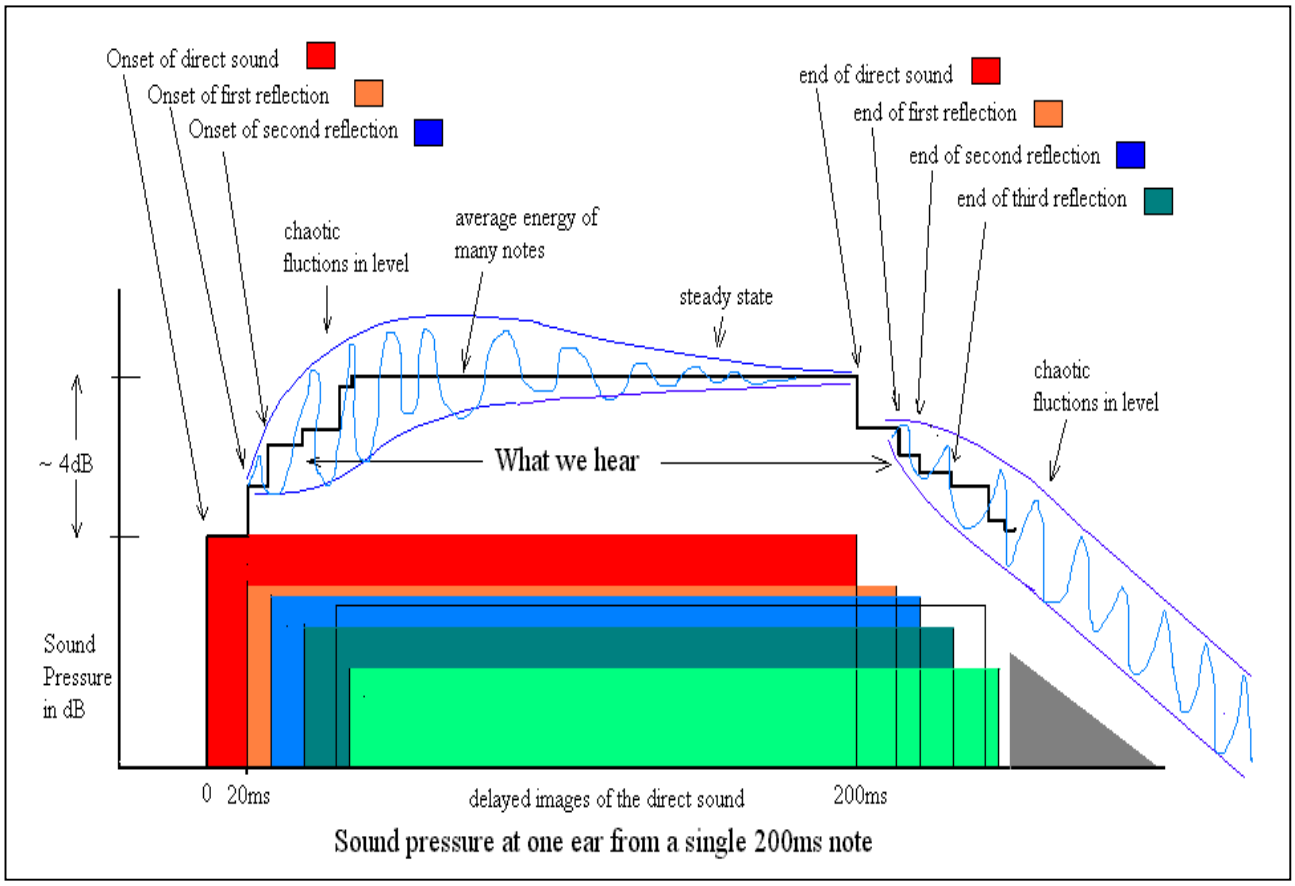

Figure 7: *With music or speech the impulse response is convolved with a note of finite length, creating overlapping images of the original sound, and significant fluctuations in sound pressure due to interference. We hear the rise and fall in level, and the difference in the fluctuations in each ear. Fluctuations during the note and for the first 50ms after the end of the direct sound are perceived as a sense of distance. Fluctuations 150ms or more after the end are perceived as reverberation and envelopment.*

So the acoustics of a particular room or hall can be divided into two separate perceptions, one associated with source distance, and one associated with reverberation and envelopment.

