

International Tonmeister Symposium, Oct. 31, 2005  
Schloss Hohenkammer

## **Improving 5.1 and Stereophonic Mastering/Monitoring by Using Ambiophonic Techniques**

By Ralph Glasgal  
Ambiophonic Institute  
4 Piermont Road  
Rockleigh, New Jersey, 07647, USA  
[www.ambiophonics.org](http://www.ambiophonics.org)  
[glasgal@ambiophonics.org](mailto:glasgal@ambiophonics.org)

**Abstract:** It is desirable that clients judging a recording at a session or a mastering engineer evaluating mic balances, panning algorithms, center channel level, virtual sound positioning, or ambience levels have a control room monitoring system that is uncompromised by the inherent defects of the stereo triangle or the 5.1 speaker array. Keeping the ITDs, ILDs, and pinna cues, captured by the microphones, intact when a recording artist auditions the raw session or later during mastering, increases the odds of early artist approval and provides a more consistent approach to evaluating any subjective postprocessing. It is also suggested that any rear ambience channels sound more musical if convolved using the latest libraries of 3D hall/theater impulse responses than attempting to record them live. These convolved surrounds should be compared with the rear mic signals if such have been obtained during an acoustic recording session in a concert hall, opera house, or church.

### **1. Stereophonic versus Binaural Monitoring**

All human sound localization, with the eyes closed, is based on the clues provided by interaural time differences between the ear canals, interaural level differences between the ear canals and the one and two eared pinna functions. A single pinna can act as a direction finder for sounds with energy above 800 Hz or so. This is why an individual with hearing in only one ear can function almost normally. There are also dual pinna-direction-finding functions that allow localization to within half a degree, even when there is no ITD or ILD, if complex higher frequencies or transients are present. The ITD and the ILD function really well only for signals with energy below 1000 Hz. Thus where complex sound fields such as music are involved, localization is degraded if any of these parameters are missing or distorted by the recording or the reproduction method. Ideally all the three localization cues, ILD, ITD and Pinna, should be present and all be in agreement to provide physiological verisimilitude and thus a less strained monitoring experience.

Unlike everyday binaural hearing, the ability to detect the sonic illusion of phantom images between the speakers of the stereo triangle or the two frontal triangles of 5.1 differs greatly from individual to individual. Head size, pinna shapes, and other genetic aspects of an individual's hearing mechanism vary to the same extent that individuals differ in their ability to see optical illusions. Thus expecting musicians or clients to hear an adjustment a record producer makes in the same way the producer heard it is often unrealistic. But if the track being monitored is converted to a binaural-like or everyday hearing format that does not rely on stereophonic sonic illusion imaging, then all monitoring parties will likely hear the same thing and will be better able to agree on what needs to be modified. Later, such modifications will be more likely to be appropriate for a larger number of later home buyers even if they listen via a stereo triangle or 5.1 arrangement that is nothing like the monitoring system.

Unfortunately, neither the 60 degree stereo triangle nor the two 30 degree side by side triangles of 5.1 are capable of preserving all the localization cues that have been captured by the recording microphone. That is, most stereo or surround microphone arrays almost always gather more ILD and ITD than is ever heard in the monitoring room. Thus when adjustments are made in channel balance, spot mic balances, panning controls, equalization, etc. or even when a take is played back for a client, decisions are not made with all the mic captured cues being present and audible. Thus unwise adjustments may be made to compensate for monitoring anomalies that are unique to the control room system or to the ears of the monitoring engineer or his client. This is true both for recordings made with microphones or electronic music made with virtual sound software. In the following discussion we will consider a stereophonic system, but the same reasoning applies to the LCR part of the 5.1 methodology.

## **2. Stereophonic Monitoring Pitfalls**

We consider now several combinations of common microphone arrangements comparing what is captured and then what is generated during monitoring. In figure 1 a pair of slightly more than head spaced omnis records an ITD of approximately 900 microseconds for an instrument way off to the side. However, when played back over speakers spaced +/- 30 degrees the ITD sensed is reduced to 220 microseconds and thus due to the precedence effect, the cello moves from 75 degrees to 30 degrees. This may superimpose the cello over the woodwinds and your conductor will not like it. Additionally there are two audible early reflections added in reproduction that are not part of the recording. Omnis are used here for clarity, but subsequent figures show no customary mic arrangement is immune to such anomalies.

In figure 2, the cello is at 25 degrees and its recorded ITD of 200 microseconds is preserved in monitoring. The recorded ILD is 0 dB but the stereo triangle generates an ILD of about 6dB which has not been recorded, at least for the higher notes of the cello where the head shadow is significant, and similarly for violins and violas in these mid-side positions. There is also a strong early reflection created at the far ear that is delayed by over 200 microseconds and so is not well merged with the direct sound. Such a reflection is probably too frontal to enhance envelopment but may cause image widening.

