Audio Engineering Society
Convention Paper
Presented at the 112th Convention
2002 May 10–13 Munich, Germany

Contrasting ITU 5.1 and Panor-ambiophonic 4.1 Surround Sound Recording Using OCT and Sphere Microphones

Robert E. (Robin) Miller III ©2002[1]

1 FilmakerStudios, Bethlehem, Pennsylvania 18018, USA

ABSTRACT

Two multi-channel microphone techniques for natural music and sound effects reproduction are experimentally compared. Simultaneous surround sound recordings of several genres of music and ambience are made in concert hall, studio, and outdoors. Trained listeners subjectively evaluate the abilities and tradeoffs of each system to recreate accurate panoramic localization and spatial impression of opera, bluegrass with audience participating, flute quartet, brass quintet, marching bands with surrounding crowd and building echoes, and 360° "Walkabout" azimuth test. Differing speaker layouts for 5.1 and "Panor-ambiophonic" Surround are shown to satisfy two distinct listening audiences, which are further divided into home, automotive, and PC markets. An approach to recording level-setting, compatible production, and delivery formats are introduced to satisfy these diverse end uses.

STEREO IN TRANSITION

While it has brought enjoyment to many over the last half century, traditional stereo, with two speakers and a listener at points of an equilateral triangle, falls short of recreating what would have been heard during recording. Implicit problems were known in the 1930s to inventor Alan Blumlein at EMI. Perhaps because it failed to meet expectations of "images in space" analogous to stereoptic "3D" vision, stereo's market acceptance in the 1950s was driven more by ping-pong novelty and phasiness. Content so produced suffered little from manufacturers placing speakers on the sides of a short box, or from consumers placing one speaker in the living room and the other in the dining room! Surround sound offers "realism" that is more compelling of us to record and play it correctly. Why?

Two-speaker stereo suffers from imprecise localization caused by each ear hearing sound not only from its intended speaker, but also "crosstalk" sound from the speaker intended for the other ear (Figure 1). Think of watching a 3D movie without glasses - each eye sees both its image and the one intended for the other eye, destroying the illusion. Phantom images of important central soloists toggle to the nearer speaker for listeners who are off the equidistant (central) plane, and are colored by comb filtering and pinna confusion as to their frontal direction because in fact the speakers are toward the sides. Also, room ambience and other sounds that should be around and behind come instead from the frontal image plane between and beyond the two speakers, confining all reproduced sound to a 60° sector – only 1/6 of the total panorama.

To compensate for stereo's inadequate spatial impression, recordists space their microphones. Then to improve localization, they make them coincident, or amplitude-pan one microphone per instrument. So stereo "devolved" out of necessity by sound engineers monitoring over two speakers and making compromises between seemingly mutually exclusive *localization* and *spaciousness*. Engineers as well as marketers who address this "legacy" in the current evolution from two channels to five or more will more likely achieve surround's potential.

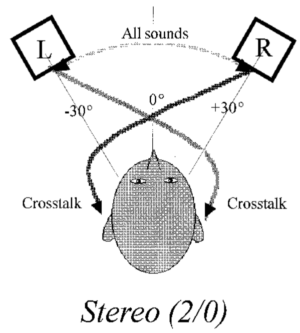


Fig. 1. Two-speaker stereo creates phantom images between the two transducers that suffer coloration of central voices and pinna confusion as to direction. All sound, including ambience, comes from the front 60°. While superior in spaciousness to monaural reproduction, stereo often falls far short of sounding "natural."

With home cinema an establishing market, ITU 775 has provided a multi-channel surround standard [1] for "5.1" speaker positions and therefore home sound receivers intended for universal replay, including music-only (Figure 2). 5.1 is positioned to replace 2.0 stereo in the home, as it has matrix surround in the cinema. Record-breaking sales of DVD players and movies with multi-channel soundtracks on DVD bode well for 5.1's popular acceptance. Broadening the reproduction soundscape to the entire 360° horizontal plane, 5.1 offers a compatible means of surround cinema and music reproduction and greatly improves spatial impression and adds envelopment, owing to two surround speakers. The author installed a "home theater" in 1999, giving his family and friends much pleasure watching movies on DVD. However for music, available in 5.1 on DTS-encoded audio CDs but mostly not recorded especially for surround (i.e. usually remixed multi-miked masters), there is potential for greater satisfaction. Investigating how much potential, how to achieve it, is it worth it, and what are the alternatives begins with a survey of 5.1's strengths and weaknesses.

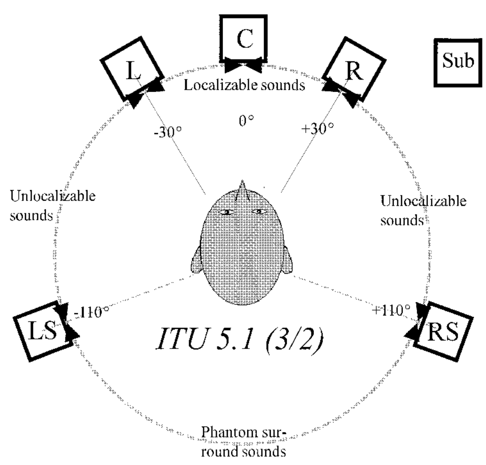


Fig. 2. ITU 5.1 (3/2) standard speaker placement creates five sets of phantom images, one between each pair of transducers, that surround the listener and makes it superior to stereo in "realism."

Derived from cinema, 5.1 localization is more precise than two-speaker stereo within the front 60° where it is best (and where trained listeners can localize sources on the order of 1°). The listening area is enlarged for sound from the center (front) speaker, benefiting, when utilized, cinema dialogue and music solos. A frontal speaker also preserves the proper tonal color compared to stereo's phantom "virtual images," especially onerous in the center due to ITD comb filtering of two identical but spaced sources and the pinna-determined source angle discrepancy. Yet while film mixers embrace this tool, some music recording engineers ignore the center channel for "artist" reasons.

For professional audio engineers who practice "surround without accompanying picture" such as music, 5.1 is an intentional compromise – but some measure of it would suggest how acceptable the compromise is. A subjective comparison with a technology that is superior in one or more ways would be useful in that determination, leading to techniques that work within its compromises. One such approach is termed *Panor-ambiophonics* (Figure 3). Using two closely-spaced speaker pairs in front and back and requiring four monaural transmission channels (two stereo), "PanAmbio" - essentially two Ambiophonic systems - is in the author's experience superior to 5.1 in accurate 360° localization, spatial impression, and envelopment with uncompromised frontal tone color (no comb filtering or pinna confusion). Bass management to redirect low frequencies from main channels, plus a ".1" LFE channel if used, are applicable to both ITU and PanAmbio, hence the 5.1 and 4.1 designations here – although these refer more precisely to the number of transmission channels, not speakers. PanAmbio's disadvantages, aside from limited popular acceptance and not being recognized by a standard, are its need for crosstalk cancellation and that it works for only one or at most two listeners, not a group. Still, regarding 5.1's qualities, PanorAmbiophonics is at least a benchmark of excellence, if not an alternative for high-quality music listening.

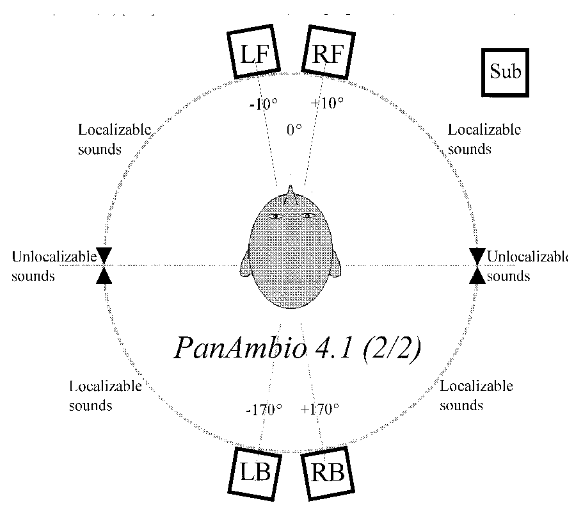


Fig. 3. PanorAmbiophonic 4.1 (2/2) speaker placement turns stereo "inside out," creating accurate images outside pairs of transducers. It can serve as a benchmark of quality for, or alternative to, ITU 5.1.

