RECORDING IMMERSIVE 5.1/6.1/7.1 SURROUND SOUND, COMPATIBLE STEREO, AND FUTURE 3D (WITH HEIGHT)

ROBERT E. (ROBIN) MILLER III[1]

[1] *Filmaker Technology, Bethlehem, PA, USA ©2006*
www.filmaker.com

Much 5.1/6.1 content, like stereo, is sourced by panning close microphone signals. The presentation pretends the "sound sources are here," where spatiality is that of the listening room and thus non-varying. However, realistic, immersive, compelling reproduction of music, movie atmospheres, and gaming effects requires preserving spatiality and directionality-dependent tone color of both direct sounds and acoustic reflections. The extent to which convolution with room impulse responses can contribute not just diffuse ambience but "realistic" tone color is considered. New recording techniques are explored, applicable compatibly to 5.1/6.1 where reproduction is a 2D horizontal circle of speakers, to stereo (including personal devices using earphones e.g. iPod), and to future 3D, where the listener is again at the center of the sphere of natural hearing.

INTRODUCTION

Recording stereo and 5.1 surround has evolved into an overly complex process, often involving dozens of microphones distributed throughout the ensemble, costly post-production, artifacted and uncompelling results, and underperformance in the marketplace. Yet a live human listener in a good seat enjoys ensemble performances with only his/her own two ears. How and why has this happened? And are there better ways?

CHALLENGES RECORDING STEREO & 5.1

In the beginning of recording when only one main microphone (later a pair for stereo) was used, recordists assumed that audio recording would emulate human hearing – using a main microphone in the best seat in the house, say in the 5th row, as a surrogate for the live listener's ears. But non-ideal microphones and the absence of compensating visual cues gave rise to perceptual shortcomings when listening to these recordings. Diametric opposites such as presence v. tone color, individual instruments v. blend, and ensemble v. hall in each case seemed irreconcilable with a single microphone in one chosen position, but a main microphone could not be in many places at once.

Distributing "spot" microphones added complexity and created more problems. Fixes had diminishing returns: EQ to fix comb filtering, separating musicians with headphones, isolation booths, or even separate sessions as though musicians' sense of ensemble was secondary. These practices had detrimental effects on the sound and on the performance as well.

Even as musicians felt that, with these complexities of recording their performances, the tail was wagging the dog, recordists experienced the dog chasing its tail. Furthermore, producers, then audiences acquired a taste for the bright, close-up sound even if bigger than life. Typically in popular music, the situation has escalated to overly processed, overly loud, less than natural results that one would think would be ever stranger to listeners as delivered fidelity approaches perfection.

Consider the dichotomies that recordists including the author for many years faced in this vicious circle:

1. Correct balance between direct sources (less dull) and acoustic ambience (less muddy) cannot be achieved without moving the main microphone closer (typically above the conductor);

2. Now that the main microphone is too near some sources but relatively further from others (obeying the inverse square law), correct balance among sources cannot be achieved without rebalancing using spot microphones;

3. Spot microphones, lacking separation, create artifacts such as comb filtering, image smearing in time, and destruction of binaural properties for personal (earphone) listening;

4. Restoring acoustic ambience (including low frequency modal enhancements) requires separate "room" microphones or artificial reverberation;

5. Spot and room microphones add complexity and cost both in recording and post-production.

When producing content in 5.1 surround, this situation seems to increase with the square of the number of channels. Still other aesthetic dichotomies arise during mixing: whether to pan spot-mic'd

instruments only to one of the five speakers, to create a phantom image between any two, or instead to envelop the listener with the spatiality of the recording environment (the concert hall, the movie or gaming scene) typically reproducing a front stage and ambience in front and back, perhaps favoring a multi-channel main microphone array rather than individual spot mics.

If content isn't compelling, it won't sell. And the decrease in sales year over year reported by recorded music labels suggests that over-processed, over-loud, same-sounding recordings are becoming passé. The author believes the solution is a change in recording methods, recognizing that:

- Recording microphones and reproduction technology for home, car, or portable listening have evolved to where compensation for limited fidelity no longer exists;
- Any listening-environment-dependent processing is better left to the listening realm, rather than embedding processed sound and its artifacts in the recording realm, as though one size fits all;
- Beyond today's delivery in stereo and mono, 5.1 surround can deliver natural envelopment and balance between direct sounds and ambience;
- Tomorrow, full-sphere 3D (with height) promises the ultimate in life-like reproduction.

Just as the market for stereo matured from an initial attraction in the early 1950s of a ping-pong novelty, 5.1 is maturing from mere cinematic fly-bys of action films. In increasing numbers of home theaters, life-like music and realistic games are possible that far exceed stereo. Yet stereo will predominate in the near future. And the ultimate in reality, full-sphere 3D, is demonstrably compelling for the future. Recording techniques for tomorrow that exhibit the scalability and compatibility to address all these needs will be explored in this paper.

HIGH SONIC DEFINITION – HSD

High Sonic Definition or HSD, is intended to be to sound what High Definition TV (HDTV) is to picture – its complement for movies or gaming. For content without picture, HSD conveys the experience of closing ones eyes at a real concert – the sound is believably live even without seeing it being made. HSD is compatible with 2-speaker stereo, personal (earphone) stereo, and 5.1/6.1/7.1 surround (ITU-R775). Furthermore, HSD in full-sphere 3D will soon be practical on modest replay equipment. Ongoing HSD developments, papers, and solutions can be found at www.filmaker.com.

HSD recording requires a special microphone and recording technique – see Fig.1. The signals captured can produce content for 3D (with height) reproduction, conventional 2D surround (speakers in the horizontal plane), stereo (including binaural and personal play devices), or mono (as of this writing still the majority format of broadcast radio and television). Furthermore, HSD encoding in a single hierarchical format on a standard "shiny" disc produces a recording that can be decoded by the user heirarchically as 3D, 2D, or stereo.

Fig.1 - HSD microphone showing front of baffled ellipsoid and discrete soundfield array mounted above (total 8 microphones).

Fig.2 - Greenwich Village Orchestra recorded in concert with HSD microphone positioned slightly beyond the critical radius.