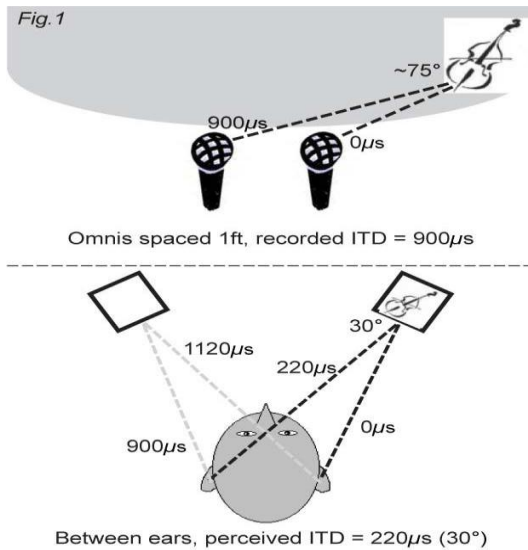


Figure1. Stereo or 5.1 Crosstalk Distorts Large Recorded Interaural Time Differences (ITD) When Monitoring.

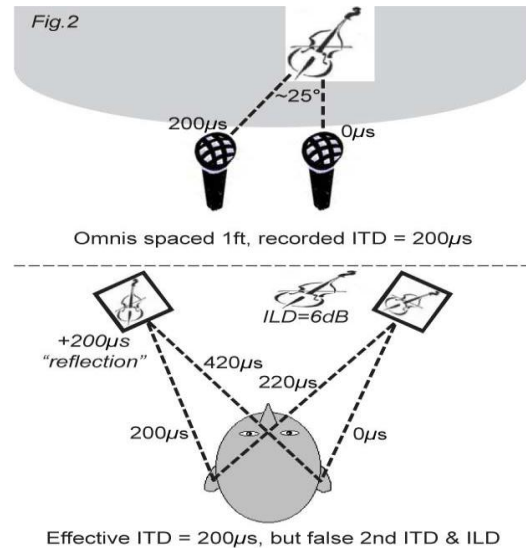


Figure2. Crosstalk Introduces a False Early Reflection and a Spurious ILD of 6dB.

In figure 3 coincident cardioids or Blumlein mics are used to record an oboe at the far edge of the stage. In this case the level difference recorded is possibly 10 dB. There is, of course, no recorded time difference. However, when one listens to the oboe, flute, piccolo or trumpet in 800Hz range, via the usual stereo monitoring system this large recorded ILD is reduced to 2 dB and a spurious ITD of 220 microseconds appears. Thus the instrument is heard at 30 degrees rather than 75 degrees and many instruments may appear to be lumped together.

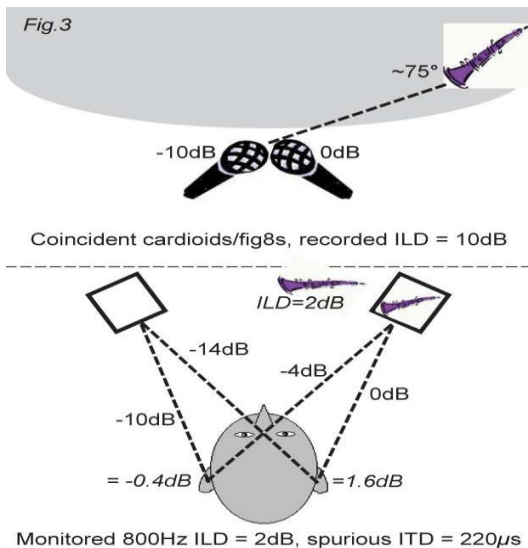


Figure3. Stereo or 5.1 Crosstalk Distorts Mid-frequency Recorded Interaural Level Differences (ILD) when Monitoring

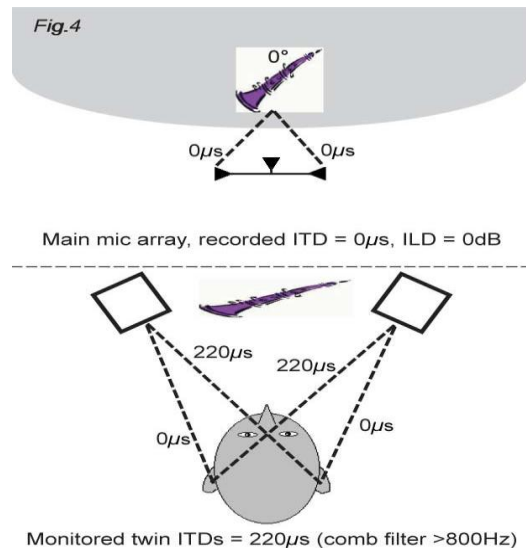


Figure4. For Central Sources at Mid Frequencies, Monitoring in Stereo Creates Two Spurious ITDs that Cause Comb.

