David Griesinger

Lexicon

3 Oak Park

Bedford, MA 01730 www.world.std.com/~griesngr

Räumliches Hören in Theorie und Praxis: Wie ergänzt man Tiefe und Halligkeit mit künstlichem Nachhall ohne Beeinträchtigung der Deutlichkeit

The Theory and Practice of Perceptual Modeling - How to use Electronic Reverberation to Add Depth and Envelopment Without Reducing Clarity

Introduction

This is a paper about the use of artificial reverberation in music recording. Although there is an emphasis on multichannel recording, the methods described work equally well in two channels. But electronic reverberation is not the solution to every problem. A recording that already has too much reflected energy will remain muddy, and a recording with inadequate separation will have poor image stability no matter what we do. By necessity we must first examine basic recording technique.

The advent of multichannel recording has brought the problems of conventional sound recording into sharp focus. Techniques we have used for years in our two channel work do not make satisfactory surround recordings, and understanding the reasons for this failure leads to a healthy questioning of our basic ideas.

The most basic of these questions is: WHO will listen to the recording, and how will they listen?

1. Is the recording going to be heard by a single individual or a group?
2. If by a single individual, does the individual sit motionless at a single ideal listing point?
3. Or is the individual allowed to move around the room?
4. If the recording is to be heard by a group, how big is the group?

Once these questions have been answered there are two others:

1. A wide front image can greatly help our ability to clearly hear individual musical lines. Do we want maximum clarity and intelligibility in our recording, or do we want to emulate a particular performance as heard from the hall?

2. How reverberant will the most likely playback room be, and how close will the listeners be to the loudspeakers?

I cannot answer these questions for all people and for all recordings. But my personal preferences are very clear, and will form the basis for this paper.

Individual listeners versus groups

Before we can begin recording we must know whether the recording will primarily be heard by a single listener at a single point, or by a group of listeners. The answer will determine many aspects of the recording method. In film recording the answer is obvious – the recording will be heard by a group. Yet for music recording we have answered the question without posing it - we have made individual listening at a single point the STANDARD.

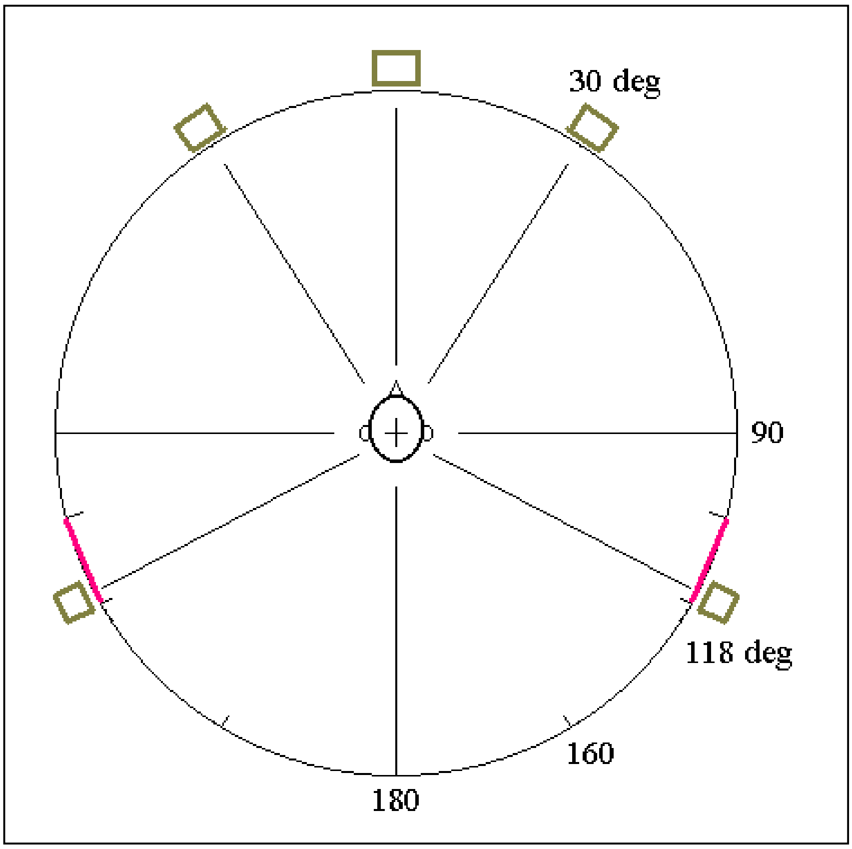

Figure 1: *The standard playback speaker arrangement for a 5 channel recording. Note that there is a single listener at a point equidistant from all loudspeakers. There is no listening area, only a listening point.*

Alas, our customers do not use this standard. 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. The market has spoken, and our individual oriented standard is out.

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 surround speaker layout is a bit silly for this task.

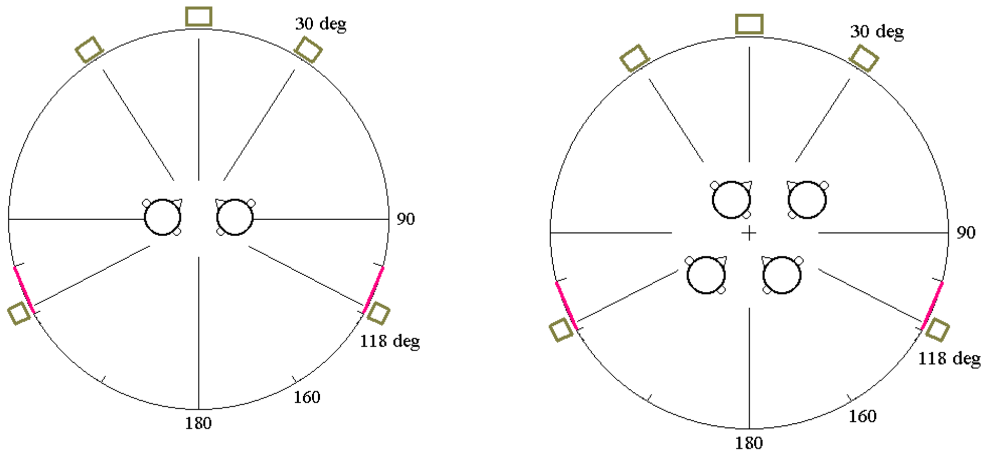

Figure 2: *What happens to our careful layout when we listen with a friend - or even with friends? Will someone - or everyone - be disappointed?*

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.

Mono was a democratic medium. It sounded equally bad over a very large listening area. Two channel stereo sounded better than mono throughout the room, but it only localizes well on a line equidistant from each speaker. Five channel surround sounds better than two channel stereo over a wide listening area, but according to our standard, it localizes well only at a single point.

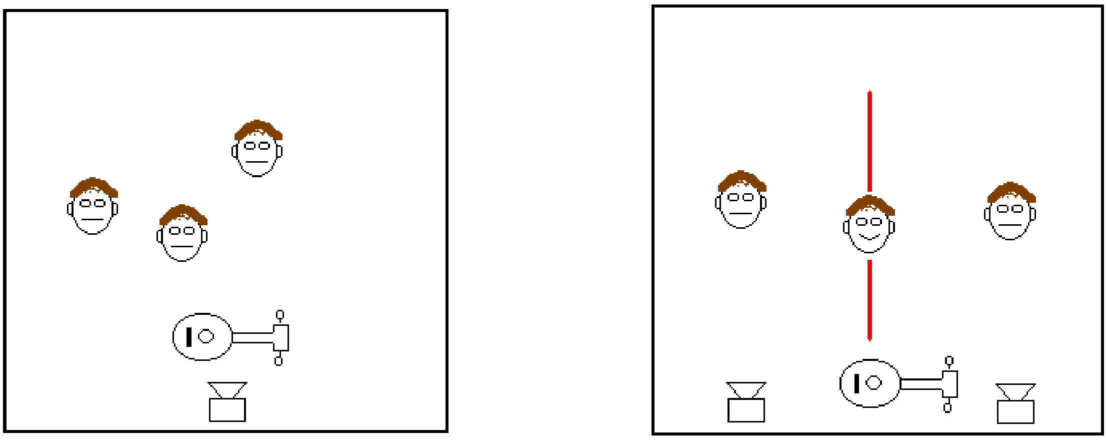

Figure 3: *Mono is democratic – it sounds equally poor everywhere. Where there is a strong center phantom image, stereo localizes well only along a line between the loudspeakers.*

We can do better. Techniques exist that create a wide and stable front image over a wide listening area, and at the same time provide a natural sense of depth and high envelopment. These techniques work even with two channel recordings.