The two perceptions are triggered by very different time windows in the reflected energy. Small rooms – where the reflected energy tends to be concentrated in the first 100ms or less, produce very little sense of reverberation, even though the amount of reflected energy can be large. But notice that it is not the reflections themselves that are audible. It is the fluctuations in the ITD and IID that are audible, and this requires that the sound pressure at the two ears should be different. Thus only reflections that come from the side of the listener will be heard! We will call these reflections lateral reflections, even though above 700Hz the optimum angle for inducing fluctuations moves toward the medial plane.

As sound engineers we need to separately control the perceptions of depth and envelopment. Recordings with too little early lateral reflections sound too present, with the sound sources in the speaker or in front of the speakers. The various voices have no blend – they seem to occupy no common space. Such a recording can sound too close and too reverberant at the same time.

We have run many experiments where lateral reflections are added in various amounts to anechoic music. Surprisingly the ideal amount of early lateral reflections is nearly the same for every individual and for every type of music. The sum of the energy in the early lateral reflections should be between ½ and ¼ of the energy of the direct sound. Recordings with too much energy in the 50ms to 150ms region sound muddy. This time range must be carefully minimized.

So it turns out there is an optimum profile for the reverberation in a recording, and it is not the profile produced by most rooms. We can create this profile through careful microphone technique. Often leakage between microphones in a multimicrophone array can produce it - or the interaction between non-delayed accents and a main mike.

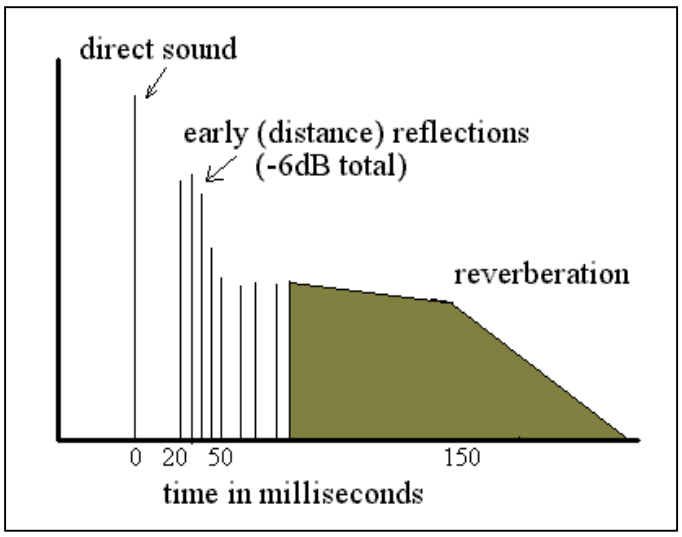

Figure 8: *The ideal reverberation profile for recordings. A strong early lateral field for producing a sense of distance, a minimum of energy in the 50-150ms region, and adequate reverberant energy after 150ms.*

Where do we put the speakers and how many do we need? I just gave a talk on this subject for an AES workshop – the slides are on my web page. To answer – five speakers are a lot better than two, and seven are better than five for many rooms. Both Tom Holman and I attempted to say what was the maximum number, The number above which there was little to be gained. We both came up with about 10 or 11.

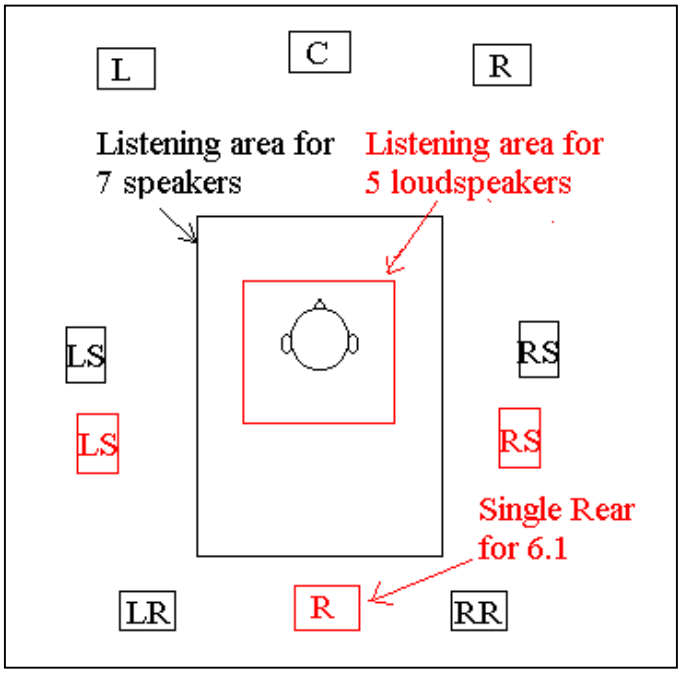

Figure 9: *difference in listening area with 4 rear speakers vs 2 rear speakers (highly room dependent.)*