In figure 4 the main mic records no level or time differences for a wideband central instrument. But upon reproduction at the console, there are two ITDs or two early reflections depending on how you view them. But more damaging is the combfiltering or the peaks and dips that occur if you move your head side to side. While not usually audible as changes in pitch or overtones, this combing causes level changes that generate ILDs at some frequencies but not others so that an instrument can appear to be off center for some notes. This combing of central sources also mimics pinna direction finding patterns further confusing localization. This combing characteristic is probably the primary cause of listeners being able to detect something is canned rather than live even when only a single instrument or voice is recorded outdoors. The rule is that a small single sound source such as a voice or harmonica is best reproduced via a single speaker. This is one reason why a mono center speaker for movie dialog is better than stereo.

In figure 5 we assume that a velocity pair recording a piccolo at the far side of the stage only outputs an audible signal on one channel. This could produce a normally large ILD upon reproduction. However, the pinna and the head shadow engendered ILD and ITD localize this monophonic signal to the loudspeaker as in everyday azimuth perception and the stage is again limited to the angle between the speakers which may unconsciously disturb the client conductor.

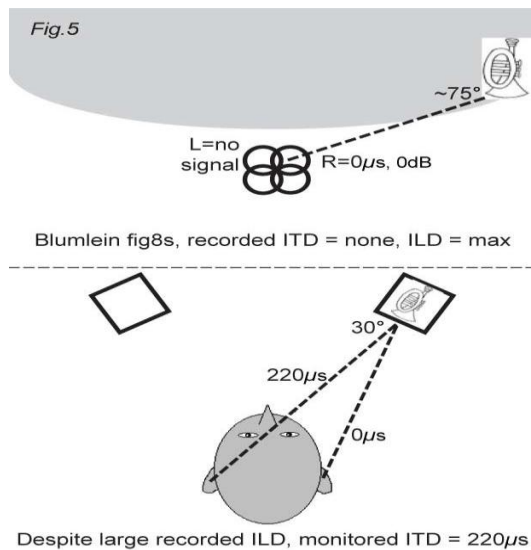


Figure5. Stereo Speaker Triangle Limits Stage Width Perception at Higher Frequencies when Monitoring.

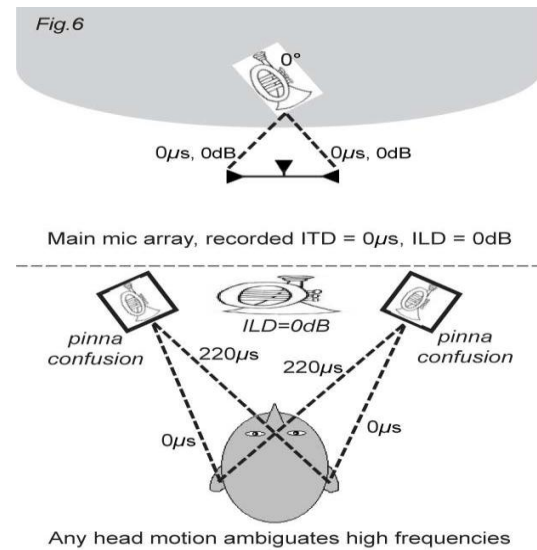


Figure6. High Frequency Central Sources Are Difficult to Localize When Monitoring in Stereo or 5.1

In figure 6, a central high frequency source is recorded and naturally has equal left and right recorded signals. Upon monitoring with speakers at 30 degrees, the pinna direction finders sense the higher overtones off to the side but the ILD is zero so the brain localizes the sound to the center, but this mechanism, like that for optical illusions, does not satisfy completely. Small head motions can also inspire doubts as to the high fidelity of the system.

It is clear that different types of recording microphones react differently with various loudspeakers that differ in crossover networks, number of drivers, time alignment, and directionality in largely unpredictable, undetected or unanticipated ways. So, in general, for a wide range of microphone arrangements, instruments, and stage locations, monitoring in stereo will inevitably introduce faults or prejudices which may lead to editing decisions which are of doubtful validity and which other listeners with quite different speakers and ears may later find objectionable.

### Establishing a Crosstalk Cancelled Monitoring Station

To avoid such pitfalls we suggest a monitoring facility that uses a binaural technology. That is one that allows the ITDs, ITDs and pinna directions to be heard as in Figure 7.

