Audio Engineering Society

Convention Paper 6959

Presented at the 121st Convention
2006 October 5–8 San Francisco, CA, USA

5.1 surround and 3D (full sphere with height) reproduction for interactive gaming and training simulation

Robert (Robin) E Miller III[1] ©2006

[1] FilmakerTechnology, Bethlehem Pennsylvania 18018 USA
www.filmaker.com

ABSTRACT

Immersive sound for gaming and simulation, perhaps more than for music and movies, requires preserving directionality of direct sounds, both fixed and moving, and acoustical reflections dynamically affecting those sounds, to effect the spatiality being presented. Conventionally (as with popular music), sources are panned close-microphone signals or synthesized sounds; the presentation pretends "They are here," where spatiality is largely that of the listening environment. Convolution with room impulse responses can contribute diffuse ambience but not "real" spatiality and tone color. These issues pertain not only to 5.1 where reproduction is a 2D horizontal circle of speakers, but to advanced 3D interactive reproduction, where the listener perceives the experience at the center of the sphere of natural hearing. Production techniques are introduced that satisfy both 3D and compatible 5.1. Independent measurement confirms that the system preserves directionality and reproduces life-like spatiality and tone color continuously in the 3D perception sphere.

1. INTRODUCTION & PURPOSE

The objective of audio reproduction for popular music is a presentation that is "in your face." In contrast, reproduction for gaming and related training simulation seeks to create the illusion "in a space." To transport subjects to another space and time – like "being there" (training scene or game action). Perhaps more than for music and movies, we must reproduce the venue as well as the sources it contains. The more appropriate and natural the recreated acoustical environment is to the scene – the more compellingly "immersive" (involving) – the more successful the content will likely be. How better to approach audio for gaming, VR, training simulation, and theme park rides is the subject of this paper.

However, to increasingly sophisticated ears, conventional audio technology in this regard is imperfect, so that a realistic impression must result from what is actually "acoustic fiction" [1]. Producing spatial audio content in 5.1 / 6.1 / 7.1 (hereafter "5.1") surround sound in ways that can suspend disbelief have been described by, among others, Theile [2,3], Holman [4], Griesinger [5], Glasgal [6,7], and the author [8,9]. But typically, the practice of surround production falls short of its greater potential by following stereo's conventions – panning closely-mic'd monaural sources (and possibly adding artificial reverberation). For critical gaming and training simulation, we must rethink these conventions. Ultimately, we must think outside the 2D circle of 5.1 surround sound to the 3D sphere of natural hearing perception.

2. PROBLEM DEFINITION

Self-imposing limits on meeting the objective of believable sound for gaming and simulation, creators source much if not most stereo and 5.1 surround content from monaural, non-spatial microphone signal(s) or synthesized sounds that are panned for surround speaker layouts. In addition, with the ITU standard 5.1 surround layout, the listener's perception is constrained to a 2-dimensional circle incapable of reproducing height information (usually acoustic reflections that contribute tone coloration according to the listener's unique HRTF). While uncorrelated artificial reverberation or ambience microphones may add a semblance of "spatiality," usually this kind of presentation, whether intended or not, is perceived by the listener as "They are here," where the dominant correlated spatiality is that of the immediate listening environment, which does not change from program to program or sound to sound, and where sounds may move but unnaturally their reflections do not. In the past considered "intimate" by music reviewers and audiences, a lack of true spatiality may be decreasingly successful in the market as increasingly discriminating gamers and training subjects as well as music listeners and moviegoers all become more demanding of more immersive experiences.

On the other hand, naturally ambient reproduction for gaming and training, also desirable for movie ambience and acoustic music (art, classical, jazz, folk, etc.), requires special techniques that capture both direct sound of sources and the acoustical reflections that are the "extension" of each sound source, such as a musician's instrument, an actor's voice, or a gun's shot. Every venue acoustically convolves each source sound with unique, position-dependent impulse responses. While in theory "dry" microphone channels may be convolved with separately measured hall impulse responses, in practice, as discussed later, it is difficult to achieve results precisely enough to suspend disbelief, especially when sounds move.

One example that avoids the problem is the training simulator that is a precise acoustic replica, such as an actual cockpit mock-up, with its actual sources of sound at fixed positions, such as audible indicators and alarms. In this case, the actual sources can be triggered.

Immersive quality differs between 2-dimensional 5.1 where reproduction is a circle of speakers more or less in the horizontal plane, and lifelike 3D, where the listener is again at the center of the sphere of natural hearing. In acoustic spaces, mostly reflection arrivals from above and below are "coded" by our individual HRTF, especially the pinnae, to be perceivable according to learned response to height information. The complex integration of each arrival over the time constant of the acoustic space (approx. $R_T/4$) results in the ultimate tonality of the sound as perceived by each individual. If upon reproduction these arrival directionalities are not preserved in 3-space, the listener is aware that they are listening to "just a recording." Instead, with 3D reproduction that preserves directionality of all these arrivals, the result is more lifelike and realistic.

A compatible 3D/2D system of recording and reproduction applicable to gaming and simulation, termed "High Sonic Definition 3D" or HSD-3D, has been described, previously termed PerAmbio 3D/2D (Patent pending) [10]. Briefly described below, the system transforms periphonic microphone signals into standard 5.1 for uncompromised mastering and distribution on ordinary six-channel media (DVD-A, SACD, DTS-ES CD, Dual-disc etc.) for conventional 2D surround replay without a decoder or additional speakers. Then at any future time by adding a decoder and flexibly positioned speakers, the full sphere of the originally recorded 3D sound is reconstituted.