Where should we put the speakers? The front speakers are placed in positions determined largely by the film industry, although there has been a lot of work on the best possible locations for three front speakers.

The position of the side speakers in 3/2 surround is a compromise. A position at 90 degrees from the front is optimal for reproducing lateral reflections below 700Hz, and also produces the most envelopment. A position of 150 degrees or more is more effective for discrete sound effects. The current 3/2 standard – with the side/rear speakers at 110 to 120 degrees – is adequate for producing envelopment, but is not back far enough to produce a satisfactory rear sound. If the side/rear speakers are direct radiators the listening area is not large.

The reason that speakers further back sound more interesting is that human Head Related Transfer Functions (HRTFs) have a very different spectrum for a sound coming from 150 degrees from the front than they do for a sound coming from 120 degrees. Thus it is difficult for speakers at 120 degrees to make a sound that sounds fully behind you. If you pan from left rear to right rear, the sound goes through the head rather than behind the head. There is considerable research that indicates that sounds from 150 degrees or more are more effective, both for producing high frequency envelopment, and for producing exciting sound effects, than sounds at 120 degrees.

The obvious solution to this problem is to add at least one more speaker. Dolby EX adds a single speaker at the rear of the listener, and drives it with a Pro-Logic matrix from the

two rear channels. This solution is not optimal. A single rear loudspeaker is usually very difficult to place in the room, and it reduces, rather than increases, the effective listening area. Also a sound effect panned to this rear speaker may often sound like it comes from the front. Front/back reversals are very common when a loudspeaker is directly behind, and rare when the speaker is at 150 degrees or so. A speaker on the medial plane also cannot produce envelopment. Thus the EX configuration may reduce the hall sound rather than increasing it.

(In our AES talk Tom Holman pointed out that some of these objections were reduced if the single rear loudspeaker is a dipole. However in this case a sound effect panned rear would tend to lose focus.)

A better solution is to use four rear speakers, and drive them with a 2/4 matrix that preserves full separation for uncorrelated signals. We have been providing this solution for some years now, and it is well proven.

How many channels are needed on CD's and other media?

Just because we want 5 or 7 loudspeakers does not mean we need 5 or 7 channels on the disk. As we pointed out, Dolby EX and the Lexicon 5 to 7 matrix decode the two rear channels of a 5.1 mix to 3 or 4 loudspeakers. In my opinion matrix technology is ideal for the rear channels. Decoding two rear channels into four can be done very successfully.

For several years we have also been marketing a 2 to 7 decoder, that converts standard two channel recordings into surround. It works well. In fact I believe that the 2 to 7 matrix produces a result that is far superior to two channel stereo on almost all material. The sound is more enveloping and has a much larger listening area. I am not alone in this opinion. I am joined by a large number of high-end audio critics.

But my own listening tests have convinced me that discrete recordings can sound better. Very often if we take a commercial 5.1 channel recording and convert it to stereo using our 5 to 2 encoder, and then play it back through our 2 to 7 decoder, the result is superior to the original discrete recording. This need not be the case. There are a lot of sound mixers out there who refuse to use the center channel for anything, and who's idea of how to use the rear channels is pretty primitive. Perhaps they listen only in the sweet spot! Passing the sound through the matrix system actually improves it. But I have also heard mixes (and made mixes) where the discrete version is noticeably better – and this is what I would expect.

However there is a problem with the 5.1 standard. The LFE channel was intended only for high energy effects at very low frequencies. But many sound mixers are using it as if it were a subwoofer channel. They put in all the bass below 80Hz. This is a bad idea.