To illustrate, a musical presentation can be captured using the HSD microphone and multi-track recorder. The HSD microphone is designed to approach perfect omni-directionality – a surrogate for human hearing – so that use of spot microphones may be reduced or eliminated. The HSD microphone is typically placed at or beyond the critical radius – further from the sound source than conventional main-microphones, that are often positioned within the critical radius to avoid muddy sound, but therefore also unduly favoring closer sources due to the inverse-square law for direct sounds. Positioned at or beyond the critical radius as in Fig.2, reverberant sound is equal to or greater than the direct

sound, giving an impression similar to live hearing in an ideal seat without the need for rebalancing using spot microphones. Spot mics may of course be used, but fewer of their signals will be necessary in mixing, so costs can be less both in setup and post-production.

Monitoring in all formats that can be derived from the HSD microphone requires a control room with all speaker configurations and switching between them, as in Fig.8. With the flexibility in post-production of the HSD-3D signals and with recordist experience and the confidence it brings, the author has found that many typical venues can be captured simply by verifying that the stereo down-mix is good for this position of the microphone, as stereo (or perhaps mono) will show the most compromise by a "wrong" mic position. (This "minimalist" approach does not imply compromise in quality; it is not unlike capturing a movie on film by an experienced cinematographer. Though he might not be able to "monitor" the results until the dailies screening next day, the latitude of the negative will have captured all that will be needed for color timing the print, transferring to video, or authoring the DVD.) 2D surround and 3D using the HSD microphone are actually increasingly less subject to degradation due to mic positioning, similar to there being more than one seat in the house that is good for live hearing, which, by more distant positioning, the HSD mic emulates more than conventional micing.

Once captured, the signals can be combined for a specific format, or processed using the HSD encoder for compatible reproduction in any format, from full sphere 3D to 2D surround to stereo or mono. With HSD, unlike some main-microphone approaches, the same recording can be released on a single disc and reproduced on headphones usind iPod and similar players, stereo speakers, 5.1/6.1/7.1, or 10-speaker 3D with height. Of course, the mixing room must indeed have all these speaker layouts and the means for switching among them to assure quality for each.

For purposes of illustrating the process, the following sections describe typical procedures in capturing audio in for concert music or sound effects for gaming or film ambience (video nat-sound) using the HSD microphone. HSD technology is intentionally hierarchical to permit capture that provides a broad range of compatible signals encompassing current standards for 5.1 and stereo plus an emergent technology for a possible 3D future. Three cases are described for content creation in 1) stereo, 2) 5.1/6.1/7.1 surround (2D), and 3) and full-sphere 3D (with height).

HSD MICROPHONE APPLICATIONS

The HSD microphone was designed for use by recording engineers to be a universal basis for multiple purposes: stereo; 5.1/6.1/7.1 (2D) surround; and 3D with height – and all distributable in compatible form from the same distribution disc. The following sections explore each of these broad applications and introduce general techniques as practiced by the author. After many years practicing conventional techniques, the HSD method requires "thinking outside the box" and a bit of re-learning, which the author has found is fully justified and greatly rewarded by the uncompromised results. In sum, recording in HSD produces stereo and surround of high quality, as well as the basic signals for full sphere 3D (with height). In hundreds of demonstrations recording engineers, musicians, and lay subjects have encouraged HSD's development with their favorable responses and comments [1].

1.1 Stereo (including iPod)

Stereo in this context denotes 2-channel (2.0) reproduction using speakers, portable devices using earphones, broadcast TV & radio, and computer multi-media (including gaming). While the techniques vary somewhat for each of these purposes, it is possible to capture signals at one time for use in all these. Conventional capture may involve multiple microphones and recording channels to be mixed (panned) to a 2-channel stereo result. Alternatively, a "main microphone" array (two or more microphones) may be mounted on a single stand or suspended, optionally augmented with "support" or "spot" mics.

Either as an approach unto itself or as the basis for surround, explored below, we begin the description of the HSD microphone with the "direct to stereo" case. An approach to stereo will be chosen by the recording engineer for simplicity of use and natural-sounding, unartifacted results. Toward this end, the basic 4-element HSD microphone offers these advantages:

- Uniform omni-directional response with frequency (a "perfect" omni);
- Nominally flat magnitude response 5~30kHz;
- Discriminates a 120º maximum front stage (e.g. sources) and 240º back stage (ambience);
- Control of balance front-to-back either during recording or in post production;
- Captures Interaural Level and Time Differences (ILD, ITD) for reproduction using either loudspeakers or earphones (including virtual headphones using crosstalk cancellation);
- Basis for surround sound, both the present 2D (ITU-R775) and future 3D (with height).

In essence, the basic HSD microphone is four microphone elements mounted at human ear positions

on an ellipsoidal head-shape (pinnaless) with baffles separating front and rear element pairs on each side, as in Fig.1. A third baffle directs non-early reflected sound from above to the back channels. Patent pending and in its fourth generation, it is cast in plastic, 24in wide, weighing 10lb. The design for a 5th version addresses issues of aesthetics and sight lines in the presence of an audience or cameras.

In its primary purpose of capturing sources and the acoustical signature of their environs, the HSD microphone is positioned at or slightly beyond the critical radius of the room. Because it acts as a "perfect omni" and thus as a surrogate for a human listening live, it can be thought of as positioned in the best seat in the house, even if elevated. This contrasts with conventional practice using main microphones based on two or three omni-directional elements (sphere microphone [2], Decca Tree), or on Blumlein figure-8 mics crossed 90º, that, if ideal, would sum to an omni equivalent. However for stereo, practical microphones must be positioned well within the critical radius, where the balance has not developed to that heard in the hall, nor intended by the performers. By the inverse square law, this positioning favors instruments close to the mic and therefore distorts the intended balance with more distant instruments. Correct direct/ambient balance and tone color including room contribution is achieved at a greater overall distance at or beyond the critical radius, where the perspective delivered by conventional omni-equivalent main microphones is too distant.

More distant positioning is the purpose of main array configurations that use directional elements (e.g. M-S cardioid+bi-directional, X-Y coincident cardioids, ORTF cardioids, Williams cardioid arrays), but these rely on critical 1st-order cosine directional elements exhibiting polar response with frequency that even in the best implementations fall short of ideal, typically deficient above 4kHz and below 200Hz. Addressing decreasing HF output suggests smaller diameter diaphragms, which exhibit lower signal-to-noise performance. Falling LF output requires bass compensation, resulting in still greater LF noise.