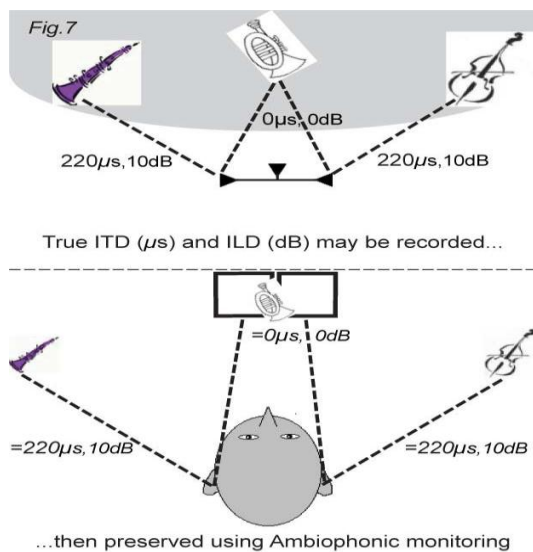


Figure7. Ambiphonic Monitoring Preserves the True ITD and ILD Captured During a Recording Session.

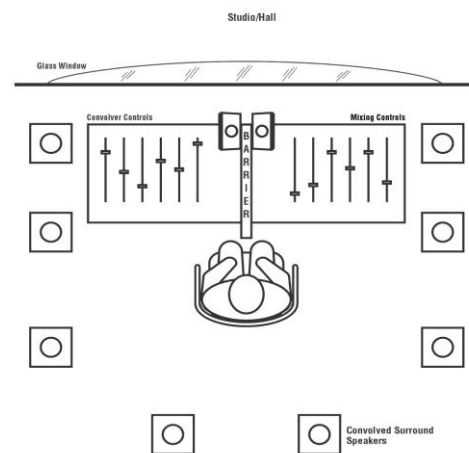


Figure8. Early Monitoring Station Using Barrier.

In figures 8 and 9a you can see an early version of such a monitoring station. Putting a simple physical barrier between two speakers directly in front of the monitoring position eliminates the crosstalk and most of the pinna confusion particularly in the central 60 degree stage area. The speakers should be head spaced on each side of the panel. The center channel in the 5.1 case is fed equally to both of these speakers. Today one uses crosstalk canceling software, which is readily available, to do the same thing without the physical barrier. Figure 9b. This method of 5.1 monitoring makes it much easier to see what happens when the center speaker is engaged. There is also no chance of a delay error between the side and the center speakers to cause errors in judgment. You can hear easily if the center channel information is compressing the width of the stage or if there are phasing effects. You can also switch to 60 degree stereo speakers plus center for a quick comparison at any point in the process.





Figure9a. Inexpensive Monitoring Station Preserves ITD and ILD



Figure 9b Software Based Monitoring Station Preserves ITD and ILD

Our experience is that musicians, listening in a binaural environment, can more easily appreciate what has been captured and are less likely to request changes, especially those that are irrational, as is quite possible, when monitoring just in stereo or 5.1.

If it sounds fine monitored this binaural way and is subsequently released without too much processing will it sound better on all those subpar stereo systems out there? I believe so but this is a subjective opinion not susceptible to proof. But common sense indicates that the great variety of systems out there will insure that the percentage of good reviews will remain about the same whether the mix is psychoacoustically pure or psychoacoustically eccentric. However, if the mix keeps the cues relatively intact, it is then possible, in years to come, for a home listener to recover this data and hear the stage with all the depth and width that the microphones did capture.

Robin Miller of Filmmaker Studios has devised a coder, Figure 10 that can convert a purist four channel recording into a 5.1 equivalent. Then at a future time a decoder can be used to fully recover the original unprocessed 360 degree surround recording. Thus one could use an advanced recording surround technology such as Ambiophonics, monitor it with full binaural realism, please the client, but still release the performance in 5.1. We believe that the 5.1 recordings made this way are superior to the recordings made the conventional way using the typical methods reviewed in the first part of the next section.