It is very easy to show that if you make the reverberation monaural below 80Hz there is less envelopment than if you keep the reverberation decorrelated and in stereo. If we are

going to separate the octave between 40Hz and 80Hz, we should use two channels, and two low frequency drivers. So we need a 5.2 standard – or we should just go back to using the main channels for full range reverberation, as they were intended.

This observation applies strongly to bass management as it is currently implemented in most receivers. We really need stereo bass management, and two low frequency drivers, at least to cover the range between 40Hz and above. So far, stereo bass management has not caught on.

So for the time being I am happy with the choice of 5.1 or 5.2 channels for distribution media. But the differences between discrete 5.1 and the output of a 2 to 5 matrix are small, and the absolutely optimum 2 to 5 decoder has yet to be made. Stay tuned.

Matrix comparisons:

Although we believe that matrix systems are capable of excellent performance while converting standard two channel material into surround, there are several such systems coming into the market, and they differ substantially in their performance. The systems include Pro-Logic 1, Circle Surround, Pro-Logic 2, DTS surround, and Lexicon Logic 7. Here are a few things to check for:

1. The width of the front image.

Alas, there is only one possible way to generate a center channel signal, and that is by summing the left and right input channels. Clearly if we reproduce a signal that is supposed to be on the left only through both the left and the center speaker, the sound image will move toward the center. It is impossible to avoid reducing the width of the front image without turning the center channel off.

All the current decoders turn the center off when there is a strong signal from just the left or the right channel. The question is – what do they do if you have a strong signal in the left, and a different strong signal in the right? (This kind of signal is the rule in both classical and popular music.) Alas, you better turn the center for this condition too, or the front width will be seriously reduced. This is just what we do – but we turn the center back on (very cleverly) when there is a vocalist. All the other systems keep the center strong under these conditions, and center width is seriously compromised.

2. The decorrelation of the rear channels

Another major difference between systems is in the decorrelation of the various outputs when there is a reverberant signal. In three of the currently available systems the rear channels are monaural under these conditions, although one of these adds a small delay modulation to one channel. In one of them (not ours) the rear channels are negatively correlated – partly mono and out of phase. This negative phase relationship is easy to see on a phase meter, and can also be easily heard in practice, as there is a strange nulling sensation in the sweet spot. As expected, the decoders that reproduce reverberation in

monaural have a build up of energy from behind when one is exactly centered between the surround speakers, and envelopment is poor.

3. The ability to reproduce stereo in the rear.

This difference does not become noticeable unless one has encoded material – such as a Dolby Surround film or music recording, or a Logic 7 encoded 5.1 track. Many such mixes have applause, or sometimes stereo music, coming from behind the listener. The matrix should be able to reproduce this signal in full stereo from the rear channels, and only one of them (guess which) does. The difference is very obvious.

So – on the subject of matrices – Caveat Emptor!

Mike technique:

We could spend hours on this subject. So here we give just the high points. Our goal is to make recordings that work well for listeners off-axis, and that produce the maximum feeling of envelopment from a 5.1 channel system.

If we want good localization off axis we must use amplitude panning and not time delay panning for all sources in the front. The panning should be from center to left and from center to right - not from left to right. We want to use the center speaker – which means that a sound image that comes from the center should be at least 6dB louder in the center speaker than it is in the left and the right. These two requirements eliminate most single point microphone arrays. It is very difficult to meet these criteria with pressure-gradient microphones, and it is impossible with omnis.

There is a further requirement on the microphone technique used for both the front left and right, and the two rear channels. The reverberation they pick up should be decorrelated. This means the left and right main microphones must use one of the combinations of patterns and angles given in Figure 10, or that they should be separated by at least the reverberation radius of the room.

The reverberation picked up by the rear microphones should also be decorrelated – and should be decorrelated with the front channels. In practice this means that the rear microphones must be separated from the front microphones by a distance of at least the reverberation radius.

If we eliminate all stereo main microphone techniques, and all closely spaced arrays, what is left? The situation is not a bleak as it looks. Most practicing engineers already space their rear microphones away from each other and from the front microphones. They also are already expert in the careful use of multi microphone technique. They use this technique for a simple reason – it works well in practice. I am only suggesting that it also works well in theory. Very few of the major recording engineers try to record surround (or stereo) from a single array. This method seems to be reserved for schools and broadcast stations.