In contrast, the HSD microphone, by assigning only a portion of the horizontal pickup circle to each of its four elements, combines mainly that narrow portion (~90º) of each element's polar pattern where its response is nearly ideal. The HSD microphone uses 6mm diaphragms, but gains 6dB SNR using boundary effects of its baffles. A pinnaless head-shaped ellipsoid, it preserves ILD and ITD that corresponds to accurate localization during stereo replay both over speakers or earphones. And using separate front and back pairs, the recordist has control of the balance, live or in post, between direct (source) and indirect (ambient) sounds, permitting positioning at or beyond the critical radius where the musical impression and tone color are correct.

Whether capturing a large ensemble or a single player, it is the acoustics, acting as an extension of each source instrument, that contribute early reflections and late reverberation arriving from all directions. Preserving the direction and timing relationships at each surrogate ear-pair of the HSD microphone is critical to life-like reproduction, as will be discussed below regarding tone color as an important purpose for surround sound – fully realizable especially in 3D (full sphere, with height). For stereo release, the four channels captured by the basic HSD microphone can be mixed pair-wise (front/back), in effect to "zoom" the listener's perspective from the 2nd row to the 15th. Mixing live or in post-production to vary from very present to very ambient, this relationship can be adjusted by song or movement, or continuously depending upon the music.

Positioning the microphone is critical as always to avoid disadvantageous room modes and reflections. However, the realm of positioning is now significantly larger, expanding even beyond the room radius, where balance is significantly better among player-to-player and player(s)-to-room. This balance is then adjustable, either live or in post-production, by varying the relative contribution of front and back element pairs. For highly mobile recording of ensemble music or gathering sound effects, only the basic HSD main microphone and a portable recorder are required. Monitoring during field recording may be as simple as using earphones and switching between front and back to assure basic quality of each pickup pair. Then the front is mixed with the back at varying levels to assure good stereo will result.

Where spot microphones are warranted such as for broadcast of a live event, the HSD microphone is used as a main mic to establish a reference for the venue and for ambience, with voice-over and support mics added in the mix. For ensemble music, the recording engineer might position and record spot mics for confidence, even if it is found in post-production that they are not needed. Whether mixed live or in post-production, spot microphones should be delayed with respect to the main microphone to minimize time distortion artifacts [3].

The HSD microphone is calibrated, using a band-limited pink noise source or simple clicker, so that each pair and the pairs together will track when using preamplifiers that have their gains ganged. Gains must track with an accuracy of a fraction of a dB. The back level pair may be set nominally with respect to the front pair simply by observing low frequency source material.

During mixing, whether live or in post-production, the back stage is typically mixed –2dB to –8dB with respect to the front pair, depending on the "zoom" effect required for the music or sound effect. As will be seen below, when surround channels are available to the

listener's ears, the back levels are mixed 2 to 3dB higher – approaching life-like balance because, owing to pinna effects, back sounds coming from their normal direction can be discriminated naturally by the listener. Because of the HSD microphone's "perfect omni" equivalent characteristic, little or no equalization is typically needed, consistent with the desirable trend in high quality audio away from overly processing sound.

1.1.1 Monitoring for 2-speaker stereo

Ruinous artifacts that exist in 2-speaker stereo are not well known today although they were to Blumlein in the 1930s. They are due to crosstalk from each widely spaced speaker to both ears, as in Fig.3. The solo human voice or instrument arrives at two different times at each ear – in effect a spurious ITD – causing comb filtering that distorts tone color note by note and that changes as the listener's head moves. In addition, a phantom in the center but actually coming from speakers toward the sides causes pinna confusion that blurs localization and widens the phantom center image.

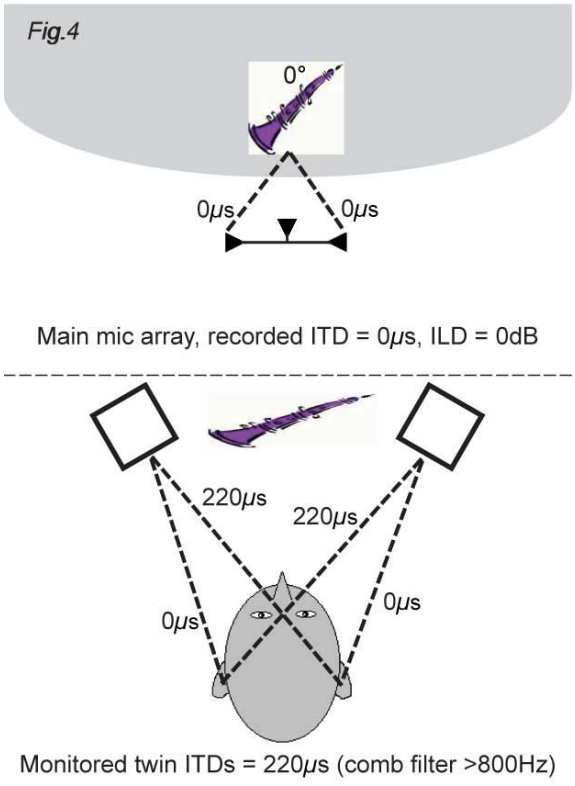

Fig.3 - Reproduction using speakers spaced 60º causes a spurious ITD of approx. 220us which results in comb filtering, distorted tone color, and blurring of important central images.

The solution, applicable both to consumer replay and to monitoring by recording, mixing, and mastering engineers is to employ closely-spaced speakers and crosstalk cancellation to create virtual headphones, as in Fig.4. Whether termed Transaural or Ambiophonics, the approach is preferable for either 2-speaker stereo or personal (earsphone) stereo for its more accurate and predictable revelation of correct mic levels, panning, and integration with ambience [4].