This paper describes experimental recordings to evaluate subjectively each of these reproduction systems in the light of the other. The objective will be to put each to its highest use. In addition, we will explore compatibilities for producing recordings that play well on both systems and a method of critical multi-channel recording level calibration. Applications are not limited to music only, but include music and natural sounds for film and broadcast.

PANOR-AMBIOPHONICS DESCRIBED

5.1 has been described elsewhere and is relatively well understood [2], so this paper will dwell more on Panor-ambiophonics. "PanAmbio" is two Ambiophonic systems, one for the front 180° and a second for the rear 180° as in Figure 3. (Note that Ambiophonics is not Ambisonics, a surround approach that uses coincident omni and figure-8s after Gerzon.) Each Ambiophonic system is two closely spaced speakers - an Ambiopole or stereo *dipole* - with crosstalk-cancellation provided by digital processing. Each Ambiopole more precisely reproduces recording angles up to 150° with reduced "angular distortion" [3, 4], which is characteristic of phantom images in stereo and 5.1, where instruments "relocate" toward one speaker or another when the listener is off the central plane. With one Ambiophonic dipole, instruments are localized more precisely, within ±5° where listening acoustics permit, failing due to pinna confusion (in this instance sounds intended for the sides coming from the front) only as they approach the extremes of a 180° wide stage. Contrast this single Ambiophonic system with conventional stereo, where all sounds are heard within 60°. Why and how Ambiophonics works – even for many existing stereo recordings - is the subject of Glasgal's papers available in AES publications [5] and at www.ambiophonics.org. Discussed here are its uses, limitations, recording techniques, comparison to, and compatibility with 5.1.

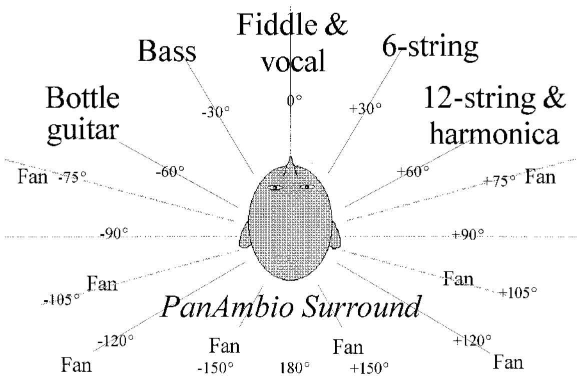


Fig. 4. PanorAmbiophonic 4.1 (2/2) reproduction localizes sources accurately within ±5°, virtually duplicating the recording session directions above. A guitar quintet and fans are placed as shown for experiments that contrasts two 360° reproduction methods. Multi-channel surround sound is more "realistic" by localizing both front stage instruments and sounds from around and behind, including antiphonal voices, audience participation, and ambience.

For surround reproduction, a second speaker dipole is added in back, and full 360° PanAmbiophonic reproduction has been demonstrated by Ralph Glasgal and the author at the 111th AES Convention, December 2001. The experimental result is precise (±5°) localization of sources around 360° (Figure 4), virtually duplicating the recording layout, although with some coloration and soft focus of voices near ±90° directly left and right. While anomalies in these two side regions are within the cone of confusion of human hearing and might be considered negligible, a PanAmbio listener is able to turn his/her head to confirm direction and tone color, just as in normal living, so the author considers these anomalies near ±90° disadvantageous. PanAmbiophony works best when reproduced in a symmetrical, "dry" (cf. recording) acoustic and with speakers at less than the critical distance (room radius) of the listening environment. With four well-matched speakers and calibrated levels, the degree of precision possible can reveal subtle errors in recording – so PanAmbio is useful also for monitoring.

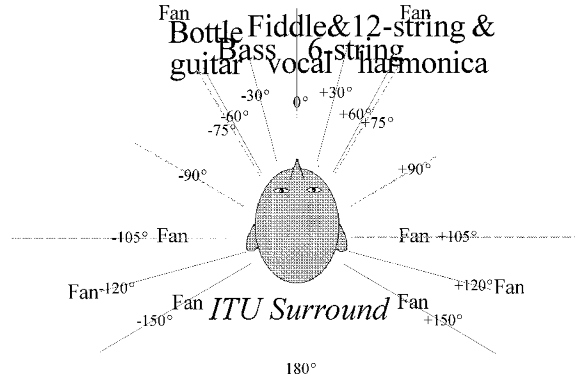


Fig. 5. Contrasted with PanAmbio, ITU 5.1 "relocates" quintet+fans by angular distortion (although less than two-speaker stereo). Original angles indicate sounds recorded at ±75° are heard at ±30° and are superimposed within the band. If precision localization is not essential to a recording, ITU 5.1 may be quite acceptable.

In contrast, the ITU 5.1 (3/2) studio and home theater speaker layout "redirects" sounds as in Figure 5, but accommodates audiences of more than two, as the "sweet area" for "important" sounds from the center speaker is increased to $1.5m^2$ [6]. Both ITU 3/2 and PanAmbio 2/2 remove the 60° confines of two-speaker stereo (Figure 6). However, PanAmbio offers a discriminating listener, or at most two, superior realism for critical music appreciation at home or automobile.

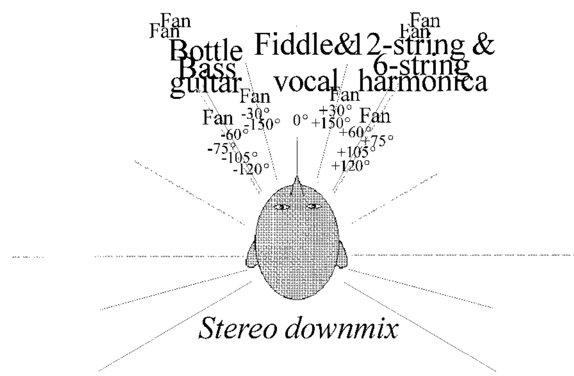


Fig. 6. Downmixing from PanAmbio or ITU 5.1 surround to two speakers by panning C to a phantom center and "folding" in back channels degrades to stereo's "hole in the middle" and all sounds in front. Multi-channel surround offers great improvement for music and digital television, as it has for the cinema.

OCT & PANOR-AMBIOPHONIC MICROPHONY

For the AES 19th International Conference on Surround Sound in Bavaria in June 2001, the author designed experiments and stereo demonstrations comparing Ambiophonic (front stage only), INA/MMA [7, 8], and OCT Optimized Cardioid Triangle [6, 9, 10, 11] of orchestra, brass quintet, and 180° "Walkabout" localization test made with an Ambiophone prototype made by the author. At the conference held at Schloss Elmau, attendees in the Ambiophonic demonstration room were able to hear recreated the nearly 120° stage width of the brass quintet session, with or without Ambisonically convolved ambience surround [12]. To demonstrate compatibility with 5.1, these recordings were also played for a large audience in the *Grosser Saal* (Great Hall - theater) using five-channel cinema speaker layout, and in autos with 5.1 and Logic Seven.

For the AES 111th Convention in New York City and this paper, the work was expanded to PanorAmbiophonic 360° reproduction with simultaneously made recordings of opera, guitar quintet with audience, marching bands, and a 360° Walkabout test [13]. Program material was chosen to represent a variety of musical genres in concert hall, studio, and outdoor acoustics. Recordings were demonstrated during Tech Tour 8 at the Ambiophonics Institute on both PanAmbio 4.0 and ITU 5.0 systems, along with a prototype automobile PanAmbio system.

Design and calibration of the microphones and their baffles are more critical for proper surround localization because the results can be discriminated. In the author's forty years of professional experience with many microphone approaches, OCT and sphere derivatives offer *both* good localization *and* spatial impression – in the past more typically an either/or choice – plus the envelopment of surround sound. For ITU 5.1, OCT uses the directional characteristics of cardioid and supercardioid microphones to image front stage sources with unambiguous phantoms among three loudspeakers (Figures 7a, b).