Experimental recordings in compatible HSD 3D and ITU 5.1 2D have been made by the author and demonstrated, including comparisons with OCT presented at the 24th AES Conference in Banff, Canada, in June 2003 and in conjunction with AES119 in New York City in 2005, including compelling 3D gaming and simulation effects. Experiments by a group of honors students at Lehigh University with the author subjectively measured the virtues of 3D reproduction using 30 subjects, described later. These new and adapted techniques for capturing, processing, and reproducing lifelike sound for gaming and virtual reality (VR) and for related training simulation and theme park rides, as well as for music and movies, are explored both for future 3D and compatible 2D ITU 5.1.

3. LOCATING SOURCES, FIXED AND MOVING, IN 3-SPACE – DISCUSSION

It is known among many investigators that a subject's ability to localize a source of sound reproduced either in stereo or 5.1 surround depends on many factors and at best is imperfect. In fact, the poor localization of phantom images characteristic of stereo has caused

many practitioners to compromise localization in favor of heightened spaciousness, which in many ways is a tradeoff to localization. For example, recordings made with spaced microphones are swimming in diffuse spaciousness, but a musical instrument will appear to jump from place to place as notes of different frequencies are played. Broadband sources appear to be spectrally broadened, if not torn.

Many have attempted to develop panning laws for both stereo and surround that position sources between the speakers in stereo and around and behind in surround, and that work for listeners both in the "sweet spot" and those in off-center seats. While localization is often compromised in music content, it is perhaps more important in gaming and related simulation, where positioning of sources in space must closely match either that on a visual display or what the trainee will encounter in reality

In addition, successful localization is essential for moving sources, whether it is a sword whipping around or enemy fighters traversing the sphere of perception. Inaccurate positioning moment by moment renders the illusion less successful. In some ways, sounds that are moving help localization inaccuracies, as the source passes from a region of inaccuracy through other positions that renders it more realistically. Subjective testing in a later section encountered this phenomenon.

It is well know that the 5.1 standard was an intentional compromise especially in terms of localization accuracy, yielding to important considerations of compatibility with cinema content sound and market acceptance. In addition, 5.1 surround consists of speakers positioned around a circle in the horizontal plane, and so does not contain height information that completes the sphere of perception. While 5.1 surround can be enveloping, full sphere 3D is more truly immersive, as is natural hearing.

Consider how a sound might move in 3-space. In the simplest case, a sound moving radially away from the listener will become softer, lose high frequency energy (due to absorption in air), and become more reverberant. A sound moving across the horizon will move through points where there are actual speakers, through many more points as a phantom, traverse differing regions of pinna coloration, are tracked by changes in interaural level (ILD) and interaural phase (ITD), increase and decrease in frequency due to Doppler effects, and be trailed by a complex of early reflections, each having all these characteristics on their own, plus might naturally come from above or below. Most complex are those sources that have vertical travel, tracked mainly by pinna coloration changes, in addition to those above. Preserving localization both statically and dynamically is essential in gaming and simulation.

4. SPATIALITY IN STEREO, 2D 5.1/6,1/7.1 SURROUND, AND FULL SPHERE 3D

Whether a performer-listener or audience-listener, almost no one enjoys totally anechoic sound, the direct-only emission of an instrument, sound effect, or voice. For this reason, we far prefer an orchestra in a concert hall, the click of a gun's safety in a cave, a signal near the helm of a submarine, or our voice in the shower. Spatiality encompasses qualities that are temporal and HRTF-related in an acoustically active sound field.

Time out for one caveat: Many audio engineers view their role as ultimately the creator of the recording who uses his/her tools to make something "better than real," if not at least novel or more saleable. This is a valid argument. However, too many recording engineers, practicing their craft only in the control room, lose their reference for what uncompressed, uncolored, natural audio sounds like [11]. What is presented below is intended as a basis for understanding, whether used for capturing reality or for intentionally departing from it.

4.1. Lifelike spatiality and Tone Color as perceived by performers & audience

It is known that musicians play differently in different spaces, which act acoustically as the "other half" of their instruments. The tone color of a musical note, sound, or speaking isn't merely altered by acoustics, it is the result of interactions between acoustics and listeners, both patrons in the audience and performers on-stage. Although the acoustics on-stage can be different than in the house, with experience, musicians learn how to interpret what they hear in order to create sound that pleases their audience.

The instantaneous differences at listeners' ears are the dynamically changing differences in arrival time, arrival direction, and individual pinna filtering. These partially correlated inter-aural differences are processed in the mid-brain and perceived in consciousness as localization, tone color, and spatiality.

4.1.1. Contradictory spatiality

If we don't have handy a fine-sounding space, recordists conventionally substitute artificial reverberation, or more advanced hall impulse response convolution. However, this "synthesized spatiality" usually treats all sources fed into it the same, which to many falls short of believable results. Also, its use might contradict the artists' intentions. Recording engineers who are not also musicians (i.e. "Tonmeisters"), actors, or Foley artists might not appreciate that these artists "play the space," performing differently for each venue. Performing in an acoustically "dry" studio may adversely affect results; why achieving ensemble is difficult when wearing headphones in separately scheduled sessions, or why dubbed dialogue sounds phony. Economics drives these expedients. But such artificiality may explain why, whether for popular or classical music, live concert recordings often excel in musical energy, even given the occasional musical clam or audience noise.

4.1.2. Spatiality that is "real"

Spatiality, the result of reverberant energy of the space, does not come from the same direction as the musicians or actors, i.e. the front stage, but from all around and even above and below. Now in the realm of psycho-acoustics, this full sphere of sound is direction-coded by our individual ears (especially the outer ears, or pinnae), processed by the mid-brain, compared to our experience in memory, and output to higher consciousness to form perception. Spatiality is important because, in most places we find ourselves, the sum of spatiality's indirect energy exceeds the direct energy from sources. A life-like recording, therefore, would be so good that we could, if we were familiar with it, identify the hall in which it was recorded. Just such a thing happened during one of hundreds of demonstrations the author has given when Mark S., an audiophile and concertgoer, upon hearing a symphony orchestra concert recording, exclaimed: "I know this hall – is it Washington-Irving High School in Greenwich Village?" It was!