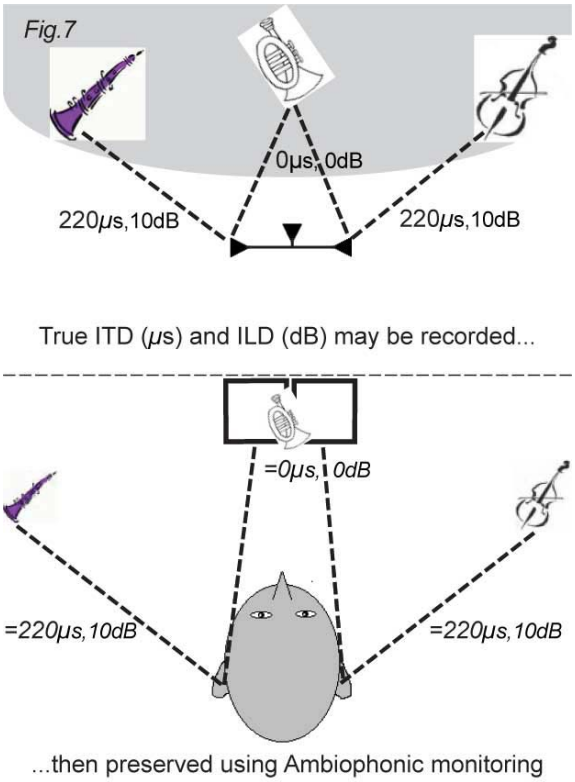

Fig.4 - Reproducing stereo using closely-spaced speakers and cross-talk cancellation solves tone color distortion and blurring, and produces a 120º wide stage with no "hole in the middle."

1.2 5.1/6.1/7.1 surround (ITU-R775)

Surround sound for movies, music, and gaming per ITU-R775 must be recognized as horizontal (2D) reproduction, where the listener in the ideal position is at the center of a circle of speakers. Justification for listening at the center of the sphere of live hearing is considered below re implementing 3D reproduction (with height).

2D ITU-R775 systems are termed "5.1" and include 5.1, 6.1 or 7.1 formats. Surround distribution uses five or six main (full-range) channels plus an optional LFE

low frequency enhancement/effects channel, so-called "0.1" because it is limited to very low frequencies (VLF typically <100Hz) that, according to the standard, are reproduced optionally without detracting from the content if missing. This LFE channel has been found to be used sparingly in movies for its intended purpose – namely for the impact of explosions and the like – where an additional 10dB of headroom is desirable. In the author's survey of DVD movie content, these effects were impulsive (not tonic), and were typically in the range of 70~120Hz [5,6,7]. This short term nature at moderately low frequencies is intended to produce a startling effect without damage to equipment or ears.

LFE is one of two signals reproduced by so-called subwoofer(s) in home theaters and in multi-media and Home-Theater-in-a-Box (HTiB) systems that use satellite-size speakers that reproduce only >100Hz. LFE is derived in post-production, however the recording engineer should capture VLF signals in main channels where appropriate, e.g. the *1812 Overture*.

1.2.1 VLF <100Hz and Bass Management

The second and more important signal reproduced by subwoofers is to redirect VLF (<100Hz) content from main channels during replay by the process of Bass Management. For home entertainment systems, it is both economical and advantageous regarding control of listening room mode response to redirect VLF to a subwoofer, along with any LFE signal [7 refs].

The author and others have argued that although a monaural LFE enhancement channel is appropriate for impulsive (non-tonic) sound effects enhancement, highest quality reproduction strongly argues in favor of binaural bass management to redirect tonic VLF >45Hz from main channels by hemisphere (left/right), termed Binaural Bass Management [5,6,7], as in Fig.5.

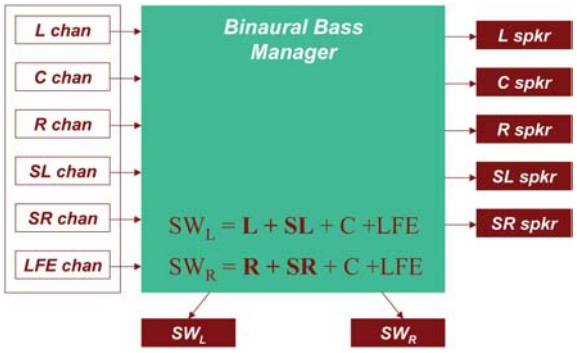

Fig.5 Block diagram of a binaural bass manager illustrates redirecting VLF from main channels to a subwoofer in the same hemisphere. Daisy-chaining four subwoofers offers additional possibilities for "positional-equalizing" listening room modes.

For binaurally redirected VLF, the recording engineer should capture in left/right microphone pairs binaural VLF sounds 45~100Hz, that are relatively uncorrelated by hemisphere [7]. Near-spaced omni microphone techniques such as HSD are able to capture natural-sounding and highly desirable lateral LF and VLF reflections in acoustic spaces; widely spaced omnis can be used to supplement the effect.

1.2.2 Native capture for 5.1 Surround Sound

Although more listener-minutes of content are consumed in stereo, there is increasing interest among providers of music and gaming content in addition to film and HDTV in recording natively in surround. The word "native" implies original capture using multi-channel surround microphone arrays, highly different from remixing multi-track originals by merely panning mono signals intended only for stereo and containing no spatial information correlated with another signal.

The author and others believe that the slow acceptance by the public for surround content without picture is that "re-panned mono" surround, lacking inter-channel spatial information, is not compelling. Soon after any initial "wow" factor that prospective consumers of surround experience, similar to the first time one heard ping pong demonstrations in stereo 50 years ago, the gimmickry of five mono speakers loses appeal. The most obvious reasons are: Sources coming from directions that seem "unnatural" to the listener are distracting; voices reproduced by speakers of inferior quality such as those typically in back are unsatisfying; and two totally correlated sounds emanating from any two speakers due to panning a mono signal can create comb-filtering that distorts source tone color and smears images. Ironically, it is this self-inflicted "phasey" result that has dissuaded many uninformed mix engineers from using 5.1 surround sound.

Several approaches have been devised for native surround recording, including many the author has tried and found quite good: those suggested by Williams and Theile (OCT) among them. Those approaches that are successful capture and reproduce two-dimensional spatiality that begins to approach live hearing, where a natural balance of correlated and uncorrelated signals is conveyed to the listeners' ears by multi-microphone and multi-speaker arrays (more than 2). In acoustic performance spaces, this includes not only direct sounds of sources, but also reflections, both early discrete ones that confirm localization and spatiality, and late diffuse reverberation that conveys the size and character of the space. (As explored later, the integration of arrivals, processed uniquely according to the HRTF of each listener, creates perception of all-important tone color.)