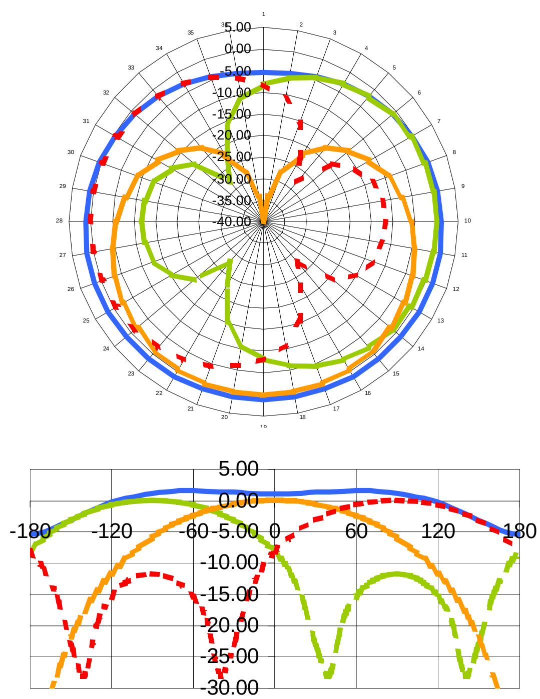


Fig. 7. Contribution of idealized OCT microphones to total energy – a) polar plot every 10° (front at bottom); b) rectangular plot.

If more than the 6dB rejection of back sounds is needed, a baffle can add 7dB. Hall sound is added to LS, RS using back-facing cardioids, or to L, R, LS, & RS using a surround reverb convolver or four-channel room microphone such as side-facing figure-8s [6, 14].

As the basis for Ambiophonic (front only) recording, where ambience is convolved from hall impulse responses for two-channel recordings, the sphere microphone [15] - a frequency-dependent analog of the human head without pinna – is shown idealized for mid-frequencies (Figures 8a, b). When baffled, its stereo characteristics ideally show nearly 10dB of back rejection (Figures 9a, b). The author's prototype *Ambiophone*, measured every 15° with filtered pink noise in a non-anechoic studio, approaches these ideal characteristics (Figures 10a, b).

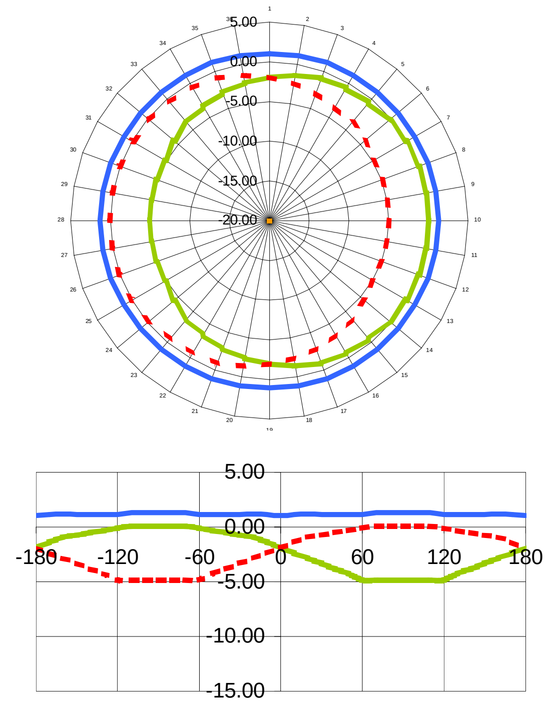


Fig. 8. Contribution of idealized sphere microphone to total energy at 1kHz (frequency dependent) – a) polar and b) rectangular plots.

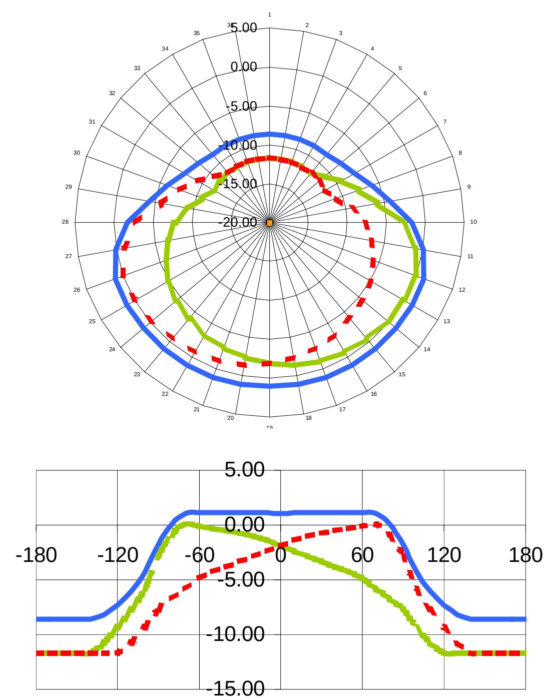


Fig. 9. Idealized Ambiophone front sphere at 1kHz – a) polar response every 10° (front at bottom); b) rectangular response.

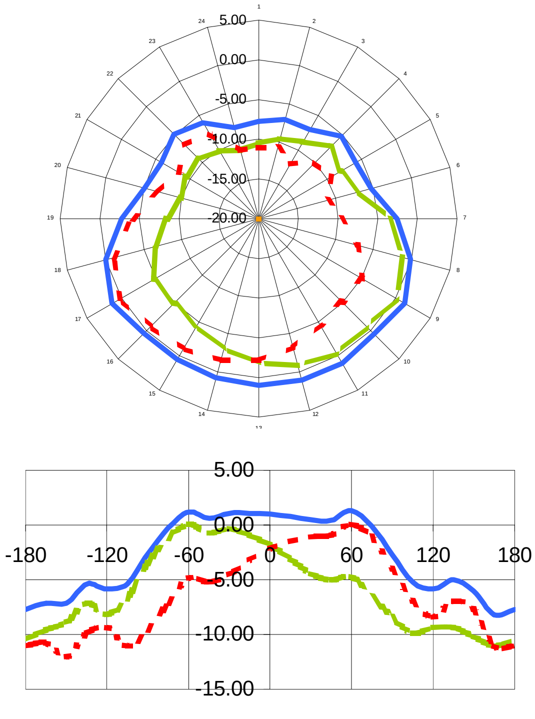


Fig. 10. Measured response, front PanAmbio prototype microphone in non-anechoic studio – a) polar plot every 15°; b) rectangular plot. Cf. previous figure. Back-angle irregularities ±90° to ±180° will be masked by on-axis response of back sphere in next figure.

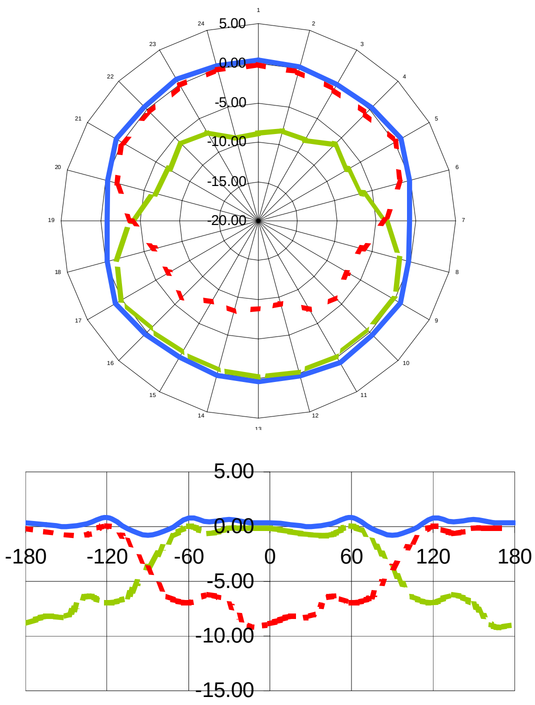


Fig. 11. Measured total response, prototype PanAmbio microphone, front + back spheres, is within ±1dB around the entire 360° – a) polar response every 15° (front at bottom); b) rectangular response.