How was Mark able to identify an obscure auditorium in a recording? The reproduction was 3D, where the spatiality of the room was compellingly conveyed in ways stereo and 5.1 cannot, because the entire sphere of sound unique to that hall was preserved. Musicians too, hearing themselves in 3D surround, exclaim: "After all the [stereo] recordings I've made, finally this is my sound." [10] For both musicians and discerning listeners, it is not only to be able to recognize the "signature" of the room, but to preserve the tone color that the room contributed to each instrument so that, upon replay, life-like tone color contributes as intended to the music. (In these cases the surround system being demonstrated is High Sonic Definition 3D (HSD-3D) using 10 speakers – see www.filmaker.com.)

All this suggests that conveying any sound in recorded form implies conveying not only the direct sound sources and preserving their provenance, but also conveying the spatiality enclosing the live listener, which in turn implies surrounding the recording listener with those same indirect sounds and preserving their provenance. Preserving spatiality may apply even more to critical reproduction for gaming and training simulation than to music and movies. In contrast, reproducing spatiality from two front speakers as with stereo does not envelop the listener in a believable way. The total contribution to spatiality of surround sound – at least the horizontal circle of 5.1 if not 3D – will be more lifelike, truer to content providers' intentions, and so more successful.

4.1.3. Spatial cues: Inter Aural Differences

Briefly, a listener's conscious receives complex signals from the ears as processed in the pons and mid-brain, and confirms spatiality by other senses, mostly what is seen. Thus for filmmakers and game developers the importance adage: "See the scene / Hear the scene."

Above approx. 700Hz, human binaural hearing detects inter-aural level differences (ILD) that are affected by source position and listener's head-related transfer function (HRTF). In circumscribing the sphere of perception by azimuth and elevation, the most active ILD cue horizontally is head-shadow attenuation. Vertically, pinnae comb filtering is the dominant player.

Below approx. 700Hz, human binaural hearing crosses over to detect time differences (ITD) – either onset time-of-arrival or phase – to a lower limit by convention of 90Hz, below which hearing is monaural. Recent studies in this "VLF" (very low frequency) range <100Hz argue for lowering that limit one octave to approx. 45Hz, with spatiality implications in recording and, for critical reproduction, a need for binaural bass management and two subwoofers [12,13,14].

4.1.4. Spatiality by impulse response convolution

Fire a gun in an acoustic space and its 3-dimensional impulse response (IR) can be measured. Then (even years apart) convolve a dry, even monaural, direct sound with the multi-channel IR and obtain multi-channel surround signals that present the sound with the spatiality one would have heard in the original space. The process is performed either in post-production or by the end consumer upon replay. If the presented space changes, or needs improvement, simply substitute new or IRs, or IRs of a better room, and re-convolve. [15]

Mathematically elegant algorithms and efficient digital signal processing (DSP) in software or hardware have made low-latency convolution practical. And it has been demonstrated to be an effective expedient for late (diffuse) reverberation. However for synthesizing spatiality and tone color (implying early reflections), it must be acknowledged that obtaining multi-channel IRs precise enough to fool even the average gamer or simulator trainee is difficult. Precision is limited by the non-ideal measurement loudspeaker when using the swept-sine technique. Also, except for very sophisticated "room simulators," IR measurements for convolution are usually taken at only one or two positions within the venue, so only one or two IR "signatures" is applied regardless of where in the venue the source is located. In natural hearing, the room effect varies infinitely for moving sounds or multiple sources, implying different and possibly changing IRs for every source position. Rather than convolution, direct recording of spatial information avoids both these problems, and therefore is preferable for critical quality.

5. CONTENT FOR GAMING & SIMULATION IN SURROUND – DISCUSSION

As stated at the outset, typically considerations of localization and spatiality above are ignored in the practice of surround production that follow stereo's conventions (panning closely-mic'd monaural sources and adding artificial reverberation). To be more successful, we must rethink conventions from the point of view of the audience in order to be true to the venue to be reproduced as well as the sources it contains.

For gaming and simulation other than the exception above where the simulator is an exact replica, consider the applicability of surround sound when the content to be produced and consumed is describable according to one of the following hierarchical forms:

1. Content suggests an enveloping nature, such as immersing the listener within a gaming scene or training environment (non-anechoic space), even if direct sources appear to be only frontal (staged);

2. Sources of sounds are distributed around 360° horizontally (e.g. limited to the ground), perhaps interactively controlled in 2D by programming or by the user, who wants them to be localized believably;

3. Coming from or moving in all directions in 3-space, both direct sources and indirect (reflected) sounds correlate naturally to recreate in the user's perception verisimilitude – a sonically life-like simulation of reality.

Item 1) describes frontally oriented content, such as movies and staged musical performances. Note that in stereo, any recorded spatiality (from room mics or reverberators) is folded into the front 60° between speakers. The only enveloping spatiality is that of the listening room, which is constant regardless of content and where sounds may "move" but unnaturally their reflections do not, which is probably inconsistent with visual cues, and which therefore is not compelling.

In 2), interactive gaming or simulation might require sounds to originate throughout 360° horizontally. Typically (as with much popular music) these are monaural channels assigned to one of 5 or more speakers or panned between two speakers to produce a phantom image (with degraded tone color and location artifacts). Ideally, these sources would be naturally colored by correlated spatiality from the other speakers, to a limited extent if 2D as in 5.1 surround. Convolved spatiality may improve 2D reproduction and may add height speakers to simulate 3D. However, if the only spatial information is that of the listening room, which is constant, probably inconsistent with visual cures, it will therefore not be compelling. (Yet most content advertised as "surround" is produced this way.)