For example, if we record an extended sound source with sufficient amplitude separation, the listening area is wide even in stereo. Adding a center loudspeaker allows us to widen the listening area for all types of recordings, even recordings with a strong center image.

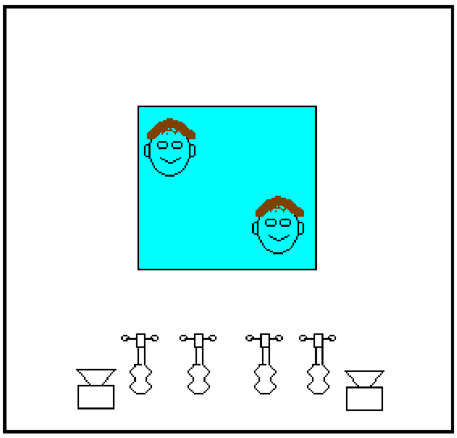

Figure 4: *Two channel recordings of extended sources with high amplitude separation – where instruments on the far left are reproduced almost entirely the left loudspeaker – will localize well over a wide listening area.*

This fact was well demonstrated at a recent conference in France, where surround recordings were played to about 200 people in a small auditorium. Most of the recordings localized poorly unless the listener was precisely centered between the front loudspeakers. Otherwise the front image collapsed into the front left or right loudspeaker, depending on which side of the audience you were sitting. The center and surround speakers did not seem to create any significant improvement over two channel playback. But on a few of the recordings the front image extended all across the stage, no matter where you were in the auditorium. The difference was obvious and electrifying. These were recordings with good amplitude separation between the three front loudspeakers.

Unfortunately the concept of the “sweet spot” is deeply ingrained in our methods of sound recording, particularly in the concept of the "main microphone". There are three problems with main microphones - and they are insolvable with current technology.

1. Many "main microphone" methods rely on time delay to provide horizontal localization, and time delay does not work unless the listener is centered between the speakers. To be specific: a technique such as ORTF, where two cardioid microphones at an angle of 120 degrees or so are spaced apart by about 18cm, has too much leakage between the left microphone and the right microphone to produce a wide front image. We can very usefully increase the amplitude separation of an ORTF array by using supercardioid or hypercardioid microphones. This technique works well for a two channel recording, but has too much phantom center for surround. Techniques that use omni directional microphones spaced less than two meters apart produce even less amplitude separation.

2. Using the center speaker correctly in a surround recording requires that the sound pressure from the phantom should be at least 3dB less than the sound pressure of the center speaker. Thus the leakage of a center sound source into the left and right speakers should be attenuated by at least 6dB. There is NO stereo technique that has this property. It would not work for stereo. Thus ANY "main microphone" technique we have used for two channel stereo is useless for surround.

3. To provide a realistic impression of the hall over a large area, the reverberation in the front left and right speakers should be uncorrelated - which means completely different. Furthermore the reverberation in the rear speakers should be uncorrelated, and at the same time it must be uncorrelated with the front speakers. As we will see later, it is possible to pick up uncorrelated reverberation from two closely spaced directional microphones if the microphone angle is carefully chosen to match the microphone pattern. However the ONLY way to make the front reverberation uncorrelated with the rear reverberation is to space the microphones apart by at least the hall radius. This does not mean that the rear microphones must be spaced more distant from the source – they will be equally decorrelated if they are simply spaced to the side.

Main microphones and the Hall Radius

The problems of typical main microphones are even more profound. We are taught - correctly - that we cannot pick up the direct sound from an instrument with a microphone that is at a distance greater than the hall radius (critical distance). (The hall radius is the distance at which the direct sound and the reverberant sound are equal in power.) Yet we are also taught to use main microphones. These two teachings cannot be true at the same time!

Many halls have a hall radius of less than five meters, and a stage house typically has a hall radius of less than three meters. Most instruments in an orchestra are more than five meters from any main microphone position. If we made a recording with the main microphone alone, the result would be unusable. Nearly all the instruments will sound too far away. Every practicing engineer knows this fact.

So we add "support microphones" to pick up the missing direct sound and add it to the recording. These microphones are close enough to the instruments to pick up the direct sound without too much reflected energy. But if the support microphones are supplying the direct sound – why are they called “support”? They are really the main microphones!

We can easily prove the importance of the “support” microphones. When the main microphone is separated from a particular instrument by several hall radii, The direct sound is almost inaudible. The total loudness for a particular instrument comes from the sum of the reflected energy. In practice we are told to bring up the level of the “support” microphone until we easily hear it. In this case the direct sound from the “support” will be stronger than the direct sound in the “main” by many decibels. If fact, if the recording is to have the clarity and separation most conductors demand in a commercial recording, the energy from the “support” microphone from any ONE particular instrument will be greater than the total energy in the “main” microphone for this instrument.

In fact, in nearly any successful commercial recording the front image is supplied entirely by the “support” microphones, and the main microphone merely supplies some of the reflected energy and reverberation. The main microphone increases the apparent distance and blend of some of the instruments – but it does not work well for all of them. There is a better way to achieve the same result.

So… we will talk about how to make a terrific recording without assuming the listener is centered between the speakers. That means finding solutions which work within the constraints listed above. 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 – but nearly all of this work assumes the listener is centered between the loudspeakers. If we want the localization to be effective over a large area, the answer is simple: amplitude panning works, and time delay panning does not.

We will localize an instrument to a particular position over a large listening area if that instrument is louder in the speaker closest to its apparent direction. Thus to reproduce a sound from the left, we play it from the left loudspeaker, and not from the right loudspeaker. If we want a source between center and left, we pan the source between the center speaker and the left speaker, and do not reproduce it from the right speaker. What is complicated about that?

And yet this is not common practice by engineers who use a “main microphone”. Consider a "Decca Tree" consisting of three omnidirectional microphones separated by one meter. Let us record a single instrument – a violin on the left. We will reproduce this sound through a three channel system, for a listener at the far right.

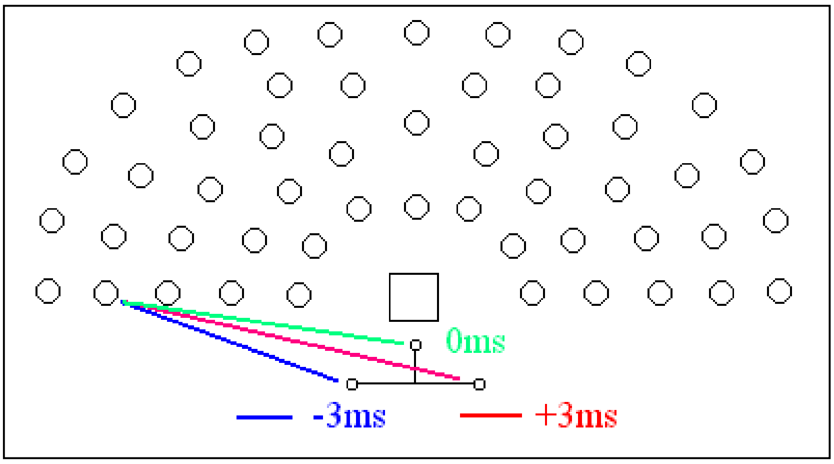

Figure 5: *An omnidirectional main microphone with a separation of 1 meter picks up a source on the left with roughly equal amplitude in all microphones, and a +-3ms delay difference.*