For given listening acoustics, the recording engineer controls whether the listener perceives either that the "musicians are here" (intimate, but all recordings so made sound invaryingly like the listener's room, which becomes boring) or that the "listener is there" (you get to travel!). This difference can be as simple as whether the listening spatiality dominates or is dominated by the recording spatiality – in essence depending upon how controlled is the listening spatiality and how completely the recording spatiality was captured.

1.2.3 HSD microphone for 5.1 Surround

The HSD microphone is the basis for 2D 5.1/6.1/7.1 (ITU-R775) surround whereby its four microphone elements provide appropriate signals for the corner speakers. Although this is not meant to imply that the Center channel/speaker is not desirable, it should be acknowledged that many practice the option of ignoring the Center path (typical for surround ambience and music). As originally intended when 5.1 was developed for cinema dialog, the Center channel can be used to anchor important central (solo) voices. This is unlike Williams or OCT methods, where stereo phantom imaging is developed between any two microphones/speakers using combinations of ILD and ITD due to the spaced directional microphones used.

As described above and in detail in a prior paper [8], the HSD microphone array uses omni-directional elements that act in sum within the horizontal plane as a perfect omni-directional microphone, emulating live human hearing. As used in HSD Surround, the Center channel is decorrelated with respect to L and R and may be created either from support (spot) microphones OR a fifth element added to the array that is directional and aimed 0º. This is approach is advisable especially for reproduction in home theaters where, with respect to left and right speakers, the center speaker typically is different in size, timbre, and plane (above or below the screen). (As will seen below, this center channel can also be derived from the HSD-3D array.)

If a natural impression is to be preserved and spot microphone(s) must be mixed into center, left, or right channels, the author advises using the RRB method of Theile that adds delay to compensate for the difference in distance from the supported source to the support mic and to the main mic [3]. For example, a spot mic 1m from a solo voice is delayed 10ms with respect to signals from the main HSD array 4m from the voice. More support mics, each given appropriate delay, may be mixed to any front channel if its source is sufficiently isolated from other spot mics so as not to create comb filtering and distortion of tone color. Sources off the median plane may be panned after appropriate delay to appropriate corner channels as long as, by precise monitoring, the result preserves the localization established by the main array. Finally, ambience and reverberation may be supplemented using distant ambience microphones (usually two or four) as described by Theile and Hamasaki, necessitating delaying all other microphones in relation to them.

1.2.4 Tone color and surround sound

When approaching the task of capturing surround sound, it is essential to recognize that the role of reflections is as important as the role of direct sound. By definition, beyond the critical radius, where the HSD microphone may be positioned, more energy is indirect than direct. Certain microphones for surround may appear to be aimed at nothing, but important reflected energy is coming from every direction around a sphere of sound. And not just as a pressure scalar, but as directional vectors of pressure gradient (sometimes called velocity) captured by a microphone pair. Native capture of 2D 5.1/6.1/7.1 (ITU-R775) surround will preserve this circle of envelopment, mapping the horizontal components of arrivals in 3-space in the horizontal plane, thereby more or less closely approximating 2D directionality in 5.1. (As will be seen below, totally preserving the 3D sphere for the listener promises localization and tone color that is even more fully immersive and lifelike – because of our pinnae.)

Just as direct sounds, reflections arriving at the ears have directional ILD and ITD (horizontal components), interpreted in the brain as localization (horizontal azimuth). Furthermore, the convolutions of the outer ear (pinna) alter the magnitude response by differently comb-filtering each arrival direction in the 3D sphere (azimuth & elevation). Integrating these pinna-colored arrivals over the room constant (approx. RT/4), the brain perceives a quick fluctuation in magnitude response that defines the "tone color" (timbre) that the performer has also heard and applied as real time feedback to playing his/her instrument and its acoustic extension, the room. The performing, the coloring of arrivals in time and 3D space, and the integrating in the listener's brain determine tone color. Tone color may be the most prized if misunderstood of psychoacoustic qualities – the essence of artistic expression, compelling communication, and perception of what is real. While localization of sources and spatiality of ambience are important, if tone color is not preserved, we know we are listening to just a recording, not live hearing.

RECORDING 3D (WITH HEIGHT)

As stated, for real-sounding surround, it is essential to preserve both the localization of direct sound and of

early reflections to preserve life-like tone color. It bears repeating: Tone color may be the most prized if misunderstood of all psychoacoustic qualities – and the essence of artistic expression, compelling communication, and perception of what is real. While localization of sources and spatiality of ambience are important, if tone color is not preserved, we know we are listening to just a recording, not live hearing.

Our ears are located in the horizontal plane, so it is correct that surround cues in this plane are more accurately localized (reduced within the cones of confusion at each side and front/back confusion for sounds on the median plane). Forward-facing, the pinnae (outer ears) favor frontal localization, so human hearing is most acute in what might be envisioned as a horizontal ellipse in front of the listener, not unlike and mostly coinciding with his/her field of vision. Horizontally within this zone, the combination of ILD, ITD, and pinna coloration result in the ability of humans to localize the provenance of sounds, direct or reflected, accurately to approx. 1º. Arrivals from above and below, while not as precise by an order of magnitude (because only pinna coloration is involved), still contribute significantly to tone color and therefore to life-like impression. Unlike vision, sounds are perceived as to tone color and localizable from every direction in 3-space. Hence, 3D reproduction, where the listener is once again at the center of the sphere of hearing, not just the circle of 2D surround, is a required step toward the illusion that sound reproduction is real.

3D reproduction has been the holy grail of audio since the recognition of height information contributed by the pinnae. In the 1960s, Gerzon developed Ambisonics, a mathematically elegant method of capturing and reproducing a life-like "soundfield" and the four-element microphone so named. The "B-format" microphone has evolved several generations and have been realized by the author and others using discrete elements. But apart from mainly academic interest, 1st order Ambisonics was a non-starter in the marketplace, due both to its complexity and the fact that it is not sufficiently accurate to reproduce a front stage, where human discrimination is most acute. High Order Ambisonics (HOA) promises accuracy when the 2nd and higher order microphones required are available.