Step 3) fully realizes the potential of surround in listening spaces that are sufficiently acoustically controlled that they do not interfere with the spatiality conveyed in the production content. Using fully periphonic (with height) 3D reproduction, the full sphere of natural hearing perception is most compelling.

5.1. Surround sound for gaming

As of this writing, game content is still released and played mainly in 2 channel stereo. Newer content is being released in 5.1 surround, and game consoles have 5.1 surround replay capability. PC audio cards and PC speaker systems are available with 5.1/6.1 outputs and speakers. Typically, surround content appears to be panned monaural sources with artificial reverberation in the surround channels. Mostly untapped is the potential to pan sound interactively as the game is played.

If, as is the case with most gaming, the subject is in a fixed position, then advanced reproduction systems, described below, e.g. Panor-Ambiophonic (PanAmbio, 4 speakers) can provide superior localization accuracy around 360° (Fig.1), and avoid tone color variations, especially of important central sounds, for greater illusion of reality. Ultimately, a 3D presentation such as HSD-3D (10 speakers) can present sounds in the full sphere of human hearing perception. Using 4 or 10 small speakers respectively, performance can be full range using bass management and subwoofer(s).

As mostly young gamers buy or build houses with home theaters, the convergence of home cinema, domestic concert hall, and gaming may mean that this purpose-built room with large screen and high quality surround sound will be used for gaming as well as movies and music. Therefore, content for the "media room" would need to measure up to the user's new expectations, now under the microscope of high definition picture (HDTV) as well as "high definition sound." Preliminary testing of 3D game sounds with 30 subjects is discussed below.

Related to electronic games is virtual reality (VR) using goggles to present visuals that rotate with user movements. In this case, headphones conveying binaural audio should also be controlled to rotate the soundfield using head-tracking by sensing head rotation.

5.2. Surround sound for simulation

Benefiting from both the creativity of game content producers and the proficiency of users gained over years of practice is training simulation for military, aviation, heavy equipment operation, etc. Unless, as discussed above, the simulator is an exact replica acoustically, 2D 5.1 surround audio offers improvement over stereo in the illusion of reality that is possible. If, as is the case for most gaming, the subject is in a fixed position, then advanced 2D reproduction systems, e.g. 4-speaker Panor-Ambiophonics, described below, can provide superior localization accuracy around 360° (Fig.1), and avoid tone color variations, especially of important central sounds, for greater reality. Ultimately, a 3D presentation such as HSD-3D (10 speakers) can present sounds in the full sphere of human hearing perception.

5.3. Surround sound for theme park rides

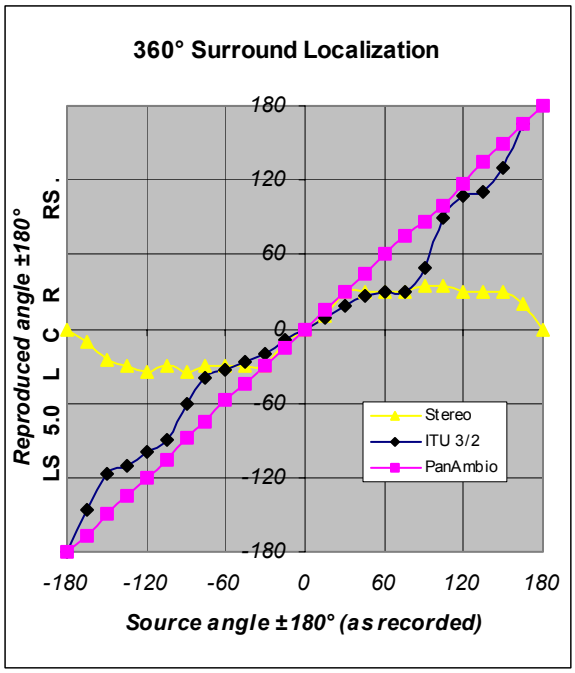
Again if the ride consists of one or two in fixed listening position(s) in a semi-enclosed moving cart, advanced sound techniques such as Panor-Ambiophonics (PanAmbio) and HSD-3D are possible. Using 4 or 10 small speakers respectively, performance can be full range using bass management and subwoofer(s).

5.4. Interactive sound field rotation

For console or PC gaming, simulation, and theme park rides, providers may want to create content that allows users to control interactively, using a joystick or similar device, both audible space as well as visual space. 360° horizontal rotation of 2D surround, either 5.1 PanAmbio (4 speakers) is possible using Ambisonic or 4.1/5.1-compatible HSD-3D techniques [8]. HSD-3D (10 speakers) is a special case where the entire perception sphere can be pitched, yawed, and rolled under control either by the user or by the program or ride.

6. 5.1 / 6.1 / 7.1 SURROUND SOUND (2D) FOR GAMING & SIMULATION

ITU standard 5.1 surround sound, born of the cinema, works best for a solidly localized front stage (L,C,R) plus decorrelated ambience around back (SL,SR). Precise localization in back of the cinema is not possible with multiple surround speakers. In home cinema and possibly with gaming pods and training simulators using one speaker for each surround channel, localization is good for sounds panned to coincide with any individual speakers, including one of the two surround speakers, but the 140° separation is not conducive to creating reliable phantom images between back speakers, nor can images be created reliably on either side between L and SL or between R and SR pairs. Surround panning that results in a correlated sound emanating from all speakers diffuses its localization and mars tone color due to comb filtering. Pair-wise panning, as though each pair of the five speakers were stereo, produces only fair results [1]. Precise localization around 360°, or indeed in the full sphere of natural 3D perception, requires a different reproduction system.