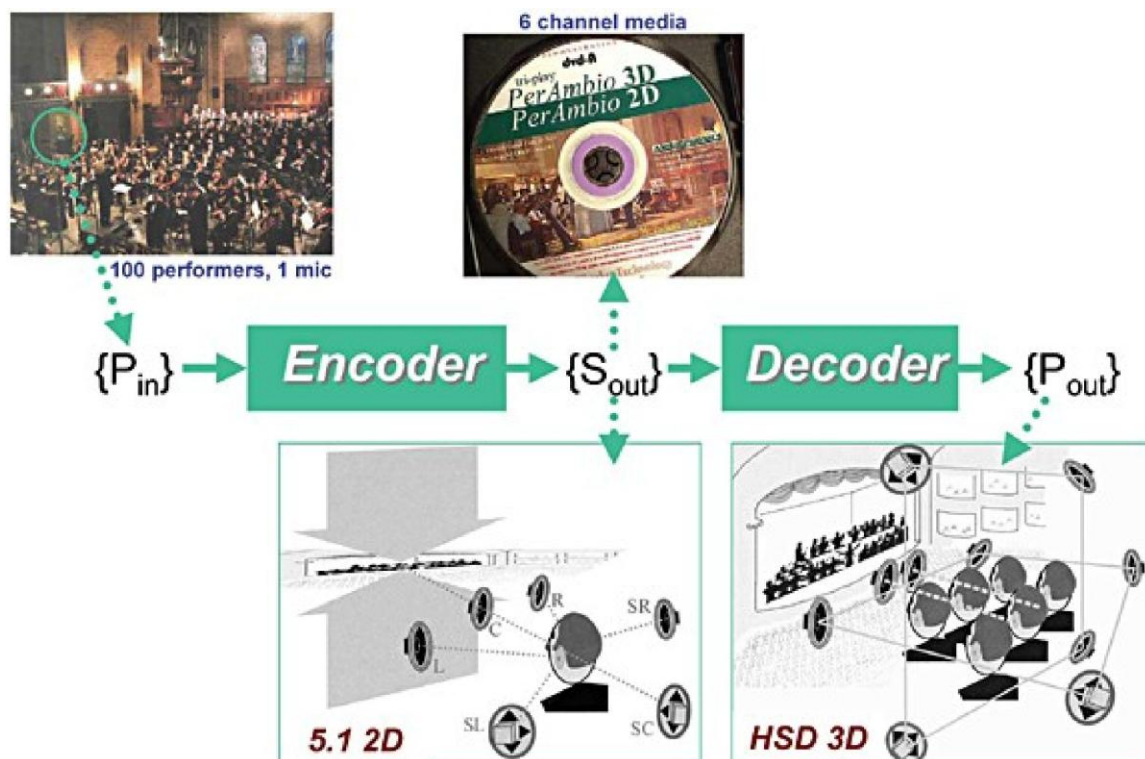


Figure10. Four Purist Microphone Channels Convert to 5.1 and Back Again

## Monitoring Speakers for Special Studios

Figure 11 shows a pair of Soundlab Electrostatic panels capable of rock concert SPL levels, working as a software crosstalk cancelled pair or Ambiodipole. Such electrostatic panels are extremely accurate transducers. Being full range, (except for low bass) they do not have crossovers, thus preserving ITDs and ILDs and making crosstalk canceling, and thus monitoring, more effective.



Figure11. Soundlab Ambipole Full Range Electrostatic Panels Operate at Very High SPLs, Have No Crossovers, Preserve ITDs and ILDs, and Don't Confuse the Pinna.

Figure 12 shows that it is possible to do accurate monitoring with very small speakers. In this case, the full-range Bose AM-5 is very directional and like the electrostatics has no crossover in the ITD, ILD region. By using an extra speaker for each additional listener, you can have more than one monitoring station in the same room.





Figure12. Inexpensive Small Speakers Act as Point Sources and Function Well as 2 Ambidipoles Side by Side for 2 Viewers

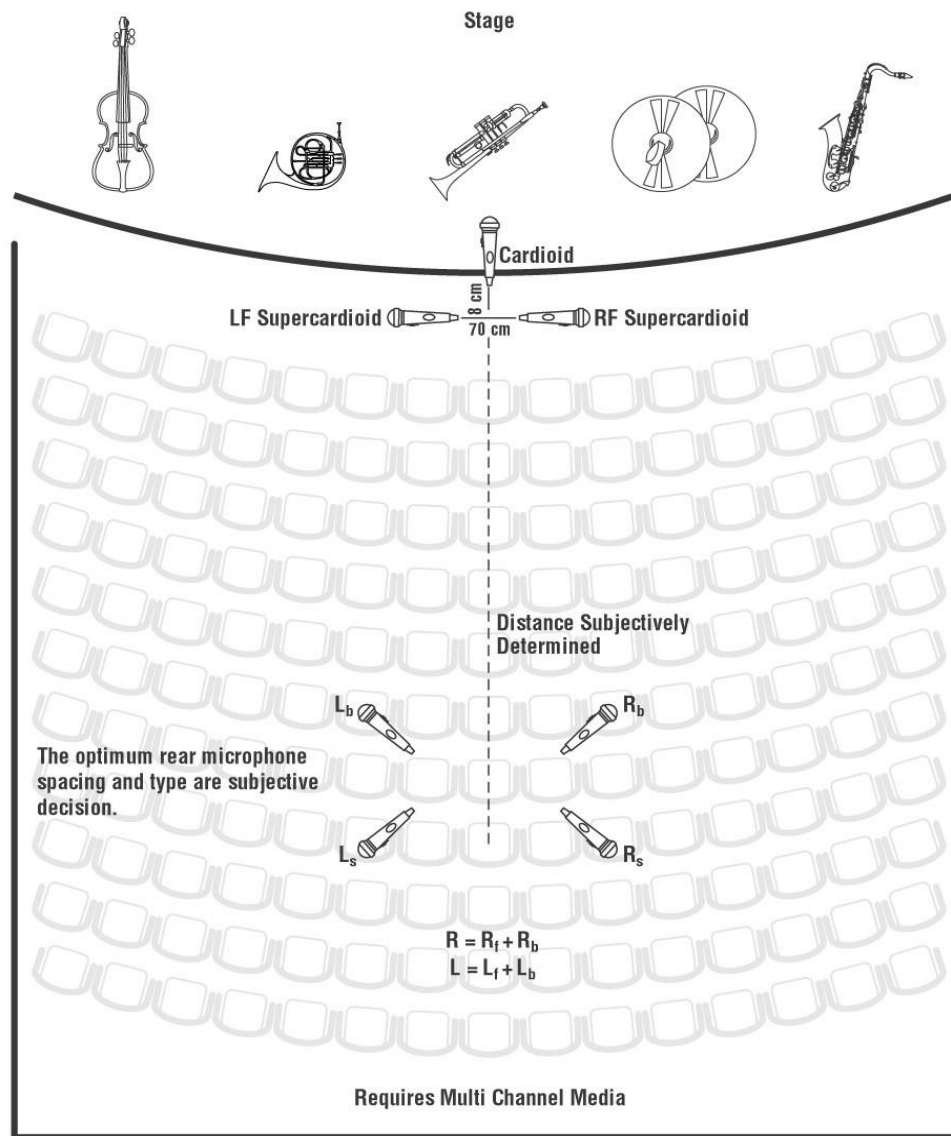
Figure 13 shows the Soundlab Majestic. The Majestic is an Electrostatic panel that can operate at 115 dB SPL and do it down to 20 Hz. It is meant for use in large studios where the utmost in fidelity is needed. Since, like the Ambipoles it has but one sound producing membrane, it is completely time coherent. Like the Surrstat (below) and the Ambipoles, the curvature limits room reflections that originate from the rear of the speaker.