Combining two such spheres for PanAmbio reproduces the full 360° horizontal plane within ±1dB (Figures 11a, b). For the reader's experimental verification, simultaneous recordings using both OCT and the author's prototype *PanAmbiophone* are available in evaluation DTS-encoded CDs, described later, along with an early consensus of subjective opinions of each.

For fair comparison, 5.1 and PanAmbio recordings were made simultaneously, both in the concert hall and the studio, using best practices in the experience of the author: OCT and dual Ambiophone (essentially two sphere microphones with acoustic baffle) comprised of small diaphragm condenser microphones (Figure 12). For ITU 5.1, the OCT array consisted of five microphones: a cardioid and two supercardioids optimized for off-axis pickup mixed with omnis to support bass reproduction. For the opera, a spot microphone was mixed according to the Room Related Balancing technique [6]. Figures 13, 14, 15 show the main array and its placement.

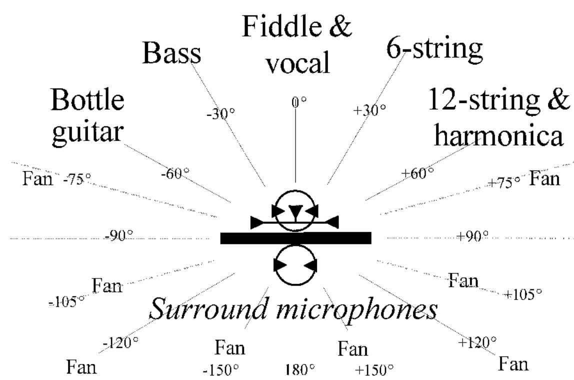


Fig. 12. Microphone arrays contrast two 360° reproduction methods. PanorAmbiophony uses twin spheres with baffle. OCT uses two supercardioids facing ±90° and cardioid facing front. Simultaneous recordings of guitar quintet + fans, opera, brass quintet, string quartet, marching bands, and "Perambiolating 360°" azimuth test were authored to companion DTS-encoded CDs for evaluation [13].

Fig. 13. Guitar quintet in studio for comparison 360° recordings. OCT and prototype Ambiophone are at right. Instrumentalists are at 0°, ±30°, ±60° and fans (not shown) at ±75°, ±105°, ±120°, ±150°.

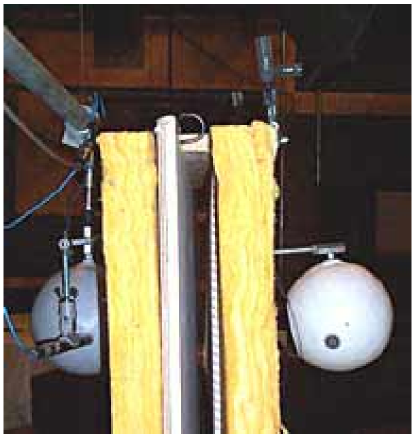


Fig. 14. OCT atop prototype "PanorAmbiophone" – twin sphere microphones separated by baffle. In the studio, the rear sphere also served as room microphone for OCT.

Fig. 15. Hoisting OCT and prototype Ambiophone microphones in the 1,000 seat opera house. Microphones are Schoeps CCM-series.

5.1 and PanAmbio mixes were made of all six recordings and encoded on DTS audio CDs for convenient replay for demonstrations and future listener tests. For music, no equalization, effects, or dynamic compression was used. In informal listening sessions, independent recording engineers and musicians involved in the recordings reported generally that, with both reproduction systems, the recordings were among the most realistic they had heard, and that in particular the localization of PanAmbio was the most accurate. We hope to verify these conclusions in future formal listening tests using trained auditioners [16]. Observing the highly accurate indication of positioning of instruments and vicarious enjoyment of the "live" performance by these critical listeners, the author feels it is safe to claim that, using these techniques, both ITU 5.1 and PanAmbio 4.1 are significantly more satisfying than conventional stereo in the realism and natural reproduction of music.

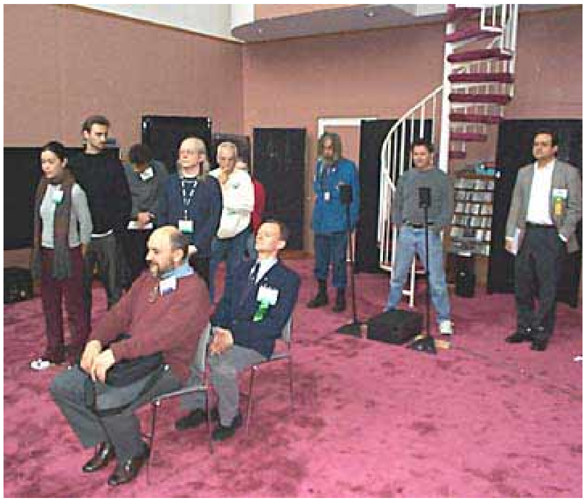


Fig. 16. Two (seated) AES 111th Conv. attendees hear PanAmbio surround at the Ambiophonics Institute. The back speaker-pair is silhouetted in front of two gentlemen in back.

SURROUND RECORDINGS FOR EVALUATION

For AES 111th, December, 2001 tour of the Ambiophonics Institute (Figure 16), the author prepared two DTS-encoded audio CDs titled *PerAmbiolating 360°* (pun intended), one in ITU 5.0 and a companion in PanorAmbiophonic 4.0 [13]. A ".1" LFE channel was considered unnecessary for musical demonstrations. Recorded in April, September, and October 2001, artists and venues were Lehigh University Opera at Zoellner Center for the Arts, and Martin Guitar Quintet, Satori Flute Quartet, & Mainstreet Brass at FilmakerStudios, Bethlehem PA, USA. Selection numbers in parenthesis ( ) below indicate pre-crosstalk-cancelled versions on the PanorAmbiophonic disc, so no special hardware is needed for evaluation – just temporarily moving four speakers (C unused) of a 5.1 layout. Except *Parade*, comparison PanAmbio and OCT 5.0 recordings were made simultaneously with OCT and Ambiophone microphones described earlier, with source locations and description of audible effect upon replay as follows:

1 (&7) Barber of Seville *Sitzprobe* - 1:58

Recording Angle 120° front, hall back

The first rehearsal with soloists, chorus, and orchestra of a mixed professional/student production. Hall is 9,200 m³ with RT=2.1s and 3.77m (calculated) room radius. Room microphones are side-facing figure-8s back 10m (no delay). A spot microphone for soloists is mixed according to Room-Related Balancing.

In the benchmark PanAmbio 2/2 playback, individual instruments and voices are distinctly localizable and widely spread, nearly equal to the 120° recording angle. The spatial impression is "natural-sounding" with front and rear stage seamlessly integrated, but dependant upon listener taste for the relative back level. In contrast, the ITU 3/2 playback over five identical speakers - 2-way with 10in (250mm) woofer – exhibits "commercially acceptable" (some listeners claimed "the best they'd heard") spatial impression and envelopment with plausible localization, albeit across a compressed front stage, 60° L-to-R, but over a much larger and stable listening area than either PanAmbio or two-speaker stereo.

2 (&8) Lunchbreak at Martin Guitar Blues - 1:59

Quintet 0°, ±30°, ±60°, fans sides & back

Simulating a jazz club (or "unplugged" telecast) with bluegrass quintet and audience, the studio is 500m³ with modal profile shown in Figure 17, RT=0.31s (controllable, chosen to mimic a performance space) and with players in a 120° arc of approx. the measured 3.2m critical distance (room radius). Instruments from left to right are bottle (slide) guitar, acoustic bass guitar, fiddle & vocal, 6-string rhythm guitar, and 12-string guitar & harmonica. Eight fans, positioned as shown in Figure 4 hoot, clap, and clink glasses.

The benchmark PanAmbio 2/2 playback has the effect, astonishing at first, of replacing the listening environment with the recording environment, achieving a remarkably natural "you are there" result – see Figure 4. In ITU 3/2 playback, the listener is enveloped in a quite plausible club atmosphere, notwithstanding the less precise localization, as shown in Figure 5.