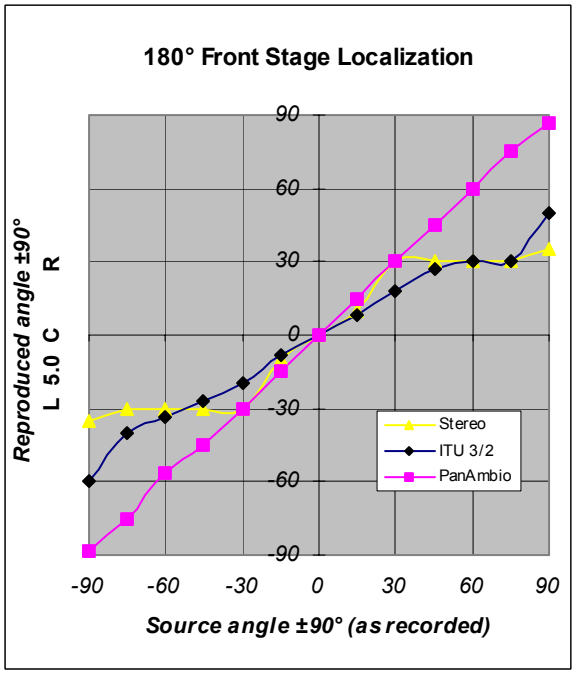


Fig.1 - Perceived localization around a) entire 360° horizontal plane and b) 180° front stage - ITU 5.1 vs. PanAmbio. ITU 3/2 is "ambiguous" at ±90°, ±105°, ±120°, and ±150°. PanAmbio approaches an ideal straight line but is "fuzzy" near ±90°.

Nevertheless, 5.1 and derivatives 6.1 and 7.1 are a great improvement over 2-channel stereo in recreating natural envelopment, if not precise localization. In addition to important approaches mainly by European broadcast entities, summarized by Rumsey in [1], Williams in [16] and since has experimented extensively to create main microphone arrays (5 microphones on one stand), as has the author in [8,9] and others in ongoing attempts to improve 5.1 spatiality in practice.

6.1. 2D alternative: 4.0 Panor-Ambiophonic (PanAmbio) surround

If the gamer or trainee is in a fixed position, such as seated before a PC display, then 360° imaging can be recreated using a pair of closely-spaced speakers in front and another in back of the subject fed by pairs of crosstalk-cancelled binaural-based signals [17]. Termed PanAmbio for short, the advantages are:

- Good localization accuracy around (except at sides coinciding with natural cone-of-confusion);
- Natural spatiality and tone color due to accurately preserving early reflections horizontally 360°;
- Undistorted tone color of important center images (cf. stereo's comb filtering of phantom images);
- Recreated sound field remains stationary with the visual display without the need for head-tracking.

Fig.1 illustrates the linearity of localization perception around 360° horizontally of PanAmbio compared to 5.1 reproduction of voice and band limited pink noise, described in detail in [17]. The approach works best with purpose-made recordings, although it is quite acceptable for much existing multi-channel music and movie content (with player set for "no center speaker"). Information and tools at www.ambiophonics.org.

7. BEYOND 2D 5.1 / 6.1 / 7.1 SURROUND – FULL-SPHERE 3D (WITH HEIGHT)

Just as with high definition images, high definition sound would make gaming content more compelling and simulator training more effective. In the present context, "high definition" refers not just to high sample rate, but more to "high spatial definition" – the life-like qualities described above of accurately localizing sources and perceiving "real" tone color and spatiality in agreement with vision. Under interactive control of the user or programmed in the simulator, sounds with or without an associated picture can be pitched, yawed, and rolled under control either by the user or by the

program or ride. Sounds throughout the perception sphere are localized accurately and vary minimally in tone color so as to immerse the gamer or trainee continuously and, therefore, to enhance his/her experience or learning as though experienced for real.

Since human hearing localizes more accurately in the horizontal plane than vertically (at best on the order of ±1° cf. ±10°), it seems to follow that the jump from stereo to 2D 5.1 surround is significant, while a the addition of height for 3D would be less so. However having observed the responses of hundreds of subjects in 2D v. 3D demonstrations, and the independent tests below, the author has concluded that reproducing the full-sphere 3D of natural hearing is not just a subtle improvement. Think of 2D surround as a disc in the horizontal plane. Then imagine that disc ballooning upward and downward until it becomes a ball. 3D indeed elevates audio to another dimension.

7.1. Practical 3D (5.1-extensible)

In developing High Sonic Definition 3D (HSD-3D) [10], the intent in brief was:

- A practical approach with modest implementation costs for consumers (a decoder and 10 speakers);
- Compatibility forward and backward with 5.1 and stereo so that neither producers' libraries nor consumer's collections would become obsolete;
- Simplified production tools & techniques with cost-justifications for content providers.

Fig.4 illustrates the HSD-3 system (Pat.pending) that accurately captures 3D acoustic signals, transforms them to compatible 5.1 for distribution on standard media, and reconstitutes 3D speaker signals when the user is ready to add the decoder and extra speakers. The speaker layout is compatible with legacy stereo and 5.1 recordings by simply moving back 26% of the speaker diameter. 10 speakers is the minimum for simulating a sphere (positioned flexibly by telling the decoder where they are), but the system is scalable to 14 or 26 speakers. As with stereo and 5.1 there is a "sweet spot," perhaps even more so when listeners expect higher precision results. However, 6 listeners can be accommodated with acceptable results within a 4ft (1.25m) square. Information at www.filmaker.com.