Farina and the author have described an approach to 3D that remedies the deficiency in the front stage of Ambisonic using a hybrid with Ambiophonics (note the possible confusion in terms) championed by Glasgal and others that produces a front stage that is naturally accurate and 120º wide – double that of conventional stereo (60º). As described above for stereo monitoring, Ambiophonics uses crosstalk cancellation and closely positions its two speakers to eliminate pinna confusion for important central (solo) voices compared to stereo's problematic phantom imaging. Together with Ambisonics, the hybrid reproduces full-sphere 3D with height plus an accurate front stage. The hybrid speaker layout is illustrated in Fig.6 and Fig.7.

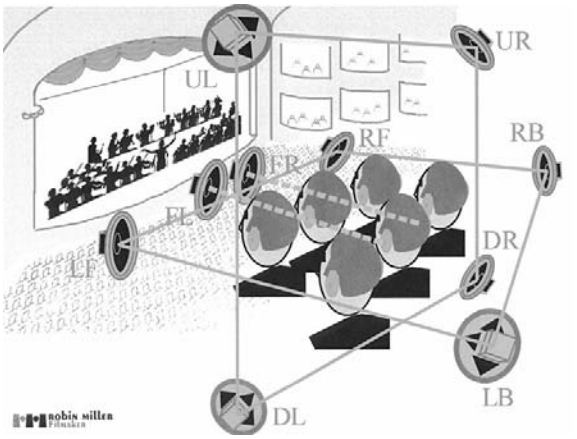

Fig.6 - HSD-3D speaker layout for compatible reproduction of legacy stereo, 5.1 2D surround, and full sphere 3D (with height).

Fig.7 - Demonstration "home theater" (7 of 10 speakers visible) for replay of stereo, 5.1, and HSD-3D (full sphere, with height).

1.2.5 HSD-3D and compatible 5.1 & stereo)

The author has developed High Sonic Definition 3D (Pat. pending, formerly termed PerAmbio 3D) that is compatible with stereo and 2D surround requiring no decoder, and using standard 6-channel media. At any time, the user can add a decoder and ten speakers to reproduce full-sphere (with height) 3D from the same disc. The HSD-3D system is illustrated in Fig.9. Results down-mixed in stereo or 2D surround are not compromised. If producers were to adopt HSD technology, they could release content in stereo (including iPod), 5.1/6.1/7.1 surround, and HSD-3D. Both producer and user libraries are not made obsolete;

and legacy stereo and surround play compatibly on the HSD-3D speaker layout [1].

The method for producing HSD-3D content begins with the HSD microphone above and adds a soundfield microphone using either a Soundfield brand model or discrete elements, as in Fig.1. As in Fig.9, six signals comprise the periphonic signals {Pin} required by the encoder, which transforms them to standard (2D) surround-compatible signals {S} for distribution using standard disc media such as DVD-A, SACD, Dual-Disc, or DTS-ES Discrete CD. This disc plays without a decoder on standard surround home theaters, or in stereo (including personal players such as iPod). When ready, the user purchases an HSD-3D decoder, perhaps embedded in an advanced multi-channel audio receiver or gaming console. Ten speakers may be flexibly positioned by telling the decoder where they are in 3-space. (An advanced receiver with measurement microphone might perform automatically this speaker locating, along with level and delay calibration.) The 10 speakers for HSD-3D are shown in Fig.6 & Fig7.

1.2.6 Recording HSD-3D

Recording HSD-3D is at once more complex and simpler than conventional recording. It is simpler (less costly) for the reasons already described in the sections above on stereo and 2D surround recording. Because the "perfect omni-directionality" of the HSD main microphone horizontally, the array may be positioned at or beyond the critical distance of the recording space, obviating or reducing the need for spot microphones and their associated artifacts and costs in labor in deploying, recording, editing, and mixing. 3D enhances perception from envelopment to full immersion. Capture simply adds to the basic HSD microphone a soundfield array to the same microphone stand and adds four channels to the hard drive (total 8). The complexity is associated with 1) precise calibration of the entire array, and 2) precise mic positioning and monitoring for the best final results, now that even untrained ears can discriminate the obviously life-like result now possible. After more than 40 years as a recording engineer practicing conventional methods, the author managed to relearn the new techniques required for HSD-3D, so he assumes that most others also will be capable of it!

1.2.7 Producing compatible HSD-3D/2D

Fig.10 shows the process flow for HSD-3D production, linking the three phases separated in time for Record, Post, and Consumer replay. Already described, the Record phase captures periphonic signals from the HSD-3D microphone as digital audio data using multi-channel audio software, such as on a Digital Audio Workstation (DAW). A transportable recording kit built in 2003 is seen in Fig.2, although a battery-operated portable recorder (Zaxcom Deva) has also been used, and a laptop-based solution is being implemented. In most cases, the equipment has been packaged in from four to eight cases or bags, transported in a small car or van, and operated by the author alone or with one assistant.

The Post-Production phase includes editing, mixing, and mastering for distribution to the consumer using digital media (e.g. 6.0 DTS-ES Discrete CD, DVD-V, DVD-A/DualDisc, Internet datastream). In a purpose-built control room shown in Fig.8 that also includes bass management and two subwoofers, monitoring is switchable between stereo (2 speakers or earphones), surround (5, 6 or 7 speakers), and 3D (10 speakers). In general, techniques common to DAW operation apply in selecting, assembling, minimally processing, and mixing tracks for the master. In particular depending on the content, one of six recording modes is selected that transforms the 3D signals into 2D for replay without a decoder, and for reconstituting the 3D signals when a decoder and height speakers are added [1].

Fig.8 - Multi-format control room (7 of 19 speakers visible) for mixing stereo, 5.1, and HSD-3D (full sphere, with height).

Fig.11 illustrates the HSD-3D encoder and decoder (Pat. pending). The encoder is a software-based plug-in used in the DAW above. As described in prior papers, it "maps" 3D signals into 2D according to one of six or more "modes" labeled i, j, k, i', j', k' [1]. The figure shows dual monitoring in stereo/2D surround or 3D, requiring hardware DSP-based decoder and 3D speaker layout matrix that emulates these functions in the consumer replay phase. The decoder changes mode either manually or using metadata embedded within the distributed media, which may be updated from time to time via Internet connection. DSP reconstitutes the 3D signals from the 6-channel disc and processes a scalable number of speaker feed signals for the 3D speakers

(minimum 10, 14, 26, or more). Speakers may be placed flexibly and the decoder told where in 3-space they are positioned. It is hoped that, using a calibration microphone and DSP-based process, the consumer's layout will be calibrated automatically, both when initially installed and whenever a significant change has been made in the listening environment. A small HSD-equipped control room with 19 speakers for stereo, 6.1 surround, and HSD-3D is shown in Fig.8.