Figure13. The Soundlab Majestic is the Ultimate Monitoring Loudspeaker with an SPL Capability of 115 dB.

### **Recording and Monitoring the Surround Channels**

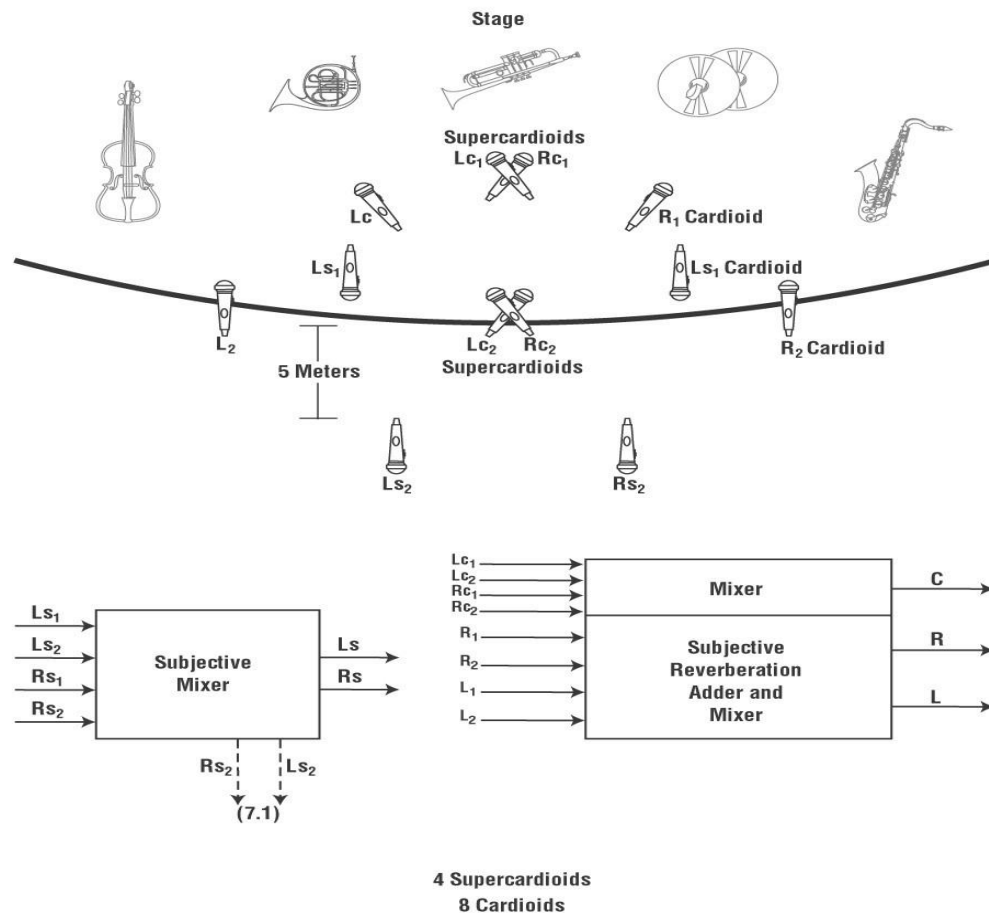
The fact that there are almost as many methods of recording surround sound as there are recording engineers is indicative of the fact that no method is psychoacoustically valid. Figures 14 and 15 show two methodologies for recording live music in a hall by Theile and Griesinger respectively. In practice, such methods are constantly being adapted but mic layouts like these illustrate the problems being encountered.