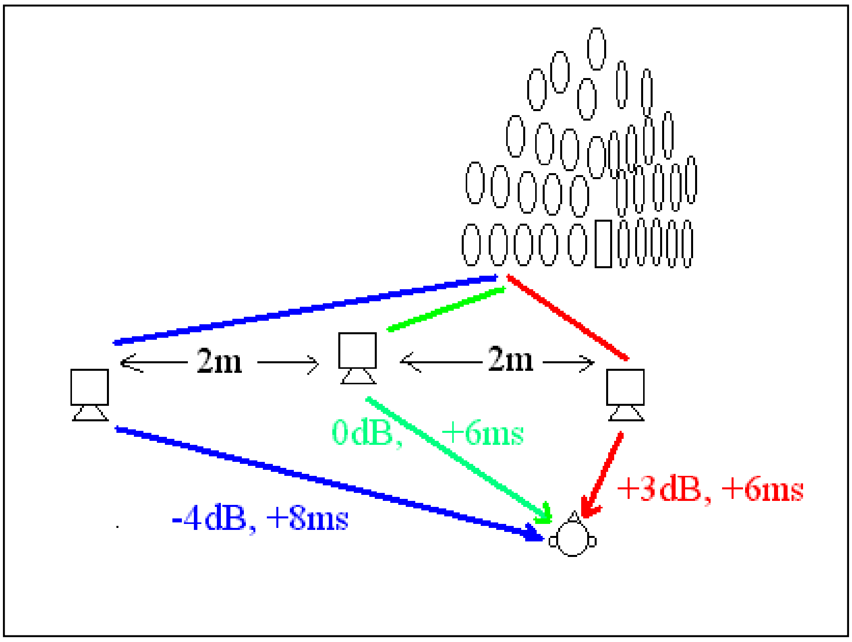

Figure 6: *A listener on the right will hear the entire orchestra from the right speaker. Signals from an instrument originally on the left will be louder, and arrive sooner, from the right loudspeaker.*

A source on the left of the orchestra is picked up by the left microphone first, followed in 3ms by the microphone for the center, and 6ms later by the microphone for the right. For an off-axis listener the amplitudes are highly biased in the direction of the nearest loudspeaker.

The time of arrival of the leading edges of each note will cause a significant error in localization, and the amplitude errors make the time error worse. The listener in our example will hear the entire orchestra coming from the right speaker. As the listener moves toward the middle of the playback area the orchestra will begin to spread into the space between the center speaker and the right speaker, and finally near the center line sound will appear to come from the left speaker. Unless you are near the center line the image is artificially narrow and poorly localized.

The collapse of the sound image into the nearest speaker is not an inherent property of stereo or surround. It is a result of our using delay panning instead of amplitude panning to create a sound image.

If we space the microphones more widely, an entirely different sound picture results. Now an instrument on the left is much stronger in the left speaker than it is in the right speaker, and the left speaker signal arrives at the listener first, not last. The combination is very effective. Now the orchestra spreads nicely between the speakers, even for a listener far off the center line.

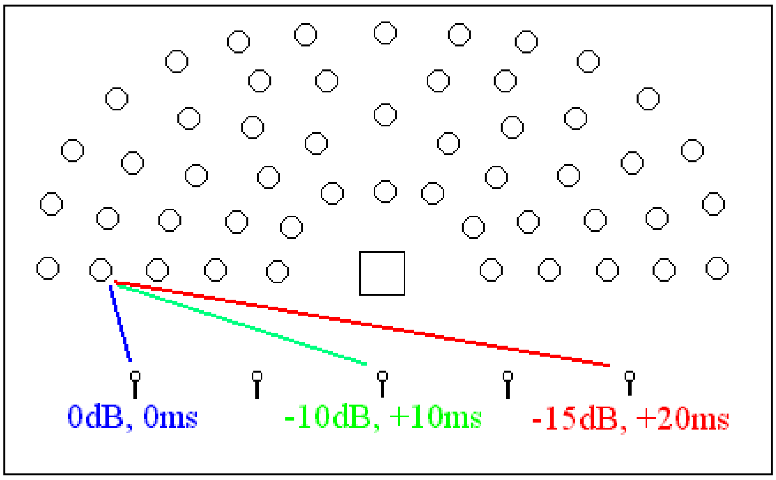

Figure 7: *Now use a spaced omnidirectional array. Notice there is an amplitude difference as well as a time difference in the signals received.*

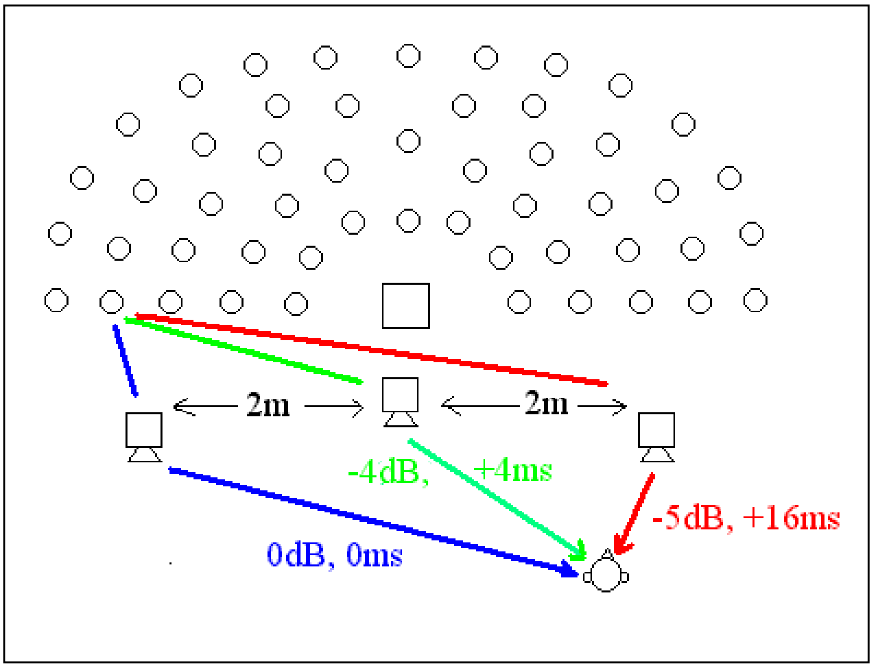

Figure 8: *With a spaced microphone pickup, an instrument on the left is heard coming from the left loudspeaker, even for a listener on the far right. Signals from the left speaker are earlier in time and stronger than the leakage signals from the other loudspeakers. The leakage is perceived as early reflections.*

The advantages of a "main microphone" array

If the listener is in the sweet spot most "main microphone" techniques reproduce horizontal localization well. In fact, they often can give a more evenly spaced image than a spaced technique as in figure 7.

But the major advantage for most users is the idea that these techniques produce a more natural sense the depth of the image. Sound sources appear to be behind the loudspeaker basis, and not in the loudspeakers themselves.

But the impression of distance or depth is not uniform across the image. Careful listeners will notice that instruments to the far left or far right of the microphone array often seem closer to the listener than instruments in the center, even though the instruments in the center are closer to the microphone.

The perception of the depth of a sound image is not a mystery. The perceived distance of a sound source depends on early lateral reflections, which means reflections that arrive at a listener from other directions than the direction of the source. A main microphone array can provide these reflections – but it only does it for some of the sources, particularly those near the center.

The major point of this paper is that you can achieve the same or better results through careful addition of early reflections generated electronically, and thus achieve both a natural sense of depth, and good horizontal localization over a wide listening area.

Enlarging the size of the playback room

So far we have been talking about horizontal localization. But horizontal localization is not necessarily the most important advantage of two channel or multichannel reproduction. Enlarging the apparent size of the playback room is probably more important to most of our customers.

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.

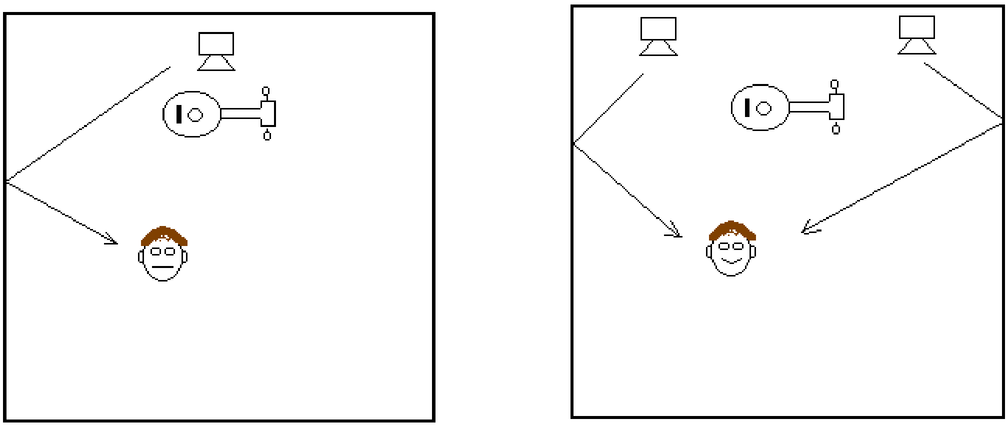

Figure 9: *With a single loudspeaker there is no possibility of reproducing the spatial properties of the original space. With two – plus a recording with decorrelated reverberation – a few of these properties will come through.*

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.