3 (&9) Mozart Wrap-a-Rondo in F - 1:42

Flute quartet ±20°, ±60°, room back

A chamber quartet in the 500m³ studio with modal profile shown in Figure 17, RT=0.31s (controllable, chosen to mimic a recital hall) and with players in a 120° arc the measured 3.2m critical distance (room radius) - from left: violin, viola, cello, and flute.

The benchmark PanAmbio 2/2 playback is a bit unsatisfying in its unequal representation of directional (string) and omni-directional (flute) in the live studio, possibly because the system's capability has created higher expectations. In contrast, the ITU 3/2 seems more acceptable in this regard, although the author feels that, in a commercial recording situation, a retake should be indicated with adjustments to acoustics and positioning. It is included on the evaluation CDs to study these error conditions.

4 (&10) Sousa's Fairest Brass - 2:37

Brass quintet 0°, ±30°, ±60°, room back

Recorded April, 2001, for AES 19th International Surround Conference, June, 2001, in the 500m³ studio with modal profile shown in Figure 17, RT=0.31s (controllable, chosen to mimic concert stage-house) and with players in a 120° arc of approx. the measured 3.2m critical distance (room radius) but with ORTF room microphone. Instruments from left: 1st Trumpet, French horn, Tuba, Trombone, and 2nd Trumpet.

In benchmark PanAmbio 2/2 replay, the more directional instruments are slightly narrower than their recorded positions across the total 120° stage due to an earlier prototype Ambiophone (larger diameter sphere). The rearward-speaking French horn, as might be expected, is only vaguely correct. In contrast, the ITU 3/2 replay is "commercially present," although images are confined to the 60° front L/C/R speakers. Both envelop the listener with room ambience.

5 (&11) SPL Setup & *PerAmbiolating* 360° - 4:36

Voice ea 15°; quartet ±45°, ±135°

The "Walkabout" was recorded in the 500m³ studio with modal profile shown in Figure 17, RT=0.31s and with the announcer *perambiolating* (pun intended) the twin baffled sphere microphone array at a radius of 2.5m. To parallel real-world conditions and the recordings above, studio acoustics were adjusted to replicate the stage house of a concert hall, with early reflections <15ms limited to those from horizontal planes (the floor), so their virtual "images" arrive at the same horizontal angle as their direct sound [17].

The benchmark PanAmbio 2/2 replay localizes announcements to the nearest 5° around all 360° with some "fuzziness" near 90° on each side. Accompanying bursts of filtered pink noise are more difficult to locate, but provide data for Figures 10a, b and 11a, b. In contrast, the ITU 3/2 replay exhibits maximum error of 45° (75° each side is solidly reproduced by a speaker at 30°) as is illustrated in Figures 18 & 19a, b. In both systems, the quartet, now surrounding the array at the corners of a square, are difficult to localize for reasons postulated above.

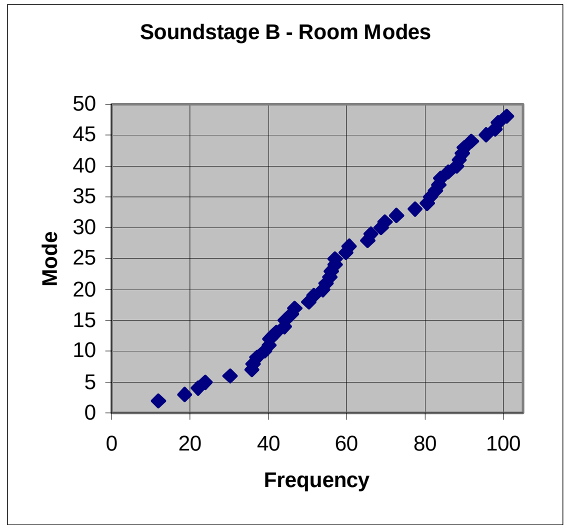


Fig. 17. Lowest 50 eigentone modes of studio where experimental recordings were made. RT=0.31s (controllable).

6 (&12) Marching Bands on Parade - 3:40

Subject Angle 180°; recreated surround

Unlike others above, this excerpt illustrates "up-producing" surround from a 2-channel stereo field recording using editing and mixing of original and additional processed tracks such as for film mixing. For ITU 3/2, L/C/R is derived after Gerzon [18]. To evaluate creative potential in post-production, surround is six effects tracks derived from the original stereo, edited and processed to simulate crowd and building echoes. The illusion has been successful with all trained listeners to date.

In benchmark PanAmbio 2/2, the result is plausible envelopment of a listener standing on the sidewalk while bands march by in the street, beginning extreme right and continuing to extreme left, with cheering and building echoes around and behind. Groups of instruments are heard to move smoothly (no perceptible angular distortion) across right-of-center through center to left-of-center to a degree of realism that the listener can readily imagine it. In contrast, ITU 3/2 replay of course is confined to the 60° triangle, but creates a satisfactory illusion nonetheless. In further contrast to traditional two-speaker stereo replay, the ITU 3/2 result exhibits less angular distortion, with no perceptible "hole in the middle."

SURROUND RECORDING METHODOLOGY

While these demonstrations involve acoustic sources, principles apply to ambient popular instrumentation. Expect them to lack the resources, retakes, and approval layers of a commercial release. Risks were taken for purposes of discovery and testing limits of techniques in order to serve artistic purposes to follow. Even high-end reproduction systems will be tested by the tracks' raw dynamic range; there is significant content at 15Hz in the opera track by an enthusiastic student bass drum player and believer in subwoofers!

Except *Parade*, recordings were made with no level compression, effects, or equalization (except filtration for OCT lows). No panned mono spots except opera principals (Room-Related Balancing using time delays). The sole exception, "Parade of Marching Bands," is a single stereo Ambiophone synthesized to PanAmbio and ITU 5.0 surround in post-production. Intended to test "up-mixing" from a simple two-channel field recording, the synthesis comprises six stereo tracks of crowd loop, spot crowd FX, and delayed & low EQ'd "building echoes" to create the surround

. Replay of the ITU 5.1 evaluation disc requires a DVD player and 5-channel receiver capable of decoding DTS to five speakers in the standard ITU-R BS.775 layout. Replay of the PanorAmbiophonic disc requires two pairs of closely spaced speakers at ±10° and ±170° and crosstalk cancellation using DSP, mechanical barriers, or pre-crosstalk-cancelled cuts 7~12 on the PanAmbio evaluation CD, as summarized in Appendix A.

Recording Level Calibration

For any multi-channel production, recording levels are critical and must be maintained – or their changes precisely controlled - in post-production and distribution. In essence, to preserve localization, the record-to-reproduce chain must exhibit constant relative channel levels from instruments to ears. Microphones vary in sensitivity even within the same model, and preamplifiers often have uncalibrated variable gain. Once analog signals reach studio level, usually +4dBu (ref .775 vrms) - or –15 to –20 dB FS depending upon the digital standard chosen - levels can be preserved by good practice. The comparison recordings above relied on the technique in Appendix B to calibrate multi-channel recording, beginning with acoustic source levels.

Once tracks are recorded, levels can be preserved or varied in post-production according to artistic choices. For OCT where omnis have been recorded for bass compensation of the supercardioids, the author has found in-band gain identity is a good starting point (used in the experiments above) after low-pass filtration at 100Hz. Similarly, room array contribution can begin at identity gain with front channels and then varied to taste. Note that spot microphones mixed using Room Related Balancing often contribute sufficiently several dB below identity with front channels (as demonstrated in the opera experimental recording).

Compatible Surround Production

The market for ITU 5.1 surround music seems assured as of this writing, yet enthusiasts exist for whom "compromise" is not to be heard, literally. This niche is typically high-end and would likely be interested in PanAmbio as an alternative for personal listening.