8. INDEPENDENT QUANTITATIVE & QUALITATIVE MEASUREMENT

As part of the Integrated Business-Engineering (IBE) program at Lehigh University, a team of six honors students conducted a survey involving 30 student subjects familiar with gaming. In groups of two seated in the focus positions ("sweet area"), each evaluated to what degree they perceived positional agreement of 38 spoken announcements as the voice "moved" from the last position to the next throughout the sphere of 3D listening. Subjects were asked to rank the end positions on a scale of 1 (strongly disagree) to 5 (strongly agree).

Characteristic of the HSD-3D reproduction system used, the voice could move along paths independent of the 10 discrete speaker locations. Some end positions were co-located with a speaker while others were not. Thus, the test was intended to demonstrate whether, using HSD-3D, a sound could be perceived as coming from virtually any point on the sphere of hearing perception. As described above, the ability of any system of reproduction to preserve the directional provenance in 3-space of both direct and reflected sounds implies lifelike localization, spatiality, and tone color.

8.1. Test results for gaming & simulation

More detailed results will be processed by the IBE Team for presentation of this paper at the AES 121st Convention. As of this writing, the overall average of 30 subjects times 38 responses ranked agreement a 4.39 out of a possible 5. Subjects expressed that end positions where they gave lower rankings were not necessarily associated with regions where no speaker existed. Instead, they were positions where the human hearing system is naturally ambiguous, including directly downward due to torso effects, or within the "cone-of-confusion" at each side. However to confirm provenance of sounds within these regions, subjects could turn their heads, as one does in normal hearing.

8.2. Test details; "motion" effects

As shown in Fig.2, the series of announcements were not fixed at points in the perception sphere, but exhibited motion from the last position to the next. While this animated characteristic might influence perception accuracy, it also might reveal unnatural tone color changes. Motion also more closely simulates practical application, where for example, gaming or simulator sounds could move throughout the perception

sphere either as programmed or as controlled by the subject using a joystick in real time. Consensus was that subjects clearly perceived this moving through space and that it added to the quality of the "content."

The subjects were also presented subjective materials including musical excerpts and swordplay sound effects. Again, detailed analysis is incomplete as of this writing. However consensus was that subjects found the 3D presentation both compelling and desirable.

HSD-3D localization, beginning FRONT & CENTER (az0°, el0°).
NOW MOVING THRU the SoundSphere to LEFT in FRONT (45,0).
Again moving through the SS thru FRONT & CENTER (0,0)...
...and continuing across the SS to RIGHT in FRONT (-45,0).
NOW RISING within the SS to RIGHT and UP in FRONT (-45,45).
...moving across the SS to CENTER UP in FRONT (0,45)...
Again moving across the SS to LEFT and UP in FRONT (45,45)...
...continuing leftward around the SS to LEFT and UP (90,45).
Moving across the SS thru DIRECTLY OVERHEAD (0,90)...
...and continuing across the SS to RIGHT and UP (-90,45).
On the move again descending past directly RIGHT (-90,0)...
...then continuing thru the SS to RIGHT and DOWN (-90,-45).
Now dropping thru the SS past directly DOWNWARD (0,-90)...
...and continuing below in the SS to LEFT and DOWN (90,-45)...
...then ascending within the SS and arriving directly LEFT (90, 0).
Starting again in the SS at LEFT and DOWN in FRONT (45,-45)...
...ascending within the SS toward LEFT in FRONT (45,0)...
...rising from the horizontal thru LEFT and UP in FRONT (45,45)...
...moving diagonally above, passing directly OVERHEAD (0,90)...
...continuing diagonally thru RIGHT and UP in BACK (-135,45)...
...continuing to descend thru RIGHT in BACK (-135,0)...
...finally settling at DOWN and RIGHT in BACK (-135,-45).
You may turn your head to verify we're RIGHT in BACK (-135,0)...
...and now moving around back past CENTER in BACK (180,0)...
...now stopping briefly within the SS at LEFT in BACK (135,0).
Faster now thru LEFT and UP in BACK (135,45)...
...passing diagonally OVERHEAD (0,90)...
...continuing RIGHT and UP in FRONT (-45,45)...
...and arriving at RIGHT in FRONT (-45,0).
Starting once more at FRONT & CENTER (0,0)...
...ascending again thru UP in FRONT (0,45)...
...passing directly OVERHEAD (0,90)...
...turn to verify we are UP in BACK (180,45)...
...continuing around thru CENTER in BACK (180,0)...
...descending now thru DOWN in BACK (180,-45)...
...continuing thru directly DOWNWARD (0,-90)...
...rising again past DOWN in FRONT (0,-45).
...finally returning to FRONT & CENTER (0,0).

Fig.2 – Test of voice moving throughout the perception sphere. Each announcement began at the last position and ended at azimuth and elevation shown, ranked by subjects' agreement.

9. IMPLEMENTING FULL-SPHERE 3D

The recording, encoding "transformation," and decoding "reconstitution" of HSD-3D (previously termed PerAmbio 2D/3D) is described in prior papers [8,10]. Briefly, it involves an 8-element main microphone array, DAW plug-in software encoder, and DSP decoder firmware (Pat. Pending). The system produces 6-channel recordings distributable using standard media (DVD-A/DualDisc, SACD, DTS-ES Discrete 6.1 CD or DVD-V) that are backward compatible with standard surround layouts (5.1/6.1/7.1) and stereo (including portable players e.g. iPod®). Legacy stereo and 5.1 recordings play compatibly on the 10-speaker HSD-3D layout and conform to ITU-R775 by repositioning the listener and changing speaker levels and delays. See Fig.4 for a simplified illustration of the HSD-3D system of capture, 5.1-compatible encoding, and 3D decoding for 10 speakers.