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, on some recordings we can 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.

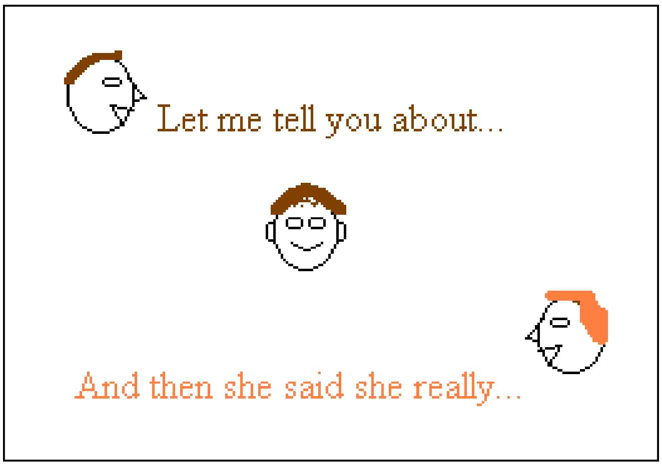

Figure 10: *We can separate the words from one person from the words of another – even without spatial cues – if the two people have different timbre. The words form separate sound streams.*

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 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 is perceived as 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".

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.

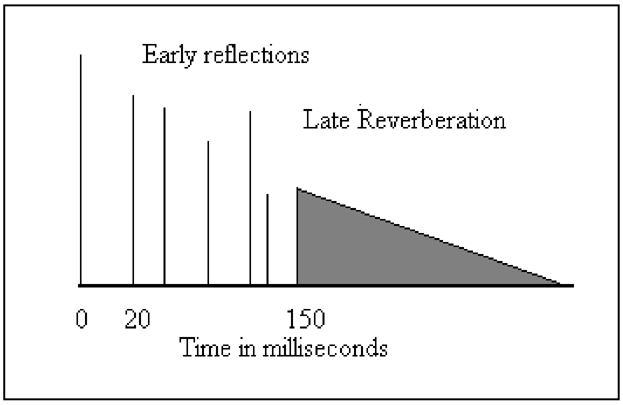

Figure 11: *The "typical" impulse response of a room seems simple, but this is not what we hear with music or speech.*

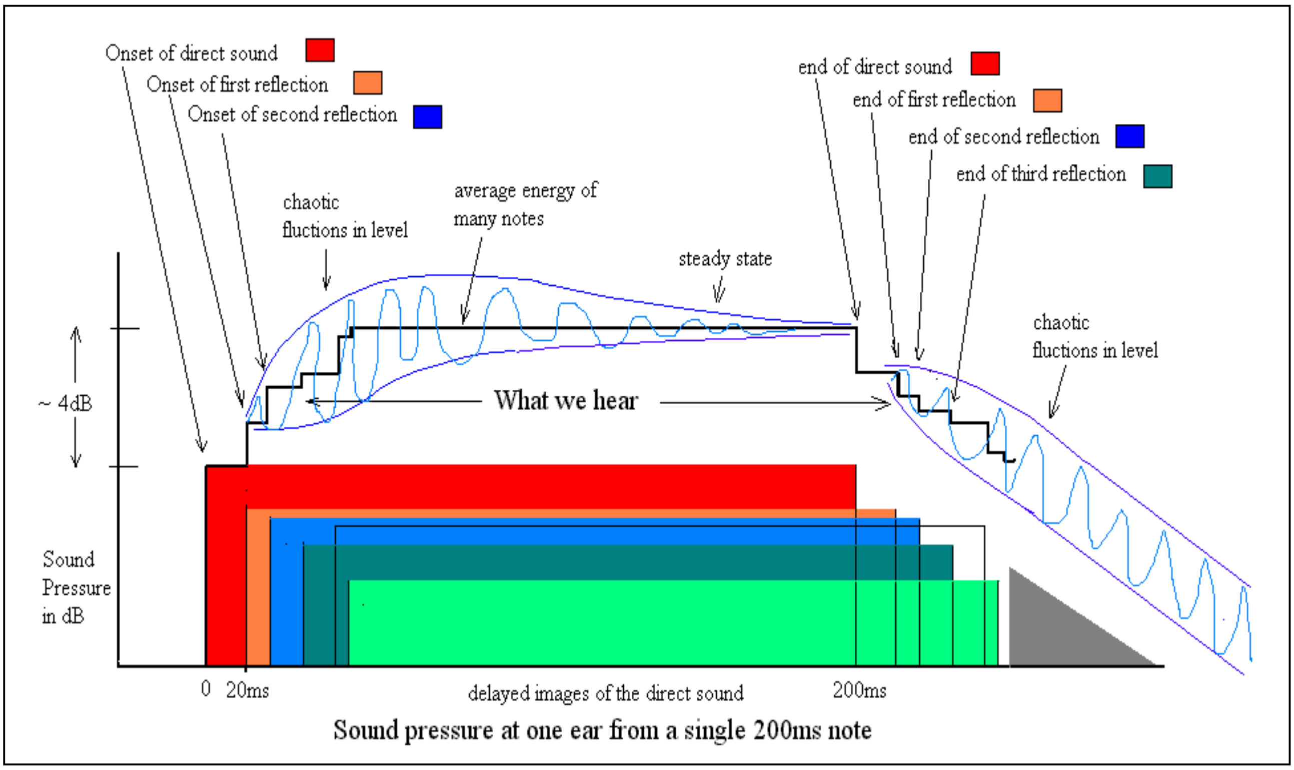

Figure 12: *With music or speech the impulse response is convolved with a note of finite length, creating overlapping images of the original sound, and significant fluctuations in sound pressure due to interference. We hear the rise and fall in level, and the difference in the fluctuations in each ear. 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.*

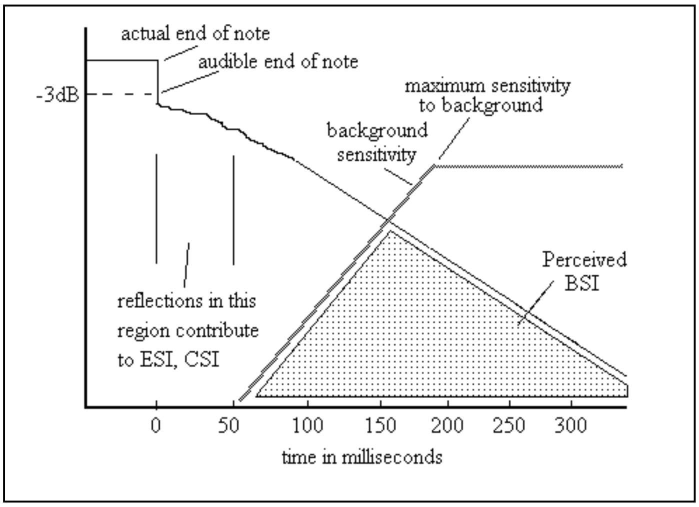

Figure 13: *Finding the end of a note takes time. Reflections that come within 50ms of the apparent end of a sound event are combined with the foreground stream, while the brain makes sure the note has really ended. Full sensitivity to the background stream occurs after 150ms.*

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 call the process by which we created this profile "Perceptual Modeling." We can create this profile through careful microphone technique. Sometimes, often by accident, leakage between microphones in a multimicrophone array can produce it - or the interaction between non-delayed accents and a main mike.