**Optimized Cardioid Triangle (OCT)**

Figure14. The OCT Microphone System Requires Subjective Decisions Dependent Upon Accurate Monitoring.

In the OCT drawing you can see that the location of the hall ambience microphones is arbitrary. Even the spacing of the hall mics and their directionality is not defined and left to the whim of the recording engineer. The Griesinger arrangement is similarly subjective and in practice almost impossible to implement. A key feature is the need for three mixers to be adjusted by ear.



**Dr. David Griesinger's 5.1/7.1 Recording Methodology**

Figure15. The Griesinger System Depends on Subjective Adjustment of Mixers

The basic problem, which neither these nor any other 5.1 recording array can solve, is that good sounding or realistic hall ambience cannot be properly recorded during a live performance or during an acoustic recording session. Compounding the problem is that the imperfectly gleaned ambience from the session cannot be mixed down to two media channels and then fed to two rear speakers with any expectation that such a mix will produce anything like a true hall experience.

A much better way to record ambience in the absence of rear direct sound is not to record it at all. Signals for any number of rear surround speakers are best derived from a library of hall impulse responses or from a venue impulse response obtained before or after the session. If you don't have to worry about capturing signals to mix for the rear channels, the main microphones can be simpler and placed more advantageously. Modern impulse response gathering tools and the processors to use them have already reached a level of fidelity that exceeds that of any live performance microphone methodology so far proposed.

The impulse response of the hall desired is then processed with the main mic signals in a mathematical operation called convolution to produce as many surround channels as you wish. A major advantage of using 3D impulse responses is that one can also easily convolve surround signals for elevated speakers in the monitoring room to further the sense of realism that musicians appreciate. Impulse responses and the software to use them are now readily available from Waves Audio and others.

Monitoring the surround channels derived from a convolver and adjusting the convolver to complement the front channels is a lot easier than working with microphone signals, that have mixed ceiling, side, rear and frontal reflections all together and are almost always contaminated with some slightly delayed direct stage sound. Figure 15 shows a live recording session with a main microphone construction that can be placed without regard to collecting sound for the surrounds. This microphone, called an Ambiphone, does also have two omni mics behind the panel and so can be used to pick up rear hemisphere direct sound such as applause or be used in movie making.





Figure16. The Ambiophone, Above and Behind the Conductor During Live Recording of Beethoven's Ninth, is Beyond the Critical Radius Without Ill Effect.

Even using a convolver with an appropriate impulse response cannot make the two speakers of 5.1 capable of delivering anything approaching a live in-hall music experience. But at least the surround ambience can be truer and uncontaminated by direct sound or by rear-hall-mic-captured ambience conflicting with the frontal ambience, unavoidably recorded by the frontal mics. If you convolve to say eight surround speakers, spread about the monitoring room, including overhead, you can have musicians listen to your tracks and their performances in much greater acoustical comfort. Someday with blue laser media you could even deliver such convolved ambience channels to the public with ease.

Figure 17 shows an Electrostatic Panel designed by Soundlab that mimics a concert hall wall when energized by convolved ambience. Several such **Surrstat** panels in a monitoring studio can provide a convincing "you are there" soundfield.



Figure17. The Surrstat Electrostatic Panel from Soundlab Allows Surround Speakers to Behave More Like Concert Hall Walls.

Figure 18 shows how a psychoacoustically advantaged monitoring/mastering studio could be setup. It allows for binaural monitoring and convenient comparisons of that with a stereo or surround downmix. Figure 19 shows the details of a coder to convert an Ambiophonic 3D recording to a 5.1 compatible mix and a decoder to recover the original Ambiophonic recording when desired.