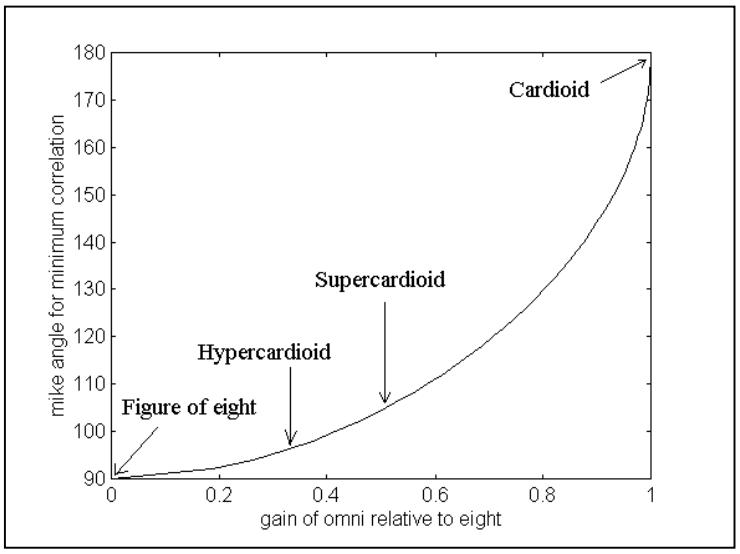

Figure 10: *Optimum mike angle for reverberation decorrelation for various microphone patterns. Note that Hypercardioids are decorrelated with an angle of 98 degrees, and supercardioids with an angle of about 107 degrees. Two cardioid microphones will be decorrelated only if they point opposite directions.*

It is possible to get the appearance of decorrelation by separating two microphones in a reverberant field. The appearance is deceptive. The decorrelation is highly frequency dependent.

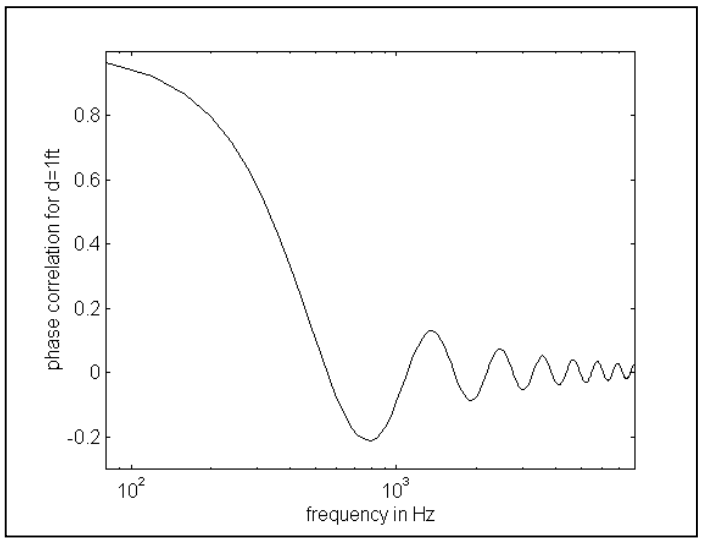

Figure 11: *The amount of correlation between two microphones separated by 25cm in a reverberant field. Calculated in three dimensions. Note the high correlation at 100Hz, and the negative correlation at 800Hz. The separation and the frequencies vary inversely, so a pair separated by 2.5m would have a negative correlation at 100Hz. (But only if the reverb radius were greater than 2.5 meters.)*