The ideal reverberation profile

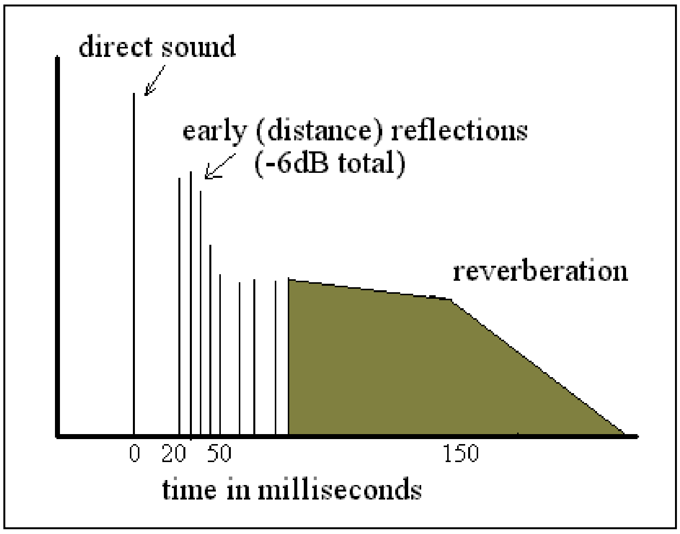

Figure 14: *The ideal reverberation profile for recordings. A strong early lateral field for producing a sense of distance, a minimum of energy in the 50-150ms region, and adequate reverberant energy after 150ms.*

With figure 14 we finally come back to the main topic of this paper – how we can make a recording with excellent envelopment, perspective, and depth without reducing clarity or intelligibility. We can do it by carefully adding decorrelated early reflections in the 20 to 50ms time range, minimizing the reflected energy in the 50 to 150ms region, and supplying sufficient decorrelated reverberation in the time range of 150ms and more.

The ideal direction for reflected energy

But there is one more essential aspect of the reverberation – the direction that it comes from. If a sound comes from in front of the listener, early reflections that come from the front, and in fact in any direction in the "medial plane" are inaudible. The medial plane is a plane surface perpendicular to a line between the ears. It includes the directions front, back, up and down. It includes all directions that do not in some way come from the side.

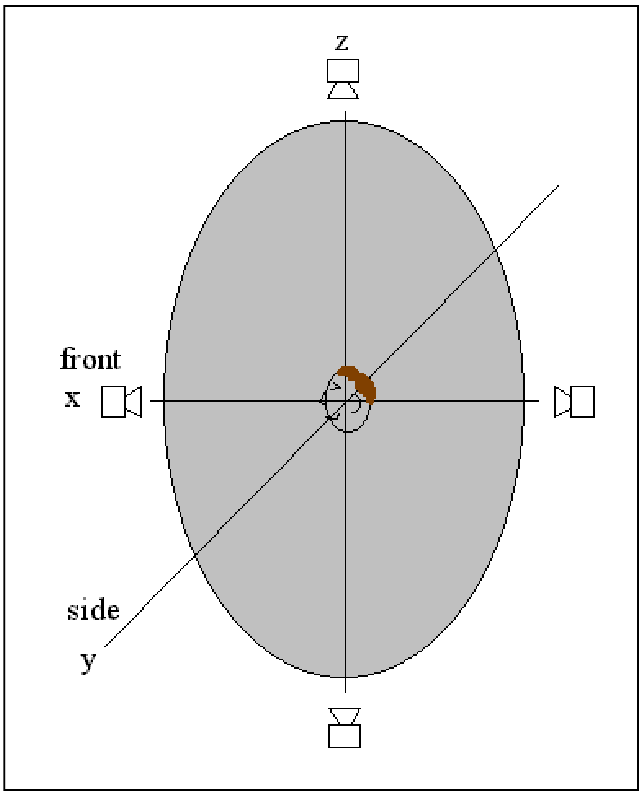

Figure 15: *When a sound source is in the front (along the x axis) early reflections coming from anywhere in the x-z plane (the medial plane) are inaudible or undesirable. Reflections coming from the side (the y axis) are audible and effective for creating a depth perspective.*

So if we want a sound to be perceived as distant we must add early reflections, and they must come from some other direction than the direction of the direct sound. This is quite difficult with conventional two channel techniques. Instruments far to the left tend to create reflections in the same channel as the direct sound. These reflections are masked. Early reflections in the right channel are usually not generated, so no depth perspective is created. These instruments always sound too close.

Should the reflections for a source in the front come from the front or the rear? Because both of these directions produce nearly identical interaural fluctuations, it is usually impossible to tell the difference. But if the reflections come both the front and the rear, and they are all decorrelated, the effect is most natural. The listener perceives the source as being at some distance in a natural acoustic space. The sense of depth and space is uniform over a large listening area, and it does not change as the listeners rotate their heads.

In a surround recording when the source is on the right, early reflections should come primarily from the left front and right rear loudspeakers. But the effect is more natural – particularly for a listener facing forward, if the reflections also come from the left rear. Reflections in the 20ms to 50ms range that come from the same direction as the source are either inaudible, or they add coloration to the sound without adding depth.

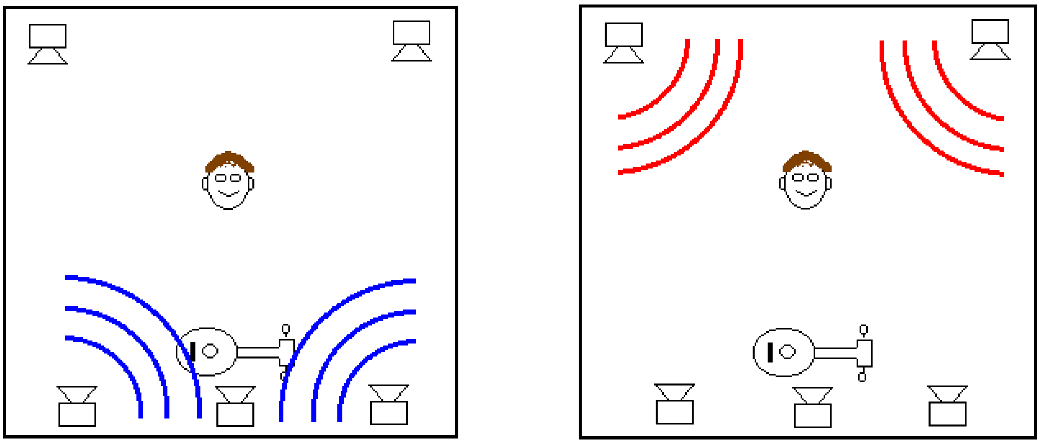

Figure 16: *It is not possible to tell the front/back direction of reflections in the 20ms-50ms time range. Decorrelated reflections from the rear sound identical to decorrelated reflections from the front. Both add an acoustic impression in front of the listener. But ideally reflections should be decorrelated and come from both front and rear at the same time.*

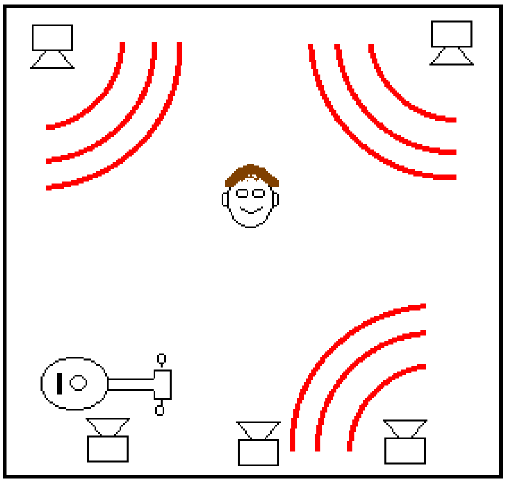

Figure 17: *When the sound source is on the right, the early reflections should come primarily from the right rear and the left front, but reflections coming from the left rear add to the naturalness of the depth impression. These reflections are difficult to generate with microphone technique, but can be generated with a surround reverberation device.*

Why we need more than two speakers to reproduce the hall

For reproducing the sound of a hall, five decorrelated speakers are better than two - and seven are better than five. Professor Boone in Delft reports 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.