9.1. 3D production, encoding, & decoding

Productions in compatible 2D/3D involve either original HSD-3D recordings or synthesizing using DAW tools (in development) and monitoring using dual speaker layouts (Fig.3). Production cost savings reflect a system that, as a design goal, captures natural sound so as to require less post-production manipulation. Techniques are described in more detail in [8]. Fig.5 illustrates the production, editing/mixing, and mastering processes.

The software encoder for distribution and 5.1/6.1/7.1 compatibility is being developed as a DAW plug-in. The hardware decoder, embedded in DSP in audio receivers or processors, has been realized for an Analog Devices SHARC prototype. Fig.6 illustrates the encoder and decoder. Additionally, under interactive control of the user or programmed in the game or simulator, 3D sounds can be pitched, yawed, and rolled.

Fig.3 - Multi-format control room (7 of 19 speakers visible) for mixing stereo, 5.1/6.1, and HSD-3D (full sphere, with height).

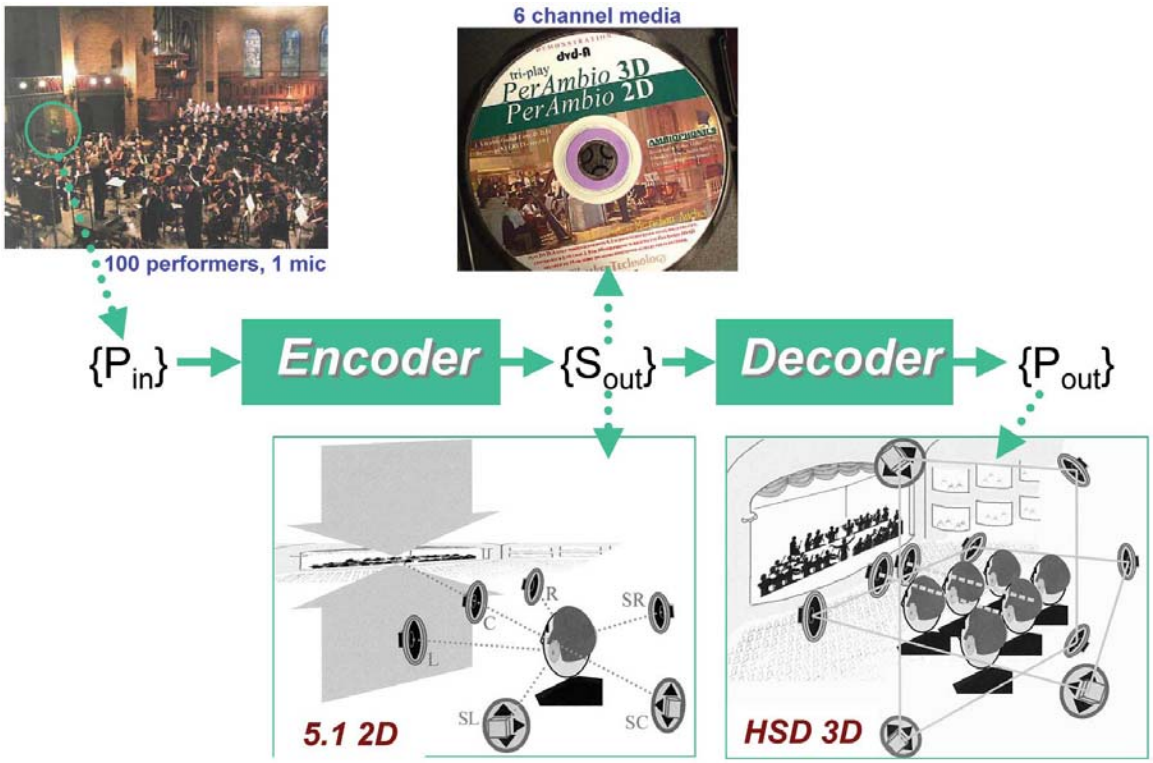


Fig.4 - High Sonic Definition (HSD, Pat. pending) 3D system. Signals from the HSD microphone are encoded to 6-channel media, playable in 5.1~7.1 without decoder. When ready for full sphere (with height) 3D, the user adds a decoder and speakers (10 total).

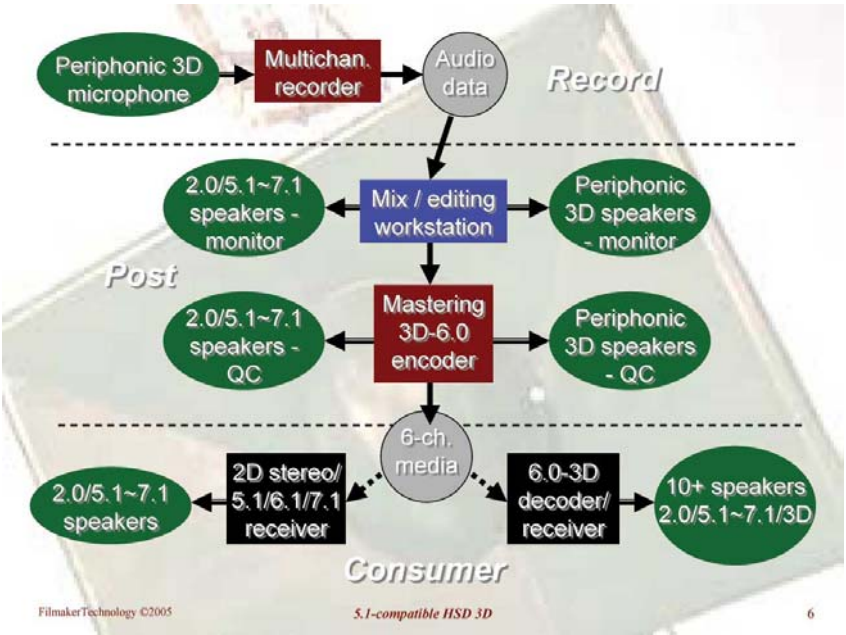


Fig.5 - Process flow for recording, post-production, and consumer replay of 5.1/stereo-compatible HSD-3D.