The consumer phase in Fig.11 shows two options depending upon the status of the user's system: conventional 2D surround or with the HSD decoder and additional speakers for 3D. The disc as purchased plays without decoder on any 5.1/6.1/7.1 surround (2D) or stereo layout. Then when ready to upgrade to 3D, the user adds height speakers and the HSD-3D decoder, such as might be embedded on a DSP chip within an advanced multi-channel audio receiver. As above, the decoder changes mode either manually or using metadata embedded within the distributed media, which may be updated from time to time via Internet connection. DSP reconstitutes the 3D signals from the 6-channel disc and processes speaker feed signals for the 3D speakers (minimum 10, 14, 26, or more). Speakers may be placed flexibly and the decoder told where in 3-space they are positioned. It is hoped that, using a calibration microphone and DSP-based process, the consumer's layout will be calibrated automatically, both when initially installed and whenever a significant change has been made. A listener in the HSD demonstration room with 10 speakers and modest home-cinema/gaming capabilities is shown in Fig.7.

CONVOLVED AMBIENCE V. TONE COLOR

If impulse responses representing room acoustics can be captured, they can be convolved with relatively "dry" recordings of sources to yield immersive results. Ambiophonics championed by Glasgal [9] has demonstrated that a vast number of existing 2-channel stereo recordings in existence can be successfully reproduced in surround by ambience convolution.

Small in file size, 2D or 3D hall impulse responses such as those collected in some of the world's finest halls by Farina [10] can be distributed on digital media or downloaded from the Internet and stored in a library within the consumer's system. While computationally intensive, multi-channel convolution is readily accomplished in DSP chips or PC-based media centers. Unlike conventional "reverb," convolution has great potential to exactly recreate ambience (not direct sources) without committing recording equipment or media channels to deliver surround sound.

A limitation is the quality of IR that can be captured, as the ear is sensitive to artifacts introduced by the process, which is inexact. In particular, the room IR has superimposed on it the IR of the measurement speaker, which is not ideal. It is critical that the tone color of the speaker must be deconvolved from the hall IR as measured, but in experiments the results of convolved ambience is not subjectively equal to ambience recorded directly and requiring media channels as described in this paper. Also, the tone color imparted by convolved ambience varies with the user's selection and might not be that intended by the musicians.

Still, the potential of "up-mixing" stereo to an acceptable form of surround such as for archival recordings is highly desirable functionality. Solutions in development for implementing an "Ambiovolver" for legacy stereo, also including cross-talk cancellation for the Ambiophonic front stage, can be found at www.ambiophonics.org.

CONCLUSIONS

Audio sourced by panning close microphone signals pretends the "sound sources are here," where spatiality is that of the listening room, and so is largely invarying. For more varying realism and immersive reproduction of a more natural world of acoustic music performances, movie atmospheres, and gaming effects, recording and reproduction that captures and preserves directionality of both direct sounds and acoustic reflections is critical. At a typical live listening perspective beyond the critical radius of the space, these indirect sounds have energy that exceeds direct sounds, and include both directional reflections occur within the time constant of the space that are critical for accurate localization and tone color, and non-directional diffuse reverberation. Non-ambient recordings may be convolved with hall impulse responses for a "first approximation" of surround ambience, if not more "realistic" tone color, without the need for recording surround directly or dedicating media channels. Where sources as well as important reflections come from other than the front, and for unartifacted tone color and ambience, direct surround recording, on dedicated media channels, and preserving directionality during reproduction is critical. New recording techniques are described, applicable compatibly to 5.1/6.1/7.1 where reproduction is a 2D horizontal circle of speakers, to stereo (including personal devices using earphones e.g. iPod), and to future 3D reproduction, where the listener is again at the center of the sphere of natural hearing. Termed High Sonic Definition (HSD), this technology offers recording engineers new tools and content producers new opportunities.

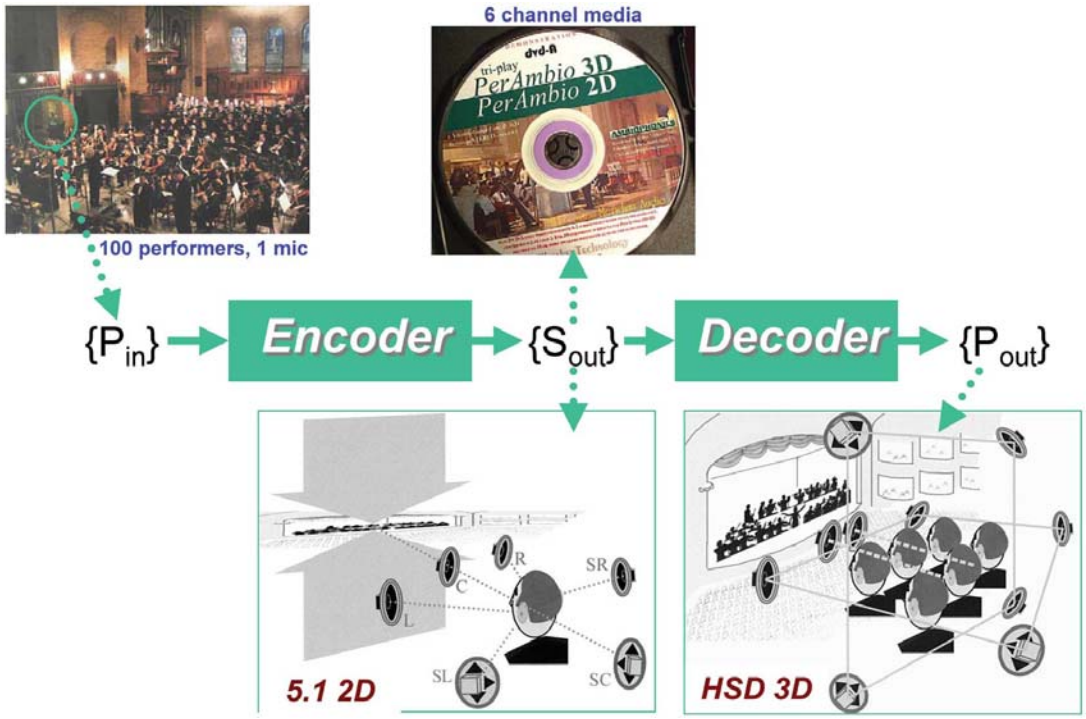

Fig.9 - High Sonic Definition (HSD, Pat. pending) 3D system. Signals from the HSD microphone are encoded to 6-channel media, playable in 5.1~7.1 without decoder. When ready for full sphere (with height) 3D, the user adds a decoder and speakers (10 total).