Like it or not, people move their head when they listen. Recent experiments with head tracking with earphones (at the IRT and elsewhere) have proved that head motion is an essential part of sound localization and externalization. We have been studying objective methods of measuring the spatial properties of small rooms, and we have found that a single binaural measurement does not produce results that match subjective impressions. It is necessary to make at least two measurements – one with the binaural head facing forward, and one with the head facing to the side.

We judge the spatial properties of a room – and of a recording, by moving our heads. A single pair of loudspeakers playing decorrelated reverberation can reproduce some of the spatial properties of the original hall – but only in the lateral direction, and only for a listener who is facing forward.

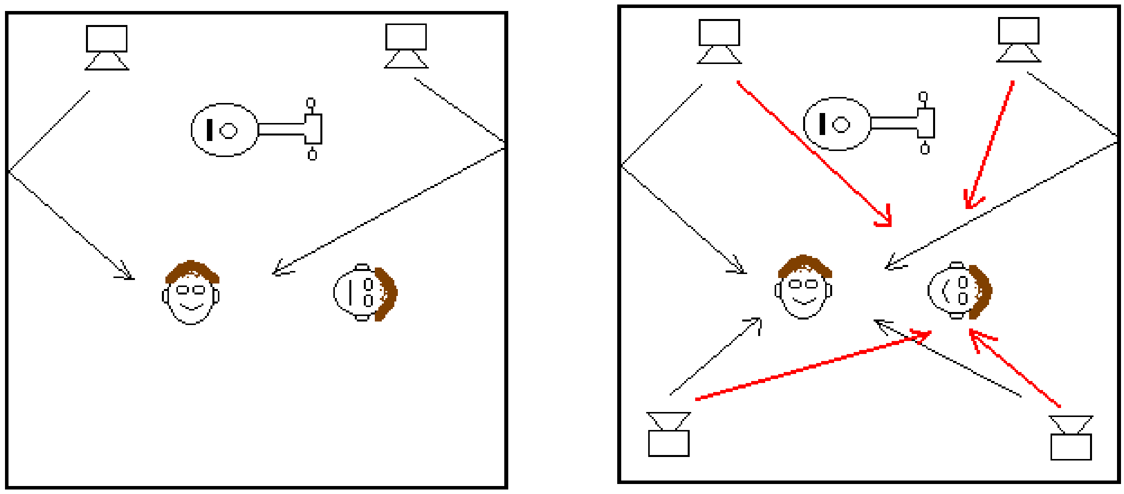

Figure 18: *A single pair of speakers can recreate the spatial fluctuations in the original hall only in the lateral direction. A listener facing the side hears only the size of the playback room. At least four speakers with DECORRELATED reverberation are needed to enlarge the playback room in all directions.*

When we add an additional pair of loudspeakers the spatial fluctuations of the original hall are reproduced both for listeners facing forward and for listeners facing to the side, and the space seems much more natural. But this only happens if the reverberation in all four loudspeakers is different. For example, if the front speakers and the rear speakers are driven in parallel, the room collapses to mono for the listener facing sideways. Likewise, if the rear speakers are driven in parallel, the spatial properties are reduced for a listener facing forward. All speakers must be decorrelated from each other!

Enlarging the size of the listening area

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.

The first thing to do is to add a center speaker. This permits frontal localization over a wider area and eliminates the “listening line”. Adding and additional pair of speakers in the rear allows real images to come from the sides and rear of the listeners, and enlarges the area where the hall reproduction is uniform and natural.

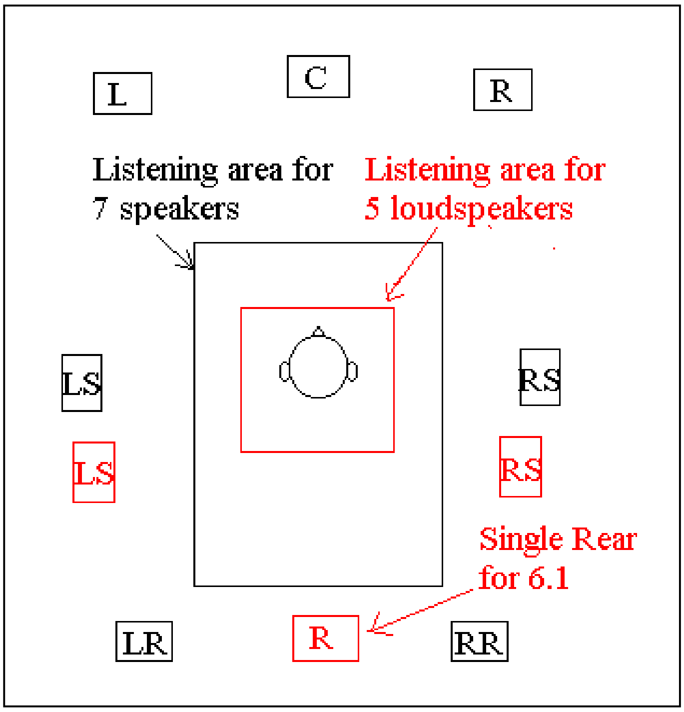

Figure 19: *A seven channel speaker layout – if all speakers are decorrelated for reverberation – increases the effective listening area compared to a five or six channel layout.*

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 the rear speakers are not back far enough to produce a satisfactory rear sound effect. If the side/rear speakers are direct radiators the listening area is not large, because the reverberation field in the rear is not very uniform.

The reason that sound effects further back sound more exciting 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 a recent AES workshop, 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, and it works well.

Mike technique for optimum depth, imaging and hall

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. Even for the front channels only it is difficult to meet these criteria with pressure-gradient microphones located on a single stand, and it is impossible with omnis.

There is a further requirement on the microphone technique used for the two front channels 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 20, 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 stand. This method seems to be reserved for schools and broadcast stations. They will simply have to catch up with the rest of us.

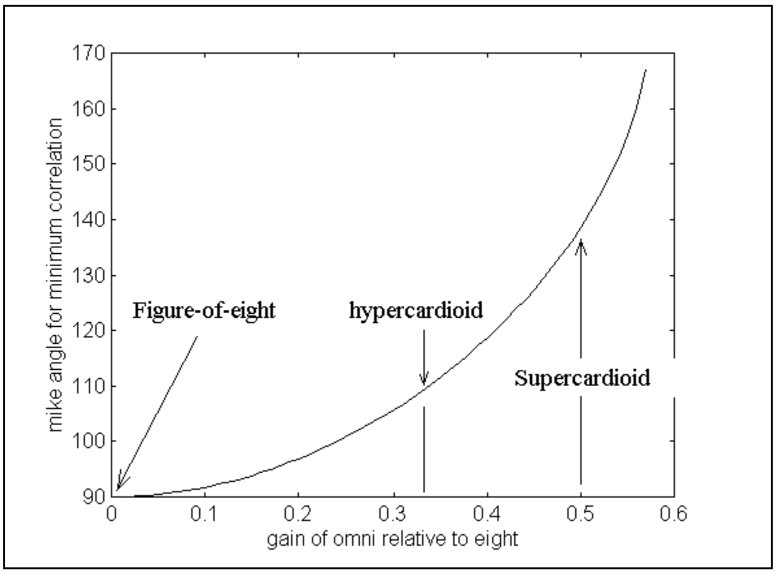

Figure 20: *Optimum mike angle for reverberation decorrelation for various microphone patterns. Note that Hypercardioids are decorrelated with an angle of 110 degrees, and supercardioids at an angle of about 138 degrees. Two cardioid microphones will never be decorrelated even if they point opposite 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. The pattern may look decorrelated on a phase meter, but still be highly correlated at frequencies below 500Hz. The only solution is to use one of the patterns and angles shown in figure 20, or to space the microphones more widely.

Figure 21 shows the correlation between two omnidirectional microphones separated by 25cm in a three dimensional reverberant field, assuming a reverberation radius that is much larger than the microphone spacing. Although the microphones appear to be decorrelated above 1000Hz, there is considerable correlation (or anti-correlation) below 1000Hz. It is decorrelation below 300Hz that gives life and richness to bass tones, and enlarges the apparent size of the playback room. If the reverberation radius is small – as in a reverberation chamber – the correlation will be less. Experiments on the correlation of microphones as a function of distance depend strongly on the reverberation radius of the space.