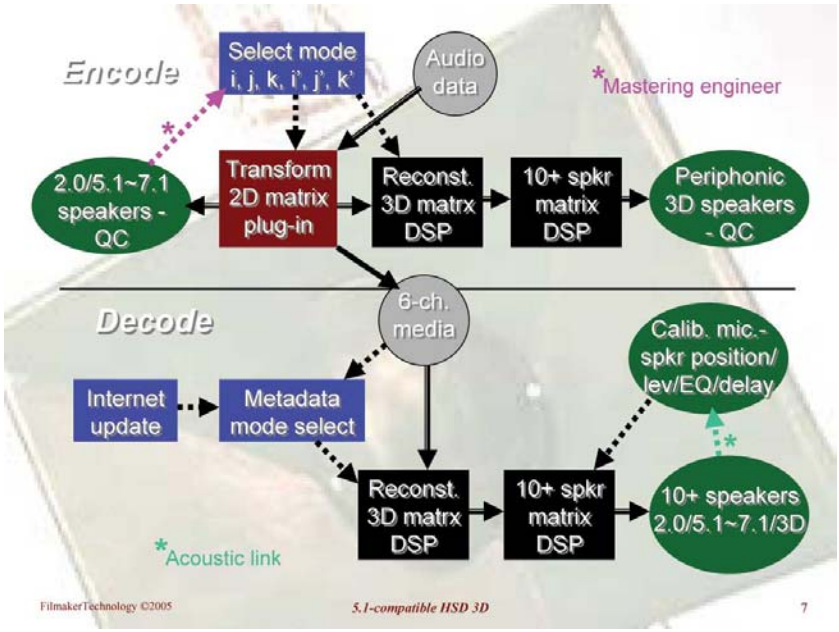


Fig.6 - software encoder (producer) and hardware DSP decoder (consumer) for 5.1/stereo-compatible HSD-3D (Pat. pending).

10. FUTURE WORK

More rigorous testing is planned, such as forced choice using voiced or other sounds that do not identify position and thereby possibly bias the results. However, the 30 subjects above indicated that most positions were clearly unambiguous, and so would probably not change appreciably even if the announcement hadn't identified direction. More involved sound effect production is also planned in order to demonstrate subjectively the advantages of both 2D surround (5.1/6.1/7.1) and 3D.

11. CONCLUSIONS

Spatiality, including localization and preservation of tone color of sound sources and the venue that encloses them, perceived by listeners throughout the full sphere (with height) of human hearing, may be more important for gaming and related training simulation, virtual reality, and theme park rides than for music and movies. In all cases, preserving arrival directionality is key not only for direct sounds, but also for early reflections (and possibly not at all for late, diffuse reverberation). Individually interpreted by listeners' unique HRTF and learned ear-brain systems, preserving directionality results in life-like tone color, suspending disbelief that it is "just a recording," but rather compellingly "real." Life-like results require preserving spatial arrival directionality at least in the horizontal plane as with 2D 5.1/6.1/7.1 reproduction, if not the full sphere 3D reproduction of natural hearing. Independent tests evaluated to what degree 30 subjects perceived positional agreement of 38 spoken announcements as the voice "moved" from the last position to the next throughout the sphere of 3D reproduction, ranking agreement a 4.39 out of 5.0 on average. These tests, along with hundreds of comparison demonstrations at FilmakerTechnology comparing 2D 5.1 with compatible HSD-3D show that, in contrast with non-spatial, monaural sources panned to multi-channel speakers, capturing and reproducing true spatiality using multi-channel surround systems has great potential for life-like gaming entertainment or training simulation content that is more compelling, and therefore more successful.

12. ACKNOWLEDGEMENTS

This work was supported by Lehigh University IBE program and Prof. John B. Ochs, and by Ralph Glasgal and the Ambiophonics Institute. Trade names are those of their owners.

13. REFERENCES

[1] Rumsey, F, "Spatial Audio," Focal Press ISBN 0-240-51623-0

[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] Holman, T, "5.1 Up and Running," Focal Press ISBN 0-240-80383-3

[5] Griesinger, D, "The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings and Their Relationship to Microphone Techniques," proceedings of AES19th Int'l Conf., June 2001, Schloss Elmau, Germany

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

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

[8] Miller, R, "Recording immersive 5.1/6.1/7.1 surround sound, compatible stereo, and future 3D (with height)," proceedings of AES 28th International Conference, Piteå, Sweden, 2006

[9] 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

[10] 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.

[11] Katz, Bob, "Mastering Audio: The Art and the Science," Focal Press ISBN 0-240-80545-3

[12] 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

[13] 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

[14] 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

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

[16] Williams, M, "Multichannel Sound Recording Practice Using Microphone Arrays," proceedings of AES 24th Int'l Conf., June 2003, Banff Canada

[17] Miller, R, "Contrasting ITU 5.1 and Pano-ambiophonic 4.1 Surround Sound Recording Using OCT and Sphere Microphones," AES 112th International Convention, May 2002, Munich Germany, preprint #5577

AUTHOR

Robin Miller, BSEE, AES, SMPTE is a musician, an orchestrator, and a filmmaker recognized by 52 awards including The Peabody. He has more than 40 years experience in music recording and mixing films and television specials. As president and CTO of FilmakerTechnology, he develops advanced entertainment technologies.