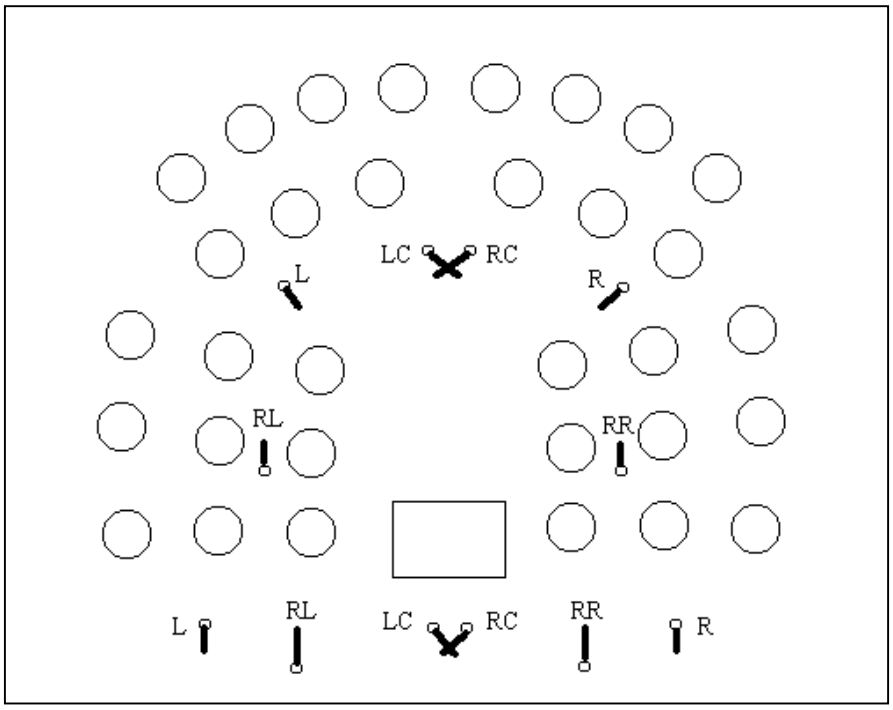

Figure 12: *Example of a simple microphone array for a large orchestra. Two supercardioid pairs are used to pick up a left center and a right center channel, along with spaced cardioid microphones for the left and right channels. The two mikes over the string sections (RL and RR) can be mixed into the back channels to create a "conductor's perspective." Two cardioid microphones are shown pointing rear(the RL and RR mikes behind the conductor) to pick up the hall sound. These are about 5m behind the conductor.*

The Market for surround

Marketing is not my field of expertise. I can however say that our home theater systems are selling well, and at least in the USA it is hard to find an audio system that is not at least a 5 channel Pro-Logic system. Two channel stereo is now a very small market. So far only Dolby Digital has a significant market for discrete surround recordings, and this is almost entirely for films.

Last I heard, discrete music mixes on DVD were held up for several reasons - one of which is cost. It turns out to be very expensive to produce a DVD surround disk. Just the extensive liner notes and graphic material cost more to produce than a two channel CD. A mass market for music surround seems difficult to develop.

But there is hope for at least some types of surround. For the last several years I have been working in combination with Harman Motive on Logic 7 matrix systems in automobiles. The killer application for surround systems is in cars. Remember I tried to show how a sense of a larger space could be created through a combination of early and late lateral reflections? Logic 7 matrix can do this in a car, and the results are dramatic.

Automobiles are an impossible space for sound. No listener is anywhere near a sweet spot, and the playback room is tiny. There is no better place for a surround system with an enormous listening area, and a very high ability to recreate envelopment. When you switch between two channel stereo and Logic 7 surround in a good listening room there is a noticeable and worthwhile improvement. If you do the same experiment in a car the difference is night and day.

With the surround on the front image falls into place, and the individual voices move from about 5 inches in front of your face to beyond the windscreen. At the same time the side walls of the car seem to disappear. Wow!

We have been modifying cars for some time to demonstrate the new technology to the major auto manufacturers. They love it. By the year 2005 there could be as many as a million cars with Logic 7 matrix systems on the roads. This is a market - a sizeable market - for matrix encoded surround CDs. Will record companies jump to release their catalogs in matrix surround? It seems unlikely, but it could happen. We know that 5 to 2 encoding produces two channel CDs that sound very good on standard two channel equipment, and yet play through a Logic 7 decoder in a way that is difficult to distinguish from the original. The product looks, costs, and markets identically to a standard CD, and can be broadcast. If enough people want it - and I think there will be a lot of people who will - we could all be surprised.

Conclusions:

Academic discussions of surround sound often get bogged down in methods that require the listener to be at a specific point, and many music producers insist on making mixes that only work at this point. This use of multi channel technology seems misguided. It is possible, and highly desirable, to make mixes that yield a large listening area. A large listening area can be created by careful use of the center channel, and by maintaining high decorrelation in the reverberation applied to the front left and right, and the two rear channels. Ironically, a good 2 to 5 matrix decoder achieves both of these goals from standard two channel material, and often sounds better than a poorly done discrete mix. Although a market for multi channel music surround material seems to be developing very slowly, the killer application for surround systems seems to be in automobiles. A market for stereo compatible matrix surround recordings may appear in a few years.