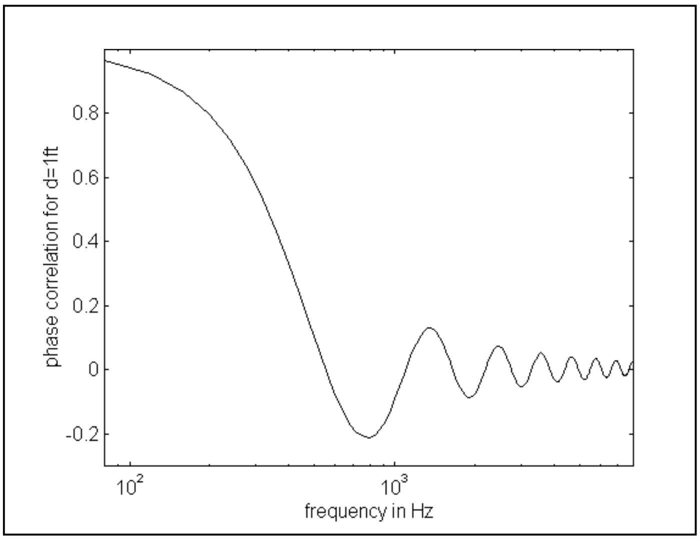

Figure 21: *The amount of correlation between two omnidirectional microphones separated by 25cm in a reverberant field. Calculated in three dimensions. Note the high correlation at 100Hz, and the negative correlation at 800Hz. The separation and the frequencies vary inversely, so a pair separated by 2.5m would have a negative correlation at 100Hz. (But only if the reverb radius were greater than 2.5 meters.)*

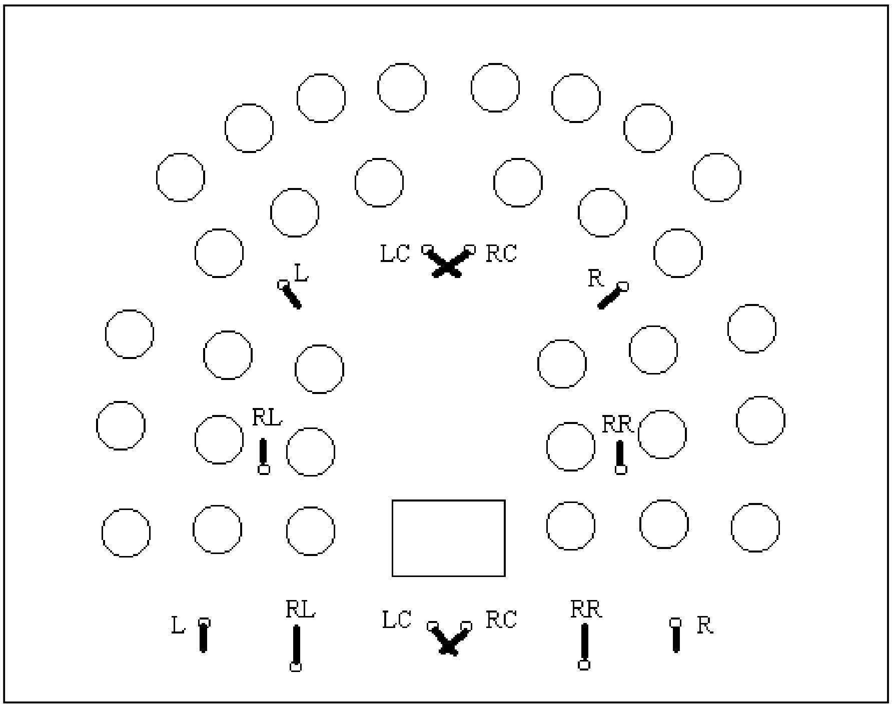

Figure 22: *Example of a simple microphone array for a large orchestra. Two supercardioid pairs are used to pick up a left center and a right center channel. These are panned center-left and center-right in the mix. There are also spaced cardioid microphones for the left and right front channels. The apparent distance and depth is controlled by adding early reflections during the mixing stage. The two backward pointing cardioid 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.

Recording with digital reverberation

In a conventional recording the main microphone is used as the pick-up for the nearest instruments. Accent microphones are added to pick up the direct sound from more distant instruments. The leakage of these instruments into the main microphone acts as an early reflection, and gives these instruments a sense of depth and perspective. The instruments close to the main microphone lack these added reflections – so they sound closer. However the direct sound from the instruments in the back row of the orchestra is too weak to be picked up by the main microphone pair. So when we add accents, the depth perception is reversed. The instruments that are furthest away sound closer than instruments at middle distances.

There is another perspective problem with conventional microphone technique. Instruments to the far left or far right of our sound field will also sound too close. The reason is that the main microphone array is too far away from these instruments – and if the main microphone array has decent left/right separation, the apparent reflection from the direct sound falls in the same loudspeaker that reproduces the direct sound from the accent. REMEMBER – an early reflection that comes from the same direction as the direct sound is INAUDIBLE or undesirable. There is no effect on the apparent distance of an instrument from such a reflection.

Figure 22 uses a pair of supercardioid microphones behind the conductor. However, this pair is not used as a "main microphone", but as a center channel pick-up. There is no main microphone in figure 22. As a consequence, nearly all the instruments will sound too close. The solution is simple: use electronics. We can generate the needed distance and depth perception by adding early reflections electronically. Since much of the direct sound comes from the front, it is usually not desirable to add early reflections to the center speaker.

A surround reverberator that has five or more inputs can be configured to supply the early reflections in just the loudspeakers that need them, regardless of the direction of the incoming sound. A reverberator with two inputs and four outputs can also be used, as long as the early reflections on each output are decorrelated from each other. This configuration is particularly easy to use with most mixing desks. A two channel reverberation send can be derived from the various microphones in the mix. By controlling the level of each instrument in the reverberation send, the apparent depth of each instrument can be separately controlled.

The result is usually better than a more conventional microphone technique. Instruments (or a chorus) in the rear of the orchestra sound behind the other instruments, as they should. Instruments on the far right or the far left no longer sound too close to the loudspeakers, but move back into proper perspective.

Hall reverberation

In figure 22 the hall is picked up as leakage in the front microphones. The hall in the rear channels is directly miked by the two cardioid microphones pointed to the rear. The lack of reverberation in the front channels is deliberate. In my experience it is almost always better to record the front channels with less reverberation than you will eventually need. The amount of hall sound in a mix is critical, and it is usually better to add a little later.

Besides, there are usually problems with the reverberation in the hall. Sometimes there is too little bass, or the hall is noisy. Small halls have too much energy in the 50ms to 150ms time range. Almost always the mix is wrong. The loudest instruments in the orchestra will be the loudest in the hall, and they will sound too reverberant in the recording. It is better to use an electronic reverberation device with at least four decorrelated outputs to add reverberation to both the front and the rear channels. We can use the decay profile suggested by Perceptual Modeling.

As in the case of the early reflections, ideally the hall should be equally loud in the front and the rear. This will produce the most uniform reverberation field in the playback room, and will make the largest possible listening area. The spectrum of the added hall reverberation is also very important. Often a shelving filter at about 300Hz is useful. One needs about 4dB more reverberant level below 300Hz than above 300Hz. This effect is very difficult to control using microphone technique, although some engineers have reported success by equalizing the surround channels only. It is better to equalize the reverberant component of all channels the same way.

In sum: we are using electronic reverberation to achieve two separate acoustic perceptions. We add early reflections to create the desired depth, and we add late reverberation to bring out the hall. Ideally the reverberation device should have a separate control for each function, and a control for the spectrum of the late reverberation.