Costs for surround production are higher than for less complex stereo, and would be higher yet providing for two surround formats. Finite dollars, bandwidth, and numbers of channels both for recording and distribution have already led the author to dual-purpose approaches with managed compromises. For AES 19th in Bavaria, June 2001, the author's experimental CD *Ambiophonic Surround Sound Demonstration* contained front-only Ambio derived from INA/MMA and OCT as well as the Ambiophone sphere. For the ITU 5.1 DTS audio CD, LS and RS channels in the studio were derived from the PanAmbio back sphere. Conversely, several room mic configurations yielded useful PanAmbio LB, RB (similar to the way Ambiophony works with many existing stereo recordings). For compatibility with two-channel stereo, "Parade" derives both 5.1 and PanAmbio from a two-channel field recording, typical of film location effects. PanAmbio pairs "fold" front to perfect stereo, equivalent to a single unbaffled sphere.

New experiments have developed during post-production an acceptable PanAmbio front stage from OCT, a 5.1 front stage from the PanAmbio front sphere, and surround for 5.1 or PanAmbio by hall convolution. More work is planned to distill these combinations to straight-forward procedures.

Compatible Surround Delivery

To facilitate both ITU and PanAmbio listening requires moving four speakers (or switching nine) – the home theater's subwoofer, receiver/decoder, and universal DVD player are the same (C unused for PanAmbio). Distribution formats for PanAmbio can be the same DTS-CD, DVD-V, DVD-A, SACD, or multi-channel broadcast using AC-3 (Dolby Digital) of the DTV standard. PanAmbio is unsuitable for large audiences such as the cinema. Requirements for PanAmbio replay are in Appendix A.

Listening Environment

For any surround approach, the listening environment is critical if precise localization is expected – especially for PanAmbio because its subtle capabilities can be more easily destroyed. Informal evaluations below were made in a control room, a home theater, a large and live demonstration room, and an automobile – with better results obtained in better acoustics. Generally, the listening room should be symmetrical or acoustically treated and "drier" than the recording venue. See Appendix A.

PanAmbio as a tool for better ITU 5.1

Ultimately, post-production decisions and approval of a final product must be made while monitoring in the delivery format. However during recording of a 5.1 production, monitoring using Ambiophonic (front only) or PanAmbio surround techniques has demonstrable operational advantages. As a location recording engineer and television "A-1," the author uses a custom portable Ambiophonic monitoring system. Its compactness suits off-base production trucks. Its "you are there" capability transforms the usually tiny, acoustically alien audio booth into the performance space. Subtle panning adjustments of spot microphones, panning errors with respect to main microphones, and phase errors are revealed and can be dealt with quickly in the heat of the session or live telecast. On replay, musical directors can discern individual voices and whether a "natural" blend and impression of hall ambience has been captured. Especially for music recording, there is often less to "fix" in post, lowering costs.

EVALUATION BY TRAINED AUDITIONERS

Only the three engineers present at the recording sessions could compare the results with the live events. The need for evaluation of recordings by trained listeners (cf. students often recruited for listening tests [16]) was observed during AES 111th Convention demonstrations, where trained and untrained attendees simultaneously perceived far different source directions.

Initial results show data from several trained auditioners, with the moderator using the form in Figure 18. Future work will require double blind analysis – the evaluation CDs divulge the angle so they can stand alone – and use a statistically larger group [19].

With ITU 5.1, trained listeners to date report critical front stage localization is compressed angularly in half, but with less angular distortion than two-speaker stereo, including less "hole in the middle," and less nearer-speaker toggling anomaly for listeners off center as shown in Figure 19a, b. When possible, sources might be positioned during recording to compensate for any objectionable "relocation." In contrast, PanAmbio reproduces original directions nearly linearly (to the nearest 5°).

***PerAmbiolating 360°* Test CDs - "Walkabout"** (to be filled out by 3rd party)** R Miller 2/02

After calibrating speakers for equal SPL, enter the approx. (within ±5°) angle of direction you hear voiced the angle at left (CCW from front). If unsure of direction, rotate head to verify. Indicate using ITALICS if direction remains ambiguous and BOLD if obviously from a speaker.

***ITU 5.0 (3/2)** - five speakers at 0°, ±30°, ±110°*

CD voiced°	RMiller	BFox	RGlasgal	You
Center 0°	**0**	**0**	0	
Left 15°	5	5	15	
30°	15	15	30	
45°	25	25	30	
60°	30	30	40	
75°	**30**	**30**	60	
90°	*60*	*60*	60	
105°	*90*	*90*	90	
120°	*105*	*105*	90	
135°	**110**	**110**	110	
150°	*120*	*120*	110	
165°	165	165	110	
Back 180°	180	180	180	
Right 165°	165	165	165	
150°	*120*	*120*	150	
135°	**110**	**110**	110	
120°	*105*	*105*	110	
105°	*90*	*90*	90	
90°	*60*	*60*	30	
75°	**30**	**30**	30	
60°	30	30	30	
45°	25	25	30	
30°	15	15	25	
15°	5	5	15	

**DTS-encoded CDs for evaluation available www.filmaker.com

***PanAmbio 4.0 (2/2)** - four speakers at ±10°, ±170°*

CD voiced°	RMiller*	BFox*	RGlasgal	You
Center 0°	0	0	0	
Left 15°	15	15	15	
30°	30	30	30	
45°	45	45	45	
60°	60	60	50	
75°	*75*	*75*	75	
90°	*90*	*90*	85	
105°	*105*	*105*	105	
120°	120	120	120	
135°	135	135	135	
150°	150	150	150	
165°	165	165	170	
Back 180°	180	180	180	
Right 165°	165	165	165	
150°	150	150	150	
135°	135	135	135	
120°	120	120	110	
105°	*105*	*105*	90	
90°	*90*	*90*	80	
75°	*75*	*75*	75	
60°	60	60	60	
45°	45	45	45	
30°	30	30	30	
15°	15	15	15	

* Using pre-crosstalk-cancelled tracks

Fig. 18. Form for moderators to report where to the nearest 5° trained listeners perceive voiced angles on the ITU 3/2 and PanAmbio 2/2 evaluation DTS-CDs [13]. Listening environment reverberation must be less than recording studio (RT=0.31s). More formal listening tests are planned.

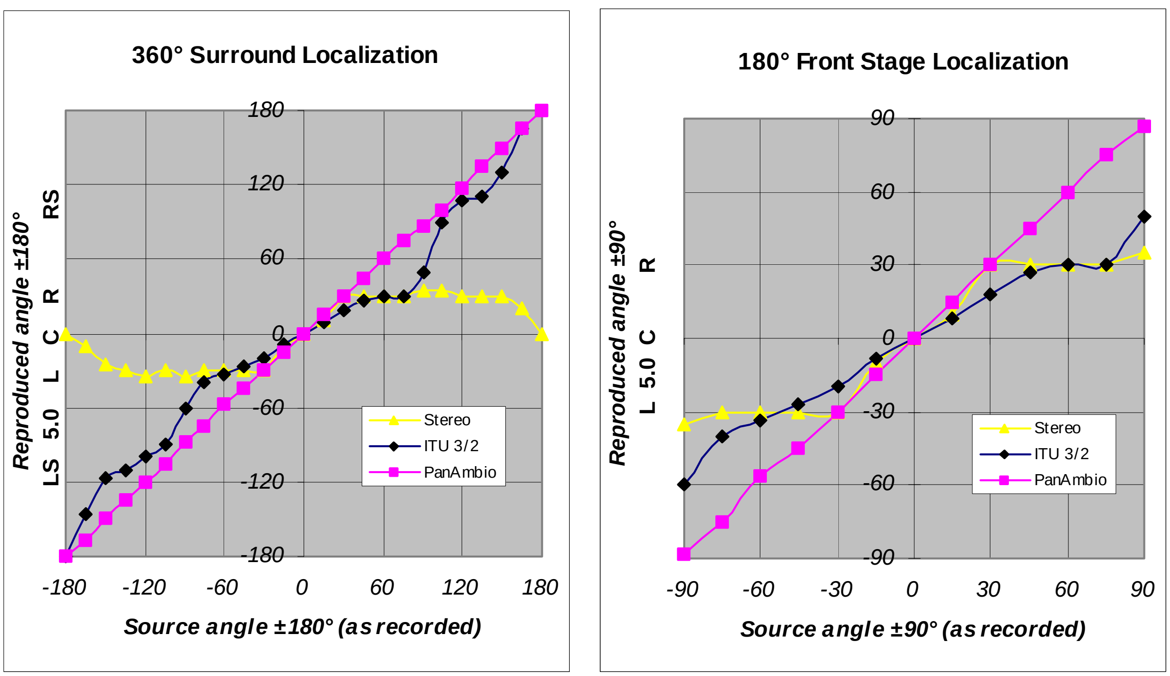


Fig. 19. Perceived localization around a) entire 360° horizontal plane and b) 180° front stage - ITU 3/2 vs. PanAmbio. Listeners reported to the nearest 5° that ITU 3/2 is "ambiguous" at ±90°, ±105°, ±120°, and ±150°. PanAmbio approaches the ideal straight line but is "fuzzy" nearing ±90°.

Note that to different degrees, both ITU and PanAmbio lack focus at the sides, within the "cone of confusion" of human hearing if the listener does not rotate his/her head. 5.1 exhibits spectral "tearing" [2] for phantoms in two 80° sectors between L & LS and between R & RS due to the HRTF of human hearing (rotating the head to ±70° restores side phantoms). Localization is reported "ambiguous" at ±90°, ±105°, ±120°, and ±150°. No front-back confusion was reported for 180° unless the listener is off center, in which case all back phantoms toggle to the nearer of LS or RS, as in two-speaker stereo – the situation addressed by the center-back channel/speaker of 6.1. Recordists should exercise care placing critical voices in these sectors.

In contrast, PanAmbio suffers ambiguity, coloration, and pinna confusion within two 30° sectors at ±90° (rotating helps confirm direction, but translating off center inhibits crosstalk cancellation). The consensus is that, toward the goal of a natural illusion of spaciousness, envelopment, *and* localization, PanAmbio is superior for critical personal listening, but ITU 5.1 is the choice for a group.

AUTO SURROUND SOUND EXPERIMENT

Today there are more high-end choices for car audio systems and a healthy aftermarket for reinstallations. Consider the traveling representative or cross-country trucker, mostly driving alone, listening to music in a fixed listening position. Consider the luxury car driver who wants the comforts of home in the cab, e.g. THX 7.1 as first certified in a Lincoln. These situations suggest a great potential market for automotive surround.

Initial models have been delivered with ITU 5.1 and 7.1 systems. These find the driver and front passenger outside the sweet area. A fifth passenger seated in the rear seat center has the best position. In addition, the small space, reflective side windows, and short distances and delay times of auto psychoacoustics significantly color the reproduced sound.

The car PanAmbio situation is promising but certainly not simpler: one or two Ambiopole pairs would be needed for each seat, so crosstalk between systems becomes a factor, not to mention issues of uncompromisingly mounting speakers behind the steering wheel, the backs of seats, etc. Electronics would need real time DSP and crosstalk cancellation algorithms unique to these speakers and this passenger compartment. For AES 111th, 2001, the author demonstrated a prototype with eight speakers - two complete PanorAmbiophonic systems for two passengers - with modest success (Figure 20). To reduce system-to-system crosstalk, speakers were placed close (0.5m) to the listeners. However, absorptive material was needed on side windows to avoid interference with the cross-talk cancellation mechanism. Bass needed augmentation due to 4in (100mm) diameter "woofers."

Still, attendees seemed impressed, possibly because the envelopment, spatial impression, and precise 360° localization of the "Walkabout" recording and width of marching bands exceeded their expectations for car sound. The systems have not been tested in motion, where road noise would be a further detriment. More work is needed in automotive surround, such as refined impulse responses and DSP, possibly unique for each listener position. However, amusement ride sound presents a workable situation for good PanAmbio if conveyances are open and carry a single rider per system.

Fig. 20. Two passengers, each with four speakers for PanAmbio surround, demonstrated in author's van at AES 111th Conv., 2001.

PC, small TV, and Game Sound

PCs including laptops, small TV sets, and video games with speakers on either side of the monitor are well suited to Ambiophonic (front only) reproduction, requiring only crosstalk cancellation software or hardware, reduced in the future from a prototype PCB (Figure 21) to a single chip. The listener is fixed in the prerequisite position, Ambio's only ergonomic caveat. Ideally, crosstalk cancellation would be done in real time by the PC or DSP based on the binaural impulse response of the exact model speakers. A simpler albeit less exact delivery method is to produce pre-crosstalk cancelled generic versions of music or sound effects for games or PCs. Examples for immediate demonstration are in streaming audio form at www.filmaker.com.

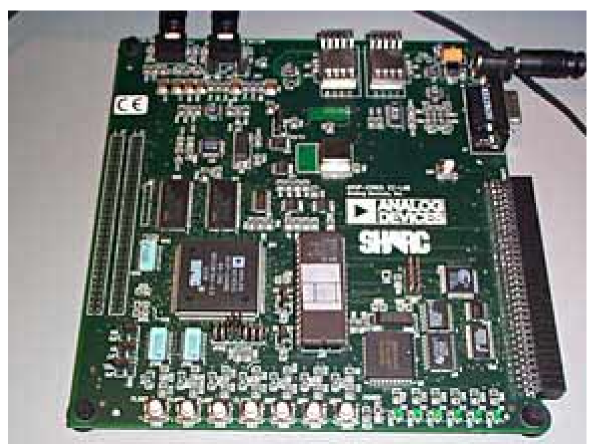


Fig. 21. Prototype real-time DSP crosstalk-canceller for one Ambiopole. The two needed for PanAmbio might be a single chip.

FUTURE WORK

The Panambiophone will evolve with the next prototype, which more closely integrates spheres and baffles. A challenge is its aesthetics for use in the presence of an audience or television cameras. While the evaluation CDs include five genres of music of wide variety, many others exist which may respond more or less well to the surround interpretations. Future sessions will include organ, choir, big band, and sound effects. Auto and amusement ride applications will be further developed, as will exploring project-specific repurposing of unused C, LFE, and CS of 6.1 formats, possibly for height if not additional surround speakers. Double-blind analysis by trained listeners has already been mentioned. Finally, the first production of an album in DVD-A, SACD, DTS-CD, or DVD-V and DTV broadcast, possibly in compatible 5.0/4.0, will be proposed to a record company or broadcaster.

CONCLUSIONS

Both 5.1 and PanAmbio reproduction systems satisfy critical listeners significantly more than conventional two-channel stereo, which has been the underlying technology for vinyl, FM, and audiocassette music reproduction for five decades, film for two, and broadcast television for one. Standardized by ITU-R 775, 5.1 is positioned to replace stereo for home audio entertainment, just as it has matrix surround for film and broadcast. Sales of DVD players and home receivers prove 5.1's popular market acceptance, even if for critical music-only reproduction it is an intentional compromise. Panor-ambiophonic 4.0 (2/2) reproduction is superior to ITU 5.1 in localization precision and therefore in directionally dependent spatial impression and envelopment for one or at most two critically positioned listeners who prefer less compromised music-only reproduction. Experiments demonstrate recording level setting and approaches for compatible production of both 5.1 and PanAmbio and for distribution using DTS-CD, DVD-A, or SACD (C unused or available for repurposing). Alternatively, Ambiophonic (front only) reproduction offers wide, precise localization with just two distribution channels – embracing many if not most existing stereo recordings - on common PC speakers, small TV sets, games, and high-end audiophile systems that add ambience by convolution. Automotive and amusement ride PanAmbio surround has been demonstrated for one or two seats, however attention must be paid to both listening acoustics and DSP issues.

ACKNOWLEDGEMENTS

The author is grateful to Ralph Glasgal for his inspiration and wisdom and to the Ambiophonics Institute for support of this work. Thanks to Jörg Wuttke of Schoeps GMBH, Buzz Turner of Redding Audio, and Jerry Bruck of Posthorn Recordings for advice and equipment, Geoff Houser and Brad Frikkers for help fabricating prototypes, Howard Moscovitz for help recording, participating musicians, and David Del Grosso and Jeff Levison of DTS for encoding the CDs. In the author's mind, acknowledgements must include teachers past and mentors present, including all those cited in the references, especially my colleague Angelo Farina. Trademarks are those of their owners, including DTS of Digital Theater Systems, Dolby Digital of Dolby Laboratories, SACD of Philips and Sony, Logic Seven of Lexicon, THX of Lucasfilm.

APPENDIX A – PANOR-AMBIOPHONIC REQUIREMENTS

To fully realize PanAmbio playback requires stringent acoustics – refer to AES Technical Document "Multichannel surround sound systems and operations" [20]. Having both ITU and PanAmbio listening requires moving four speakers (or switching nine) – the studio's bass-manager or home theater's subwoofer, receiver/decoder, and universal DVD player can be the same (C unused for PanAmbio). Distribution formats can be DTS-CD, DVD-V, DVD-A, or SACD - the market will resolve which among them succeed - or multi-channel broadcast using AC-3 (Dolby Digital) of the DTV standard. Requirements for PanAmbio playback are:

- Symmetrical or acoustically treated room and layout with RT less than recording venue and with one or at most two listeners seated at the center of speaker pairs directly front and back at a radius less than the acoustic room radius.
- Universal DVD/CD player or DTV receiver with 4-channel output (2 AES/EBU or one multi-channel encoded serial digital connection, coax or optical e.g. S/PDIF – C unused);
- Decoder (digital stream to multi-channel analog) – C unused – or integrated in 5.1 home receiver. Alternative is all-digital integration of crosstalk-canceller, bass manager, and power amplifiers below;
- Crosstalk-cancellers – two DSPs, ideally based on impulse response of speakers used, currently in prototype form (see Figure 21) – evaluation CD [13] has tracks pre-cancelled for "generic" speakers;
- Amplifier/bass manager typical of available 5.1 studio units and home receivers and accommodating multi-channel inputs (S/PDIF coax or optical) and four speaker outputs (C unused) plus subwoofer, if any;
- 4 main speakers with vertically in-line components, full range or common-woofer/satellite systems. Dual PanAmbio and ITU 5.1 requires moving four speakers (or switching nine). Subwoofer optional if main speakers are full range and producers decide the ".1" LFE channel is not needed for music;
- Calibration of channels relative of one another within ½dB at the listening position using SPL meter and filtered pink noise from test CD.

An alternative to PanAmbio is Ambio (front-only, 120~150° wide reproduced stage) that uses one crosstalk-cancellation DSP, one main speaker pair, and any number of surround speakers that are fed ambience convolved from impulse responses of any desired hall for many existing stereo recordings. In this regard, note that the anomalies of two-speaker stereo are caused by the triangular speaker placement with respect to the listener – they are not necessarily intrinsic to the recordings, whether made with widely spaced, closely spaced, coincident, or panned microphones.

APPENDIX B – RECORDING CALIBRATION PROCEDURE

The method adds to either location or studio recording kit a portable (preferably battery operated) amplifier-speaker, source of filtered pink noise, and SPL meter. It is necessary to know from experience or to measure with the SPL meter the peak sound pressure level of the performance to be recorded. Then the recordist may adapt the following generalized procedure for his/her specific equipment: Beginning at a determined distance – less than the room's critical radius, e.g. <3m - from an OCT array, position the noise source front and center and adjust its output for some Sound Pressure Level, e.g. 72dB, measured at the microphone array. Adjust preamp C gain for a digital record level 3dB (for overload safety) lower than the point below Full Scale by the amount above 72SPL that you anticipate (or have measured) the ensemble to peak, e.g. –35FS for a small orchestra that peaks at 104SPL. Move the noise source to 3m directly left of the array and set preamp L, then 3m directly right and set preamp R, then 3m from and on common axis with any room microphone array and set preamps LS & RS. Spot microphones can be similarly calibrated at a lower FS level to allow for their closer positioning. For a PanAmbiophone, only two positions, directly front and back, are needed to set preamps LF, RF and LB, RB respectively. Change levels only in all channels equally using preamps with calibrated gain adjustment, or their attenuation controlled by ganged VCAs in the case of analog consoles, or by digital controls of digital consoles. Precision of a fraction of a dB is required to preserve localization.

REFERENCES

[1] ITU-R BS.775.1: "Multichannel Stereophonic Sound System

With and Without Accompanying Picture" (Geneva, 1992-4).

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

[3] M. Williams, "Unified Theory of Microphone Systems for Stereophonic Sound Recording," AES 82nd Convention 1987, London, preprint 2466.

[4] M. Williams, "Frequency Dependent Hybrid Microphone Arrays for Stereophonic Sound Recording," AES 92nd Convention 1992, Vienna, preprint 3252.

[5] R. Glasgal, "Ambiophonics: Achieving Physiological Realism in Music Recording and Reproduction," presented at the 111th Convention of the Audio Engineering Society, preprint 5426.

[6] G. Theile, "Natural 5.1 Music Recording Based on Psychoacoustic Principles," Proceedings of the 19th International Conference of the Audio Engineering Society, Schloss Elmau, Germany, June, 2001, revised October, 2001.

[7] M. Williams, "Microphone Array Analysis for Multichannel Sound Recording," AES 107th Convention 1999, New York, preprint 4997.

[8] M. Williams, "Multichannel Microphone Array Design," AES 108th Convention 2000, Paris, preprint 5157.

[9] J. Wuttke, "General Considerations on Audio Multichannel Recording," Proceedings of AES 19th International Conference, Schloss Elmau, Germany, 2001.

[10] H. Wittek, O. Neumann and M. Schäffler, C. Millet, "Studies on Main and Room Microphone Optimization," Proceedings of AES 19th International Conference, Schloss Elmau, Germany, 2001.

[11] D. Griesinger, "The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings and Their Relationship to Microphone Techniques," Proceedings of AES 19th International Conference, Schloss Elmau, Germany, 2001.

[12] A. Farina, R. Glasgal, E. Armelloni, A. Torger, "Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music," Proceedings of AES 19th International Conference, Schloss Elmau, Germany, 2001.

[13] R. Miller, "Surround Sound" pages, www.filmaker.com, evaluation DTS-CDs and streaming PC demonstration.

[14] K. Hamasaki, T. Shinmura, S. Akita, and K. Hihama (NHK), "Approach and Mixing Techniques for Natural Sound Recording of Multichannel Audio," as presented at AES 19th International Conference, 2001.

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

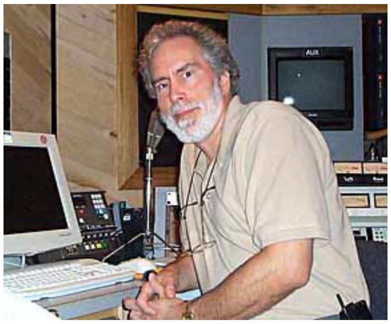
[16] B. Blesser, "An Interdisciplinary Integration of Reverberation ," AES 111th Convention 2001 as presented, paper 5468.

[17] W. Wosczczyk, G. Martin, P. Depalle, J. Corey, R. Quesnel "A Hybrid Model for Simulating Diffused First Reflections in Two-Dimensional Synthetic Acoustic Environments," Proceedings of AES 19th International Conference, Schloss Elmau, Germany, 2001.

[18] M. Gerzon, "Optimal Reproduction Matrices for Multispeaker Stereo," AES 91st Convention 1991, New York, preprint 3180.

[19] F. Rumsey, J. Berg, "Verification and correlation of attributes used for describing the spatial quality of reproduced sound," Proceedings of AES 19th International Conference, Schloss Elmau, Germany, 2001.

[20] "Multichannel surround sound systems and operations" - AES Technical Document ESTD1001.0.01-05



The Author

Robin Miller is a member of AES and SMPTE, BSEE (Lehigh University), musician, consultant, and 53 award-winning producer and studio owner with more than 40 years experience in film, television, and music sound recording and mixing.