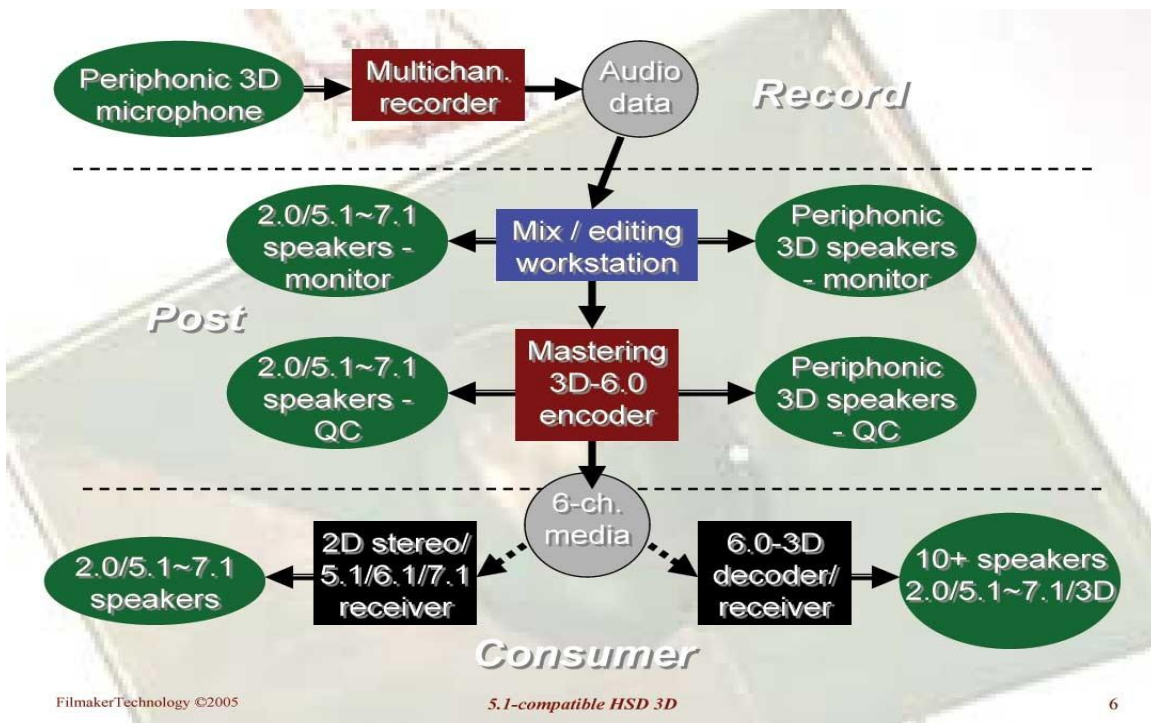


Figure18. Monitoring/Mastering System Maintains Correct ILD and ITD

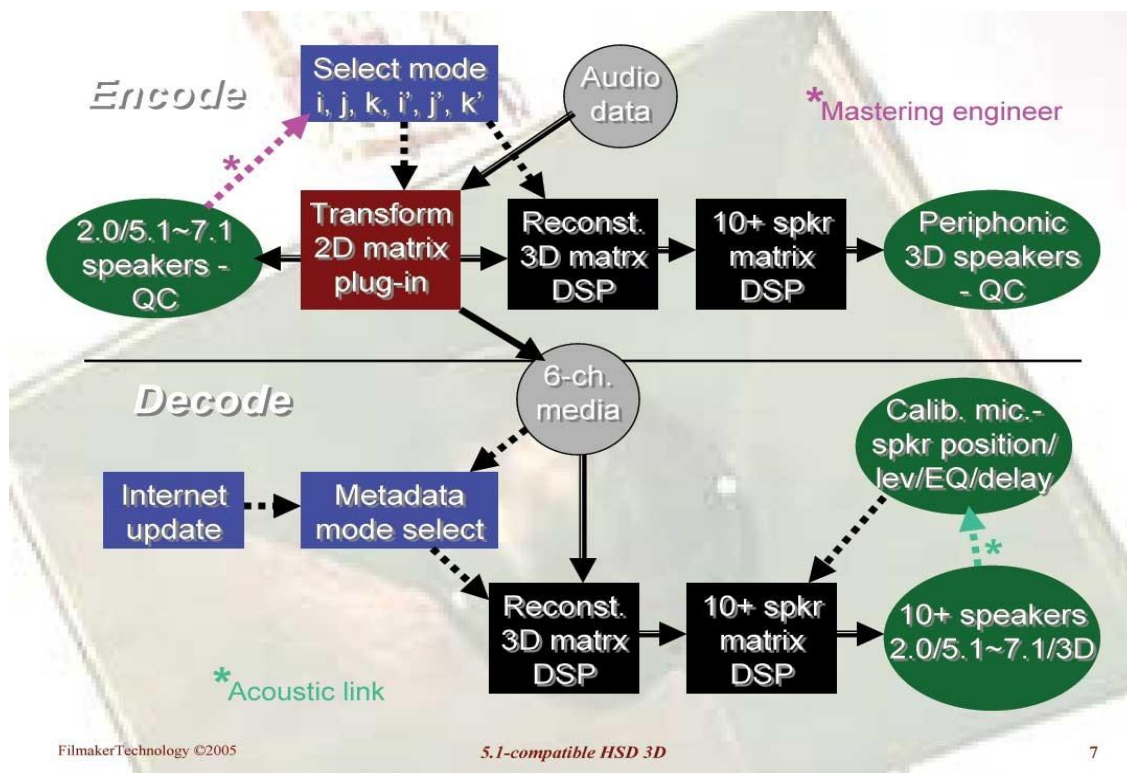


Figure19. Encoder-Decoder Processes Conversion from 3D to 5.1