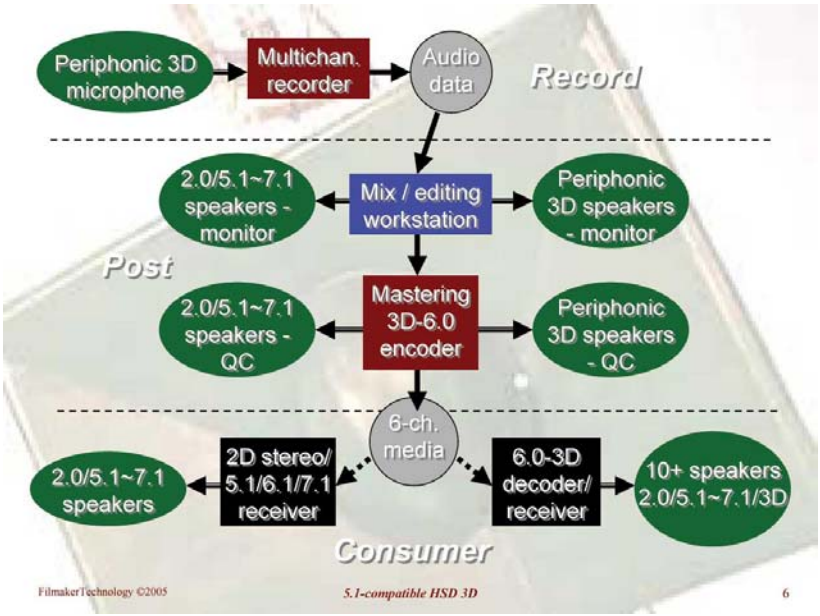

Fig.10 - Process flow for recording, post-production, and consumer replay of 5.1/stereo-compatible HSD-3D.

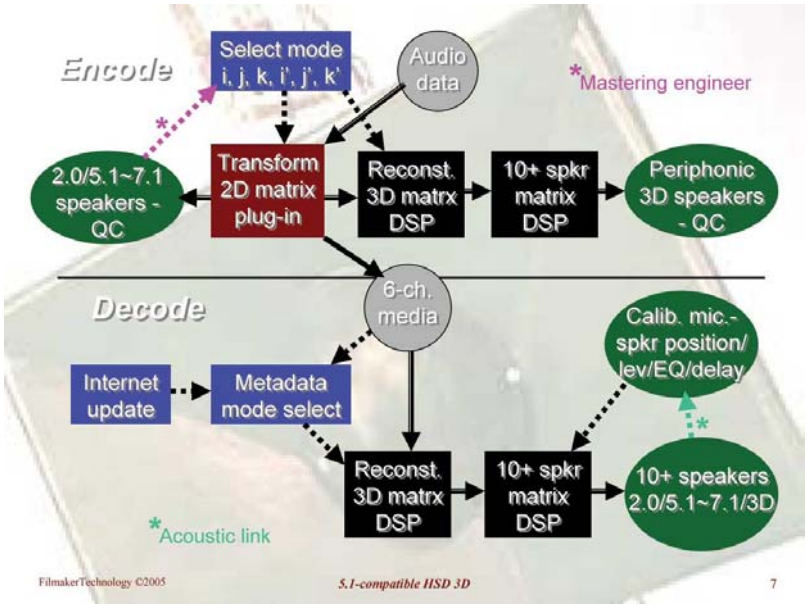

Fig.11 - software encoder (producer) and hardware DSP decoder (consumer) for 5.1/stereo-compatible HSD-3D (Pat. pending).

REFERENCES

[1] Miller, R, "Scalable Tri-play Recording for Stereo, ITU 5.1/6.1 2D, and Periphonic 3D (with Height) Compatible Surround Sound Reproduction," AES 115th Int'l Conv., New York City, USA, October 2003 – preprint #5934.

[2] Theile, G, "On the Naturalness of Two-Channel Stereo Sound." J. Audio Eng. Society, Oct. 1991.

[3] Theile, G, "Natural 5.1 Music Recording Based on Psychoacoustic Principles," Proceedings of the AES 19th Int'l Conf., Schloss Elmau, Germany, rev. 10/2001.

[4] Glasgal, R, "Improving 5.1 and Stereophonic Mastering/Monitoring by Using Ambiophonic Techniques," International Tonmeister Symposium, Schloss Hohenkammer, Germany, Oct 2005

[5] Miller, R, "Physiological and content considerations for a second low frequency channel for bass management, subwoofers, and LFE," 23rd VDT (German Tonmeisters), Leipzig, Germany, Nov. 2004

[6] Miller, R, "Physiological and content considerations for a second low frequency channel for bass management, subwoofers, and LFE," 149th ASA /CAA Convention, Vancouver, Canada, May. 2005

[7] Miller, R, "Physiological and content considerations for a second low frequency channel for bass management, subwoofers, and LFE," 119th AES Convention, New York City, USA, Oct. 2005, preprint #6628

[8] Miller, R, "Spatial Definition and the PanAmbiophone Microphone Array for 2D Surround & 3D fully Periphonic Recording," presented at AES 117th Int'l Conv., San Francisco Oct.2004, preprint #6253

[9] Glasgal, R, "Ambiophonics: Achieving Physiological Realism in Music Recording and Reproduction," Proceedings of AES 111th Convention, preprint 5426.

[10] Farina, A; Ayalon, R, "Recording concert hall acoustics for posterity," 24th AES Conference on Multichannel Audio, Banff, Canada, 26-28 June 2003

AUTHOR

Robin Miller, BSEE, AES, SMPTE is a musician and orchestrator and filmmaker recognized by 52 awards including The Peabody. He has more than 40 years experience in music recording and mixing films and television specials.