Conclusions

1. It is possible to make two channel recordings of large sources, such as a symphony orchestra, that sound excellent over a wide listening area. To do this we use microphone techniques with high left/right amplitude separation and high decorrelation of the reverberation.
2. Using the same techniques, but adding a center loudspeaker, can create a large listening area for sources with a strong center. However a sound source in the center must be stronger in the center loudspeaker than it is in the phantom image.
3. Adding two additional speakers behind the listener improves the realism of the reverberation, and enlarges further the apparent size of the listening room. This only works if the reverberation in the front is decorrelated from the reverberation in the rear.
4. Our studies of the perception of reflections show that early reflections that do not come from the same direction of the sound source can create a perception of distance and depth. If these reflections fall between 20ms and 50ms, they do not affect the intelligibility or clarity of the direct sound. In conventional recording techniques these reflections are supplied by leakage of sound into other microphones and into the main microphone array. This process is poorly controlled and often unsatisfactory.
5. Adding early reflections electronically during the mix allows the apparent depth of a sound source to be controlled accurately, and without added muddiness. We call this technique "Perceptual Modeling". These early reflections should come from all directions except the direction of the direct sound, and should be different in each speaker.

6. Hall reverberation should be added equally to all the loudspeakers except the center speaker. The reverberation must be decorrelated in each loudspeaker, which means that a microphone technique using a specific combination of patterns and angles should be used, and/or that the microphones for the hall should be separated by at least the critical distance.
7. A reverberation device used for a surround recording should be able to separately control the level of the early and late reverberation, and to carefully minimize the energy in the 50ms to 150ms time range. All outputs should be fully decorrelated at all times, and ideally early reflections should not be generated in the same direction as the incoming sound.

References

1. Griesinger, D. “Speaker Placement, Externalization, and Envelopment in Home Listening Rooms" Presented at the 105th Conveintion of the Audio Engineering Society, San Francisico, 1998 Preprint # 4860
2. Griesinger, D. "General Overview of Spatial Impression, Envelopment, Localization, and Externalization" Proceedings of the 15th International Conference of the AES on small room acoustics, pp 136-149, Denmark, Oct 31-Nov.2, 1998
3. Griesinger, D. "Multichannel Sound Systems and Their Interaction with the Room" Proceedings of the 15th International Conference of the AES on small room acoustics, pp 159-173, Denmark, Oct 31-Nov.2, 1998

These references are available on the author’s web site at www.world.std.com/~griesngr

Appendix

The derivation of the optimum angle for minimum correlation between two closely spaced pressure gradient microphones in a three-dimensional reverberant field.

The derivation relies on a number of facts:

1. I assume that the reverberant sound field contains three independent spatial axis, and the reverberation along each axis is completely independent and uncorrelated. Note that this assumption assumes a three dimensional space - if we recorded with coincident microphones on a flat surface (the two dimensional case) the result will be different. In fact, for the two dimensional case only two figure-of-eight microphones at 90 degrees will be decorrelated.

2. I assume that any of the reverberation axis can be rotated - and the decorrelation will be preserved. In other words, if we have a figure-of-eight output along the x axis, Ox, and a figure of eight output along the y axis Oy, where Ox and Oy are decorrelated, then if

$$Oy' = Ox * \sin(a) + Oy * \cos(a)$$
$$Ox' = Ox * \cos(a) - Oy * \sin(a)$$

Then Oy' and Ox' are also completely decorrelated for all values of a. Note that this implies the simple case (where a = 45 degrees) that the sum and difference of two uncorrelated signals is also uncorrelated. These equations correspond to a simple rotation by the angle a of the x and y axis of a coincident figure of eight pair at 90 degrees. This should have no effect on reverberant decorrelation in a diffuse soundfield.

3. Now let's model the microphone pair by an omni and a figure of eight combined. We will call the output of the figure of eight V and the output of the omni as M. The microphone has a pattern determined by the amount of the omni component, so the mike output O is

O = V + m*M

Where m varies from zero (a figure of eight) to one (a cardioid). Values of m greater than one produce sub cardioid microphones, and the result is always correlated.

4. We will also assume symmetry about the y axis. Thus we consider a single microphone in a pair, with an angle from the y axis of a. If a=0 the microphone points forward along the y axis. If a=90 the microphone points along the x axis. We assume the other microphone in the pair points with an angle of -a.

5. Now we want to write an expression for the amount of independent decorrelated energy in the microphone. To find achieve zero correlation, we will compare the voltage difference between the two microphones to the voltage sum of the two microphones. When the total energy in the sum signal is equal to the total energy in the difference signal, then there will be no correlation. In this example, the difference signal is entirely composed of the velocity pickup along the x direction. The sum signal consists of the sum of the squares of the pressure signal along the x direction, the pressure and velocity along the y direction, and the pressure in the z (or vertical) direction.

6. Now we need to discuss the phase of the pressure wave compared to the phase of the running wave. We will assume here that we have running waves, and the pressure is in phase with the velocity. (Assuming we have standing waves gives the same result.) In other words, if we have a sound traveling along the positive x axis, the pressure and the velocity in the x direction will be in phase. (If we calculate the total sound power in a diffuse field using this assumption we get the classically known values of P = 1 for an omni, P = 1/sqrt(3) for a cardioid, and P = 0.5 for a hypercardioid with m=0.33.)

We will do the calculation for running waves. To do this we will take the sound power in the x axis (lateral axis) and divide it into two parts: the part that is moving in the positive direction along the x axis and the part that is moving in the negative direction. We will call the positive part Vlp and the negative Vlm. The corresponding pressure components are Plp and Plm. We will assume the left-moving sound is uncorrelated with the right-moving sound.

We will add the positive moving pressure component to the positive moving velocity, and subtract the negative pressure component from the negative velocity.

The output of the left microphone is given by:
OL = Vlp*(sin(a)+m*Plp -Vlm*(sin(a)+m*Plm

OR = -Vlp*sin(a)+m*Plp + Vlm*sin(a)+m*Plm

The sum signal, OL + OR

OL+OR = 2*(m*Plp+m*Plm)

OL-OR = 2*(Vlp*sin(a) – Vlm*sin(a))

Remember the Vlp and Vlm are independent random variables – so when we sum them we must take the sum of the squares. Remember also that Vlp and Vlm share half the power in the x direction. So if the magnitude of the total power in the x direction is VL, then the magnitude of Vlp = magnitude of Vlm, and each is VL/(sqrt(2)). We assume the same for Plm and Plp. Thus,

OL+OR = 2*sqrt((m*Plp)^2+(m*Plm)^2) = 2*m*PL
OL-OR = 2*sin(a)*VL

To find the total output of the sum signal we must add in the pick up from the other two axis, the front/back and the up/down. The up/down signal is only picked up through the pressure, so we simply add in the pressure PU. The front/back signal is picked up both with pressure and velocity. Once again we divide the front back signal into a positive going wave and a negative going wave. However this time the signs of the signals work out differently, and

Front/back(OL+OR) = 2*sqrt((m*PF)^2+(VF*cos(a))^2)

OL+OR (total) = sqrt((2*m*PL)^2 + (2*sqrt((m*PF)^2+(VF*cos(a))^2))^2 + (2*m*PU)^2)

If we now assume that PL = PF = PU = VL = VF = VU = 1 (this is simply saying the reverberation is uniform in all directions) we can set the equations for the OL-OR equal to the equations for OL+OR, and find the values for m as a function of the angle a.

Sin(a) = sqrt((m^2 + m^2 + cos(a)^2 + m^2) = sqrt(3*m^2 + cos(a)^2)

Or

3*m^2 +cos(a)^2 – sin(a)^2 = 0

3*m^2 +2*cos(a)^2 –1 = 0

The equation is not solvable for values of m larger than sqrt(1/3), or 0.57. This means that a microphone must be nearly a super cardioid (m=0.5) to give a decorrelated reverberation, even with a mike angle of 180 degrees (a = 90).

This equation has the following values, and is plotted in figure 20.

For	a =	m =	
	45	0	figure of eight
	54	0.33	hypercardioid

69	0.50	supercardioid
90	0.57	the maximum omni component.
no angle	0.1	cardioid

We see that for an opening angle of 110 degrees (a=55) (typical of x-y coincident technique) we need a hypercardioid microphone to get zero correlation in the reverberation.

Notice also however, that for two cardioid microphones pointing oppositely along the same axis the correlation is not zero. Thus the latest proposal by Dr. Theile for a main microphone, which uses two capsules separated by a meter, should use supercardioid capsules pointing forward (a = 70 degrees) if it is to have minimum reverberation correlation at low frequencies. If hypercardioid capsules were used the microphones could point forward with a= 54 degrees.