Audio Engineering Society
Convention Paper
Presented at the 114th Convention
2003 March 22–25 Amsterdam, The Netherlands

Transforming Ambiophonic + Ambisonic 3D Surround Sound to & from ITU 5.1/6.1

Robert E. (Robin) Miller III [1]
©2003
[1] *FilmakerStudios, Robin Miller, Filmaker Inc.*
Bethlehem PA 18018 USA

ABSTRACT

ITU 6.1 with six discrete full-range audio channels, implemented in DVD-A, SACD, and DTS-ES Discrete, provide the means to deliver full sphere periphonic 3D surround sound. For compatible distribution, the channels are transformed for 5.1/6.1 reproduction, but can still be fully recovered for "PerAmbio" reproduction – an Ambisonic + Ambiophonic hybrid approach, described in a prior paper, that maximizes 3D envelopment along with front stage imaging and spaciousness, while economizing the number of channels and speakers. To clarify that fewer media channels "r" are required than speakers "s" the use of MCN - multichannel numbering, in the form "r.lfe.s" is proposed. Experimental "PerAmbio 6.1.10" (10 speakers or more + subwoofer) recordings test six encoding variations applicable in home theater, virtual reality, and music-only production.

PURPOSE OF PAPER

The surround audio market is driven by upgrading by growing numbers of DVD owners from 2-speaker stereo to multichannel surround in the horizontal plane. However judging by the response to the author's "periphonic" demonstrations to date, full sphere 3D surround is compelling and desirable. To listeners it is just as astonishing when height is *present*, as, conversely, it would be astonishing in real life were 3D *absent*. Audio engineers need to consider the consequences of not including a path to 3D in production, transmission, and reproduction standards.

This paper reports a method for transforming "PerAmbio" full sphere 3D to dual-format recordings that play directly in standard ITU 5.1/6.1 horizontal 2D surround with no extra decoding or relocation of speakers. Only when users choose to recover "full sphere 3D" surround would they need the described converter that derives 10 or more speaker feeds, plus subwoofer(s). This production, transmission, and reproduction system is applicable in music, cinema, broadcast, amusement ride, and Webcast markets.

Compatible with 5.1 AND extensible to 3D

Even as we progress toward the ultimate sonic experience – fully periphonic 3D sound of natural hearing – it is imperative that systems be compatible with 2D standard ITU 5.1 surround [1]. This premise not only recognizes the popular acceptance of 5.1, it also considers the value to entertainment providers and their customers of software libraries that need not be rendered obsolete – if both parties plan for the future. The new 3D systems described here and in prior papers are not intended to compete with 5.1 – they are the next step in the evolution started by it [2,3].

Compatibility with ITU 5.1/6.1 implies that:

- New 3D recordings should be able to play without compromise on current, standard 2D ITU 5.1/6.1 surround systems, such as in home theaters;
- Existing 5.1 and legacy 2-0 stereo recordings should play (possibly better) on 3D layouts;
- Users, manufacturers, and record companies should benefit by *having choices* - and by being able to make them in steps as it suits them.

For example, if manufacturers pre-elect to offer hardware and entertainment providers software along lines outlined in this paper, then the following scenario could play out in the marketplace:

1. *Initial user step* - Foresighted consumers comprise a demand for recordings made in 3D PerAmbio 6.1 (possible today using DVD-A, SACD, and DTS-CD) and initially enjoy them on their standard 2D ITU 5.1/6.1 systems;
2. *Optional interim user step:* Serious listeners can enjoy 2D PanAmbio 4.1 recordings (compatible with 5.1/6.1) using two closely spaced speaker pairs and crosstalk cancellation [4]. At least one manufacturer will offer this capability in 2003.
3. *Ultimate user step* – The user upgrades to fully periphonic 3D using the converter described in this paper, which reconstitutes from the 6 distribution channels 10 or more unique speaker feeds. Legacy 5.1/6.1 and stereo recordings can also be enjoyed as illustrated in Fig.1

PerAmbio 6.1.10+ is extensible to 3D and scalable from 10 speakers, to 14, to a practical maximum of 26, plus subwoofer(s). Dual-format PerAmbio recordings can be released that play in 5.1/6.1 and with a decoder in 3D, making future-proof priceless recorded performances and saving having to re-record others for 3D. Capable media such as DVD-A, SACD, and DTS-CD already exist for the music industry. However, existing broadcast standards need to be implemented for more than the 5 channels of AC-3, such as 48-channel AAC of the Advanced Television Standard. A more detailed description of PerAmbio 6.1.10+ appears later in this paper, and its requirements are summarized in Appendix A.

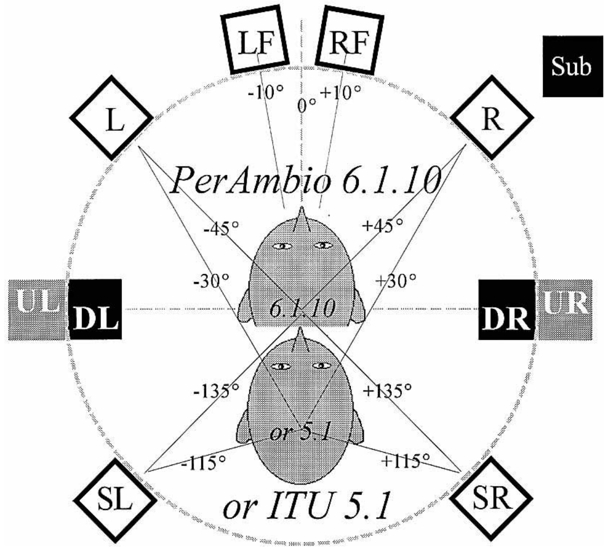

Fig.1 PerAmbio 6.1.10+ in plan view illustrates the minimum of ten speakers (plus subwoofer) – white at ear level, black on floor, gray at ceiling. By relocating the listener back 26% of the speaker layout diameter, speakers comply with ITU 5.1 angles.

HYBRID APPROACH TO DUAL-FORMAT 2D/3D SURROUND SOUND – PerAmbio 6.1.10+

In the author's experience, no single approach produces ideal surround sound. In 2D surround, ITU 5.1 including 6.1 works well for cinema-style (3-2), but 5.1/6.1 does not support 360° localization that is necessary for the natural reproduction of sources in an acoustic environment [5]. Conventional widely spaced AB or Decca Tree plus room mics fail ITU 5.1/6.1 both in frontal localization and an integrated room impression. Near-spaced directional microphones are weak in low frequencies, while OCT [4,6] suffers off-axis pick-up in the front left and right channels. PanAmbio 4.1 [2,4], an HRTF binaural-based approach using twin baffled sphere microphones (Ambiophones), delivers accurate image and spaciousness, but confines listening to the median plane [7]. In 3D, Ambisonics, like other coincident techniques such as X-Y and M-S, fails in adequate spaciousness. One can conclude that the cues necessary for human hearing of surround sound are not all neatly bundled in any one format! [8,9,10]

The solution is a hybrid approach, where cues lacking in one format are compensated for by another. Considering properties above, a system that combines HRTF-based Ambiophonics with Ambisonics, called PerAmbio 6.1.10+, could, in theory, optimize 2D or 3D envelopment + frontal imaging + spatial impression.

The system employs dual-format recording, processing, and 6-channel media so that both users and entertainment providers can upgrade from 2D to 3D at lower cost. PerAmbio combines Ambiophonics for an accurate front stage with Ambisonics for ambience. The combination is more than the sum of its parts. For example, there is no longer Ambio's limited listening area – listeners can venture freely (except very near any speaker) within both the 5.1/6.1 2D and 6.1.10+ 3D speaker layouts [2]. Surprisingly, listeners of PerAmbio 6.1.10+ in 3D can venture even *outside the speaker array* and enjoy plausible surround sound.

Fig.2 PerAmbio 6.1.10+ test microphone array in the concert hall.

Fig.3. 3D experimental array (from top) Soundfield microphone, OCT for 2D, and "Ambiophone" with directional elements instead of baffle to reject back sound and realized using Schoeps microphones.

2D surround today, 3D tomorrow from dual-format recordings

The system below provides manufacturers, entertainment providers, and home theater owners:

1) ITU 5.1/6.1 discrete replay in 2D (horizontal) surround sound for a standard home theater;
2) Periphonic full sphere 3D surround sound for home theater, PC/multi-media, game/virtual reality, training simulator, or amusement ride.

In horizontal 2D form, 6-channel distribution media are transformed for ITU 5.1/6.1 discrete replay directly with no special decoding. When the user elects to upgrade to fully periphonic 3D, the same 6 channels undergo lossless matrix conversion to reconstitute signals for 10 or more speakers.

6-channel PerAmbio 6.1.10+ microphone array

In its simplest form, PerAmbio requires an array of six microphones (Fig 2 & 3): a Soundfield microphone or equivalent (coincident omni plus three orthogonal figure-eights) and an Ambiophone – a baffled or directional form of sphere microphone [11] that favors frontal incidence. Six recording channels are needed. In certain situations and as 3D sound has been refined, up to ten microphones and channels have been used for tailoring the front stage image and preserving low frequency response.

During the recording session and through mastering, it is essential to match the levels of the channels within a fraction of a dB and to track or correct them in post. An acoustic filtered pink noise source must be recorded at several angles of incidence to the array for calibration and quality assurance.

TRANSFORM FOR DECODER-FREE 5.1/6.1

Although it would be trivial to simply deliver the six native PerAmbio channels in place of standard ITU 6.0 channels, it would repurpose these channels entirely. This paper describes a lossless algebraic matrix that can "encode" the six channels to play compatibly without any decoding, flattened to standard ITU 5.1 or 6.1 horizontal surround speaker layouts. Or instead, it can be decoded to PerAmbio 6.1.10 full sphere 3D surround by recovering the Ambisonic B-format (W,X,Y,Z channels) [12] plus a binaural-based Ambiophonic front stage. The B-format contribution can be sourced either from a Soundfield microphone or from 3D ambience auralization convolved during post-production or by the user from a library of 3D hall impulse responses. For reference, Ambisonics is summarized in Appendix B. Ambiophonics, after Glasgal et al [13,14,15], contributes a 150°-wide front stage with image accuracy of ±5° using a closely spaced "ambiopole" speaker-pair and crosstalk cancellation. The PerAmbio 5.1/6.1 transformation introduced here is linear and bi-directionally lossless. Users can have full sphere 3D surround by adding a matrix decoder and 4 or 5 speakers placed as in Fig.1.

After mastering the dual format 2D/3D recording from the six PerAmbio channels, given as {Pin} in 6x1 matrix form, a 6x6 transformation matrix {S} for standard 5.1/6.1 surround is applied to obtain the 6 media channels {Sout} as follows:

$$\{\text{Sout}\} = \{\text{S}\} \bullet \{\text{Pin}\}$$

$$\begin{pmatrix} L \\ R \\ C \\ SC \\ SL \\ SR \end{pmatrix} = \begin{pmatrix} s(L,FR) & s(L,FR) & s(L,W) & s(L,X) & s(L,Y) & s(L,Z) \\ s(R,FL) & s(R,FR) & s(R,W) & s(R,X) & s(R,Y) & s(R,Z) \\ s(C,FL) & s(C,FR) & s(C,W) & s(C,X) & s(C,Y) & s(C,Z) \\ s(SC,FL) & s(SC,FR) & s(SC,W) & s(SC,X) & s(SC,Y) & s(SC,Z) \\ s(SL,FL) & s(SL,FR) & s(SL,W) & s(SL,X) & s(SL,Y) & s(SL,Z) \\ s(SR,FL) & s(SR,FR) & s(SR,W) & s(SR,X) & s(SR,Y) & s(SR,Z) \end{pmatrix} \times \begin{pmatrix} FL \\ FR \\ W \\ X \\ Y \\ Z \end{pmatrix}$$

On a standard ITU home theater surround system, a multichannel disc (6 discrete channel DVD-A, SACD, or DTS-CD) plays {Sout} directly in 6.1. If it is 5.1, current implementations automatically add the SC information into SL and SR speaker feeds at –3dB.

When the user augments his/her system for fully periphonic 3D, a "reconstitution" matrix {P} is applied, implemented in DSP, that responds to flags in the metadata to select one of six recording modes, discussed later, to output PerAmbio 6.1.10+ as follows:

$$\{\text{Pout}\} = \{\text{P}\} \bullet \{\text{Sout}\}$$

Since matrix {P} is the inverse of matrix {S},

$$\{\text{Pout}\} = \{\text{S}\}^{-1} \bullet \{\text{Sout}\}$$

$$\begin{pmatrix} FL \\ FR \\ W \\ X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} s(L,FR) & s(L,FR) & s(L,W) & s(L,X) & s(L,Y) & s(L,Z) \\ s(R,FL) & s(R,FR) & s(R,W) & s(R,X) & s(R,Y) & s(R,Z) \\ s(C,FL) & s(C,FR) & s(C,W) & s(C,X) & s(C,Y) & s(C,Z) \\ s(SC,FL) & s(SC,FR) & s(SC,W) & s(SC,X) & s(SC,Y) & s(SC,Z) \\ s(SL,FL) & s(SL,FR) & s(SL,W) & s(SL,X) & s(SL,Y) & s(SL,Z) \\ s(SR,FL) & s(SR,FR) & s(SR,W) & s(SR,X) & s(SR,Y) & s(SR,Z) \end{pmatrix}^{-1} \times \begin{pmatrix} L \\ R \\ C \\ SC \\ SL \\ SR \end{pmatrix}$$

As confirmation that PerAmbio 3D is reconstituted without loss, an objective of the system is that

$$\{\text{Pout}\} = \{\text{Pin}\}$$

It will be shown that the requirements are met with only slight loss of SNR (1.4 to 3dB, or less than ¼ to ½ bit in 24) due to a headroom factor used to avoid clipping when random phase elements are matrixed.

Basic PerAmbio Recording Modes: i, j, k

Since ITU 5.1/6.1 is to be derived from 3D recordings – and since recording venues vary in their 3D space – the recordist must have control of how the 3D space for each recording is translated into the best result in 2D. Then in the end, it must be possible to reconstitute the 3D signals without loss.

The method provides the recording engineer with six acoustically based variations. In the interest of continuity in describing the matrix transformations, the six acoustic modes will be more fully described later. Experimental PerAmbio 6.1.10+ recordings have been made of a number of musical genres and venues both to verify the concept and to demonstrate it to so engineers can evaluate the results.

In order to be able to reconstitute height, one or more channels must be non-coplanar with the other channels. While there might be something to be said for having C coplanar with L & R, the argument is made (and results support) that C need not be coplanar with L & R. Nor need SC be coplanar with SL & SR. For 3D, non-coplanar channels increase height gain, thereby decreasing noise during rematrixing. Results in 2D are optimal because desirable non-co-planar sounds of the venue can be favored, e.g. ceiling ambience or upstage voices. In 5.1 this "compromise" is quite acceptable, since height is not reproduced in any case. In both 2D and 3D, any channel that is not coplanar with any other contributes spaciousness because the signals have greater probability of being decorrelated. Finally in 3D, human hearing resolves more acutely horizontally than vertically.

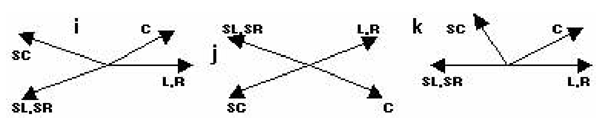

Fig.4 Illustrates in terms of elevation three basic modes I, j, & k for transformation to ITU 5.1/6.1 of amphitheater, concert with soloist & audience, and arena. Source is to the right.

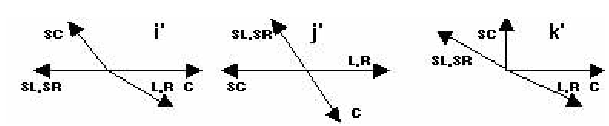

Fig.5 Shows in elevation three "tilted" modes i', j', & k' for transformation to 5.1/6.1 of opera, drama, and organ behind choir. Source is to the right.

Three basic modes are shown in Fig.4, and variants that place the microphone array high and "tilted" in Fig.5. Their applications will be described in more detail later, but range from recording opera to broadcasting sports in dual format 2D & 3D surround.

The 6x6 matrices Si, Sj, and Sk for transforming PerAmbio to ITU 6.1 are shown in Table 1, with mode elevation angles on the left. PerAmbio reconstitution matrices Pi, Pj, and Pk are shown in Table 2, with confirmation on the right that Pout=Pin for normalized full scale random phase addition of a sound emanating from 45° front, 45° left, and 45° up.

Mode "i" (inclined) — 6x6 coefs x 0.5000

°Up*	{Si} \	{Pin}: FL	FR	W	X	Y	Z
0	{Sout}: L	0.500	0.000	0.354	0.250	0.433	0.000
0	R	0.000	0.500	0.354	0.250	-0.433	0.000
30	C	0	0	0.354	0.433	0	0.250
30	SC	0	0	0.354	-0.433	0	0.250
-30	SL	0	0	0.354	-0.217	0.375	-0.250
-30	SR	0	0	0.354	-0.217	-0.375	-0.250
min/max rec	Transformation {S} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE

Mode "j" (juxtaposed l; descending C)

°Up*	{Sj} \	{Pin}: FL	FR	W	X	Y	Z
30	{Sout}: L	0.500	0.000	0.354	0.217	0.375	0.250
30	R	0.000	0.500	0.354	0.217	-0.375	0.250
-30	C	0	0	0.354	0.433	0	-0.250
-30	SC	0	0	0.354	-0.433	0	-0.250
30	SL	0	0	0.354	-0.217	0.375	0.250
30	SR	0	0	0.354	-0.217	-0.375	0.250
min/max rec	Transformation {S} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE

Mode "k" (on it's back)

°Up*	{Sk} \	{Pin}: FL	FR	W	X	Y	Z
0	{Sout}: L	0.500	0.000	0.354	0.250	0.433	0.000
0	R	0.000	0.500	0.354	0.250	-0.433	0.000
30	C	0	0	0.354	0.433	0	0.250
60	SC	0	0	0.354	-0.250	0	0.433
0	SL	0	0	0.354	-0.250	0.433	0.000
0	SR	0	0	0.354	-0.250	-0.433	0.000
min/max rec	Transformation {S} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE

Table 1. Matrix {S} coefficients to transform six PerAmbio channels in one of three modes i, j, or k into ITU-compatible form for replay without a decoder on standard 5.1/6.1 layouts. When desired, the six original PerAmbio channels can be recovered without loss for a 3D layout.

--ITU 6.1-to-PerAmbio reconstitution--> — *{B} noise degradation dB* 6.02

{Pi} \	{Sout}: L	R	C	SC	SL	SR	{Pout}
[Pout]: FL	2.000	0.000	-1.327	0.327	-1.655	0.655	1.00
FR	0.000	2.000	-1.327	0.327	0.655	-1.655	1.00
W	0.000	0.000	1.061	0.354	0.707	0.707	0.71
X	0.000	0.000	1.155	-1.155	0.000	0.000	0.50
Y	0.000	0.000	0	0	1.333	-1.333	0.50
Z	0.000	0.000	0.500	1.500	-1.000	-1.000	0.50
Reconstitute {P}	min/max	-1.655 2.000		{Pout}-{Pin}= 0			{P}x{Sout}

--ITU 6.1-to-PerAmbio reconstitution--> — *{B} noise degradation dB* 6.02

{Pj} \	{Sout}: L	R	C	SC	SL	SR	{Pout}
[Pout]: FL	2.000	0.000	-1.000	1.000	-2.000	0.000	1.00
FR	0.000	2.000	-1.000	1.000	0.000	-2.000	1.00
W	0.000	0.000	1.061	0.354	0.707	0.707	0.71
X	0.000	0.000	1.155	-1.155	0.000	0.000	0.50
Y	0.000	0.000	0	0	1.333	-1.333	0.50
Z	0.000	0.000	-0.500	-1.500	1.000	1.000	0.50
Reconstitute {P}	min/max	-2.000 2.000		{Pout}-{Pin}= 0			{P}x{Sout}

--ITU 6.1-to-PerAmbio reconstitution--> — *{B} noise degradation dB* 7.27

{Pk} \	{Sout}: L	R	C	SC	SL	SR	{Pout}
[Pout]: FL	2.000	0.000	-1.464	0.845	-1.691	0.309	1.00
FR	0.000	2.000	-1.464	0.845	0.309	-1.691	1.00
W	0.000	0.000	1.035	-0.598	1.195	1.195	0.71
X	0.000	0.000	1.464	-0.845	-0.309	-0.309	0.50
Y	0.000	0.000	0	0	1.155	-1.155	0.50
Z	0.000	0.000	0.000	2.309	-1.155	-1.155	0.50
Reconstitute {P}	min/max	-1.691 2.309		{Pout}-{Pin}= 0			{P}x{Sout}

Table 2. PerAmbio reconstitution matrix {P}, by which the six original channels can be recovered for fully periphonic 3D surround.

Initially, Gerzon "TriField" coefficients were used to derive Center information from L & R [16], but redundant results conflicted with the frontal imaging of the hybrid approach, along with narrowing the front stage, especially when played Ambiophonically.

Correcting 5.1/6.1 speaker placement

No decoder is necessary for normal 5.1/6.1 replay. However, a "smart decoder" would allow users to correct for non-standard speaker placements in ways not possible with conventional multichannel recordings. In this case, PerAmbio would be reconstituted as though for 3D, but then re-transformed with differing speaker azimuths and elevations.

Digitally tilting 3D space – modes i', j', k'

Tilting the Ambisonic image offers the recording engineer flexibility, doubling the number of applicable transformation modes, and aligning the Ambisonic image vertically with the L,R image. Tilting is usually associated with raising the microphone, suspended or on a high stand. Lower gives more the perspective of sitting in a front row, while raised gives a "balcony" perspective and puts the microphone above audience and camera sight lines. Raised and in 3D, we can discern vertical localization of upstage v. downstage sources, such as pit orchestra v. singers, organ above chorus, playing field sounds below crowd, etc. Tilting does not necessarily mean physically tilting the array, but electrically changing an Ambisonic array's angle of inclination, "elevation" Φ, using a tilting matrix designated {B'}.

Using Ambisonic B-format, "tilting" at angle Φ° from horizontal can be implemented either during recording or in postproduction using

$$w' = w$$
$$x' = x*\cos \Phi° + z*\sin \Phi°$$
$$y' = y$$
$$z' = z*\cos \Phi° - x*\sin \Phi°$$

Or in matrix form by tilting {S} to {S'} by {B'}

$$\{S'\} = \{B'\} \bullet \{S\}$$

For correct horizontal orientation of the reconstituted Ambisonic B format, an "untilting" matrix must be employed upon replay

w’’ = w’
x’’ = x’*cos Φ° + z’*sin Φ°
y’’ = y’
z’’ = z’*cos Φ° – x’*sin Φ°

Or in matrix form, untilting the PerAmbio 3D reconstitution matrix {P} is the inverse of untilting {S}

$$\{P\} = \{\ \{B'\}^{-1} \bullet \{S'\}\ \}^{-1}$$

As examples, for mode “i” the 6x6 tilting matrix {Bi’} is shown for –30° tilt in Table 3. The “untilting” inverse matrix $\{Bi\}=\{Bi'\}^{-1}$ is shown for +30° in Table 4. Transformation to 2D ITU5.1/6.1 with a –30° tilt is shown in Table 5. Reconstitution of PerAmbio 3D by +30° “untilting” back to horizontal is in Table 6.

Tilt {Bi’}	FL	FR	W	-30° X	Y	Z
L’	1.000	0	0	-0.063	-0.25	-0.358
R’	0	1.000	0	-0.062	0.25	-0.358
W’	0	0	1.000	0.088	0	0.153
X’	0	0	0	0.866	0	-0.500
Y’	0	0	0	0	1.155	0
Z’	0	0	0	0.375	0	0.650

Table 3. Tilting matrix {Bi’} for “tilting” the mode i soundscape –30° provides three additional recording modes i’, j’, or k’.

Untilt {Bi}	L’	R’	W’	+30° X’	Y’	Z’
L”=L	1.000	0	0	-0.125	0.217	0.455
R”=R	0	1.000	0	-0.125	-0.217	0.455
W”=W	0	0	1.000	0	0	-0.236
X”=X	0	0	0	0.866	0	0.667
Y”=Y	0	0	0	0	0.866	0
Z”=Z	0	0	0	-0.500	0	1.155

Table 4. “Untilting” matrix {Bi} restores to horizontal six discrete transformed channels on DVD-A, SACD, or DTS-CD (+30° shown).

Mode i’ high tilted		-30 ° tilt		{S’}={S}x{B’}			
°Up*	{Si’} \	{Pin}: FL	FR	W	X	Y	Z
-30	{Sout}: L	0.500	0.000	0.354	0.217	0.375	-0.250
-30	R	0.000	0.500	0.354	0.217	-0.375	-0.250
0	C	0	0	0.354	0.500	0	0.000
60	SC	0	0	0.354	-0.250	0	0.433
0	SL	0	0	0.354	-0.250	0.433	0.000
0	SR	0	0	0.354	-0.250	-0.433	0.000
min/max rec	Transformation {S’} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE
Mode j’ high tilted							
°Up*	{Sj’} \	{Pin}: FL	FR	W	X	Y	Z
0	{Sout}: L	0.500	0.000	0.354	0.281	0.308	-0.071
0	R	0.000	0.500	0.354	0.281	-0.308	-0.071
-60	C	0	0	0.354	0.313	0	-0.325
0	SC	0	0	0.354	-0.438	0	0.108
60	SL	0	0	0.354	-0.062	0.433	0.325
60	SR	0	0	0.354	-0.063	-0.433	0.325
min/max rec	Transformation {S’} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE
Mode k’ high tilted							
°Up*	{Sk’} \	{Pin}: FL	FR	W	X	Y	Z
-30	{Sout}: L	0.500	0.000	0.354	0.217	0.375	-0.250
-30	R	0.000	0.500	0.354	0.217	-0.375	-0.250
0	C	0	0	0.354	0.500	0	0.000
90	SC	0	0	0.354	-0.023	0	0.460
30	SL	0	0	0.354	-0.185	0.500	0.179
30	SR	0	0	0.354	-0.185	-0.500	0.179
min/max rec	Transformation {S’} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE

Table 5. {S’} matrix coefficients transform six PerAmbio channels according to one of three tilted modes I’, j’, or k’ into ITU-compatible form for replay without a decoder for 5.1/6.1 layouts (–30° tilt shown).

--ITU 6.1-to-PerAmbio reconstitution-->				{B} noise degradation dB 7.27			
{Pi’} \	{Sout}: L	R	C	SC	SL	SR	{Pout}
{Pout}: FL	2.000	0.000	-1.244	1.155	-1.821	-0.089	1.00
FR	0.000	2.000	-1.244	1.155	-0.089	-1.821	1.00
W	0.000	0.000	0.943	0.000	0.943	0.943	0.71
X	0.000	0.000	1.333	0.000	-0.667	-0.667	0.50
Y	0.000	0.000	0	0	1.155	-1.155	0.50
Z	0.000	0.000	0.000	2.309	-1.155	-1.155	0.50
Reconstitute {P}	min/max	-1.821 2.309		{Pout}-{Pin}= 0			{P}x{Sout}
--ITU 6.1-to-PerAmbio reconstitution-->				{B} noise degradation dB 6.02			
{Pj’} \	{Sout}: L	R	C	SC	SL	SR	{Pout}
{Pout}: FL	2.000	0.000	-1.372	0.461	-1.256	0.167	1.00
FR	0.000	2.000	-1.372	0.461	0.167	-1.256	1.00
W	0.000	0.000	1.179	0.707	0.471	0.471	0.71
X	0.000	0.000	0.667	-2.000	0.667	0.667	0.50
Y	0.000	0.000	0	0	1.155	-1.155	0.50
Z	0.000	0.000	-1.155	-1.155	1.155	1.155	0.50
Reconstitute {P}	min/max	-2.000 2.000		{Pout}-{Pin}= 0			{P}x{Sout}
--ITU 6.1-to-PerAmbio reconstitution-->				{B} noise degradation dB 9.80			
{Pk’} \	{Sout}: L	R	C	SC	SL	SR	{Pout}
{Pout}: FL	2.000	0.000	-1.647	2.003	-1.928	-0.428	1.00
FR	0.000	2.000	-1.647	2.003	-0.428	-1.928	1.00
W	0.000	0.000	1.035	-1.142	1.468	1.468	0.71
X	0.000	0.000	1.268	0.808	-1.038	-1.038	0.50
Y	0.000	0.000	0	0	1.000	-1.000	0.50
Z	0.000	0.000	-0.732	3.089	-1.179	-1.179	0.50
Reconstitute {P}	min/max	-1.928 3.089		{Pout}-{Pin}= 0			{P}x{Sout}

Table 6. {P’} matrix reconstitutes original PerAmbio 6.1.10+ channels for full periphonic 3D surround and “untilts” to horizontal (+30° shown).

Refining the transformation matrices

The 6x6 transformation to ITU 5.1/6.1 described above with a headroom margin of 6dB has the possibility of clipping. However unlikely to be noticeable, with a safety factor of 0.5, random phase components could add instantaneously to 117% in the L or R channels. The safety gain of 0.5 also lowers L & R replay levels 6dB. Furthermore, the 6x6 matrix reconstitution of PerAmbio, where the safety gain of 0.5 is made up, causes loss of SNR of a minimum of 6dB to a maximum 9.8dB, or from 1 bit to just over 1.6 of 24 bits. In addition, the six tilt and untilt matrices for i, j, & k are all different.

There is a solution: Since Ambiophonics’ has high localization accuracy horizontally, and because it involves crosstalk cancellation, it is advisable to assign only the Ambio pair to L,R and not include Ambisonic information. This prevents conflicting imaging cues arriving at L,R speakers from two different pickup systems. Whether a price, believed to be acceptable, is paid in reducing the number of speakers reproducing ambience information from six to four in 2D requires further investigation (eight or more speakers are used for ambience in 3D). Also since Ambiophonics is largely insensitive to a change in inclination over the range of angles normally associated with tilting, say 0 to –45°, the Ambisonic components may be electrically tilted independently over this range.

The preceding suggests simplifying both the transformation matrix {S} and reconstituting matrix {P} from 6x6 to 4x4, reducing the number of DSP coefficients from 36 to 16 as shown in Table 7 & 8.

Tilt matrix {B'} and "untilt" matrix {B} are now also 4x4 and are respectively the same for all modes, as in Table 9 & 10. With no ambience added to L,R there is no opportunity for clipping during transformation of random phase components. Furthermore, during reconstitution of PerAmbio 3D, noise from making up the more gentle safety factor of 0.85 results in loss of SNR of only 1.4dB for modes i, i', j, & j' and no more than 3dB for k & k', which is just over ½ of 24 bits lost for these two modes. L,R levels are now reduced only 1.4dB compared to 6dB for 6x6 matrices. The 4x4 transformation modes are shown in Fig.6 & 7.

Mode "i" (inclined) — 4x4 coefs x 0.8500

°Up*	{Si} \ {Pin}:	FL	FR	W	X	Y	Z
0	{Sout}: L	0.850	0.000				
0	R	0.000	0.850				
30	C			0.601	0.736	0	0.425
30	SC			0.601	-0.736	0	0.425
-30	SL			0.601	-0.368	0.638	-0.425
-30	SR			0.601	-0.368	-0.638	-0.425
min/max rec	Transformation {S} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE

Mode "j" (juxtaposed I; descending C)

°Up*	{Sj} \ {Pin}:	FL	FR	W	X	Y	Z
0	{Sout}: L	0.850	0.000				
0	R	0.000	0.850				
-30	C			0.601	0.736	0	-0.425
-30	SC			0.601	-0.736	0	-0.425
30	SL			0.601	-0.368	0.638	0.425
30	SR			0.601	-0.368	-0.638	0.425
min/max rec	Transformation {S} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE

Mode "k" (on it's back)

°Up*	{Sk} \ {Pin}:	FL	FR	W	X	Y	Z
0	{Sout}: L	0.850	0.000				
0	R	0.000	0.850				
30	C			0.601	0.736	0	0.425
60	SC			0.601	-0.425	0	0.736
0	SL			0.601	-0.425	0.736	0.000
0	SR			0.601	-0.425	-0.736	0.000
min/max rec	Transformation {S} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE

Table 7. 4x4 matrix {S} (lightly shaded) transforms six-channel PerAmbio 3D into one of three modes I', j', or k' for ITU 5.1/6.1 replay without a decoder, or with a converter back to PerAmbio 3D.

--ITU 6.1-to-PerAmbio reconstitution--> {B} noise degradation dB 1.41

{Pi} \ {Sout}:	L	R	C	SC	SL	SR	{Pout}
{Pout}: FL	1.176	0.000					1.00
FR	0.000	1.176					1.00
W			0.624	0.208	0.416	0.416	0.71
X			0.679	-0.679	0.000	0.000	0.50
Y			0	0	0.784	-0.784	0.50
Z			0.294	0.882	-0.588	-0.588	0.50
Reconstitute {P}	min/max		-0.784 1.176		{Pout}-{Pin}= 0		{P}x{Sout}

--ITU 6.1-to-PerAmbio reconstitution--> {B} noise degradation dB 1.41

{Pj} \ {Sout}:	L	R	C	SC	SL	SR	{Pout}
{Pout}: FL	1.176	0.000					1.00
FR	0.000	1.176					1.00
W			0.624	0.208	0.416	0.416	0.71
X			0.679	-0.679	0.000	0.000	0.50
Y			0	0	0.784	-0.784	0.50
Z			-0.294	-0.882	0.588	0.588	0.50
Reconstitute {P}	min/max		-0.882 1.176		{Pout}-{Pin}= 0		{P}x{Sout}

--ITU 6.1-to-PerAmbio reconstitution--> {B} noise degradation dB 2.66

{Pk} \ {Sout}:	L	R	C	SC	SL	SR	{Pout}
{Pout}: FL	1.176	0.000					
FR	0.000	1.176					
W			0.609	-0.352	0.703	0.703	0.71
X			0.861	-0.497	-0.182	-0.182	0.50
Y			0	0	0.679	-0.679	0.50
Z			0.000	1.358	-0.679	-0.679	0.50
Reconstitute {P}	min/max		-0.679 1.358		{Pout}-{Pin}= 0		{P}x{Sout}

Table 8. 4x4 matrix {P} (lightly shaded) reconstitutes six PerAmbio 6.1.10+ channels without loss for fully periphonic 3D surround.

		Tilt {B'}	W	X	Y	Z
-30	° tilt	W'	1.000	0	0	0
		X'	0	0.866	0	-0.500
		Y'	0	0	1.000	0
		Z'	0	0.500	0	0.866

Table 9. 4x4 {Bi'} matrix "tilts" recording modes i', j', or k' for ITU 5.1/6.1 replay. Any angle may be used (-30° shown).

Untilt {B"}	W'	X'	Y'	Z'
W"=W	1.000	0	0	0
X"=X	0	0.866	0	0.500
Y"=Y	0	0	1.000	0
Z"=Z	0	-0.500	0	0.866

Table 10. 4x4 {B} matrix "untilts" back to horizontal B-format recording modes i', j', or k' for PerAmbio 3D (+30° shown).

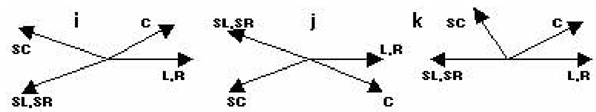

Fig.6 illustrates in terms of elevation three basic modes I, j, & k for 4x4 transformation to ITU 5.1/6.1 of amphitheater, concert with soloist & audience, and arena. Source is to the right.

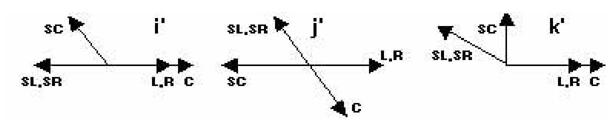

Fig.7 shows three "tilted" modes i', j', & k' for 4x4 transformation to 5.1/6.1 of opera, drama, and organ behind choir. Source is to the right.

The revised 4x4 transformation matrix {S} with –30° tilt is shown in Table 11. The new 4x4 reconstitution matrix {P} with +30 "untilt" is shown in Table 12.

Mode *i' high tilted*

°Up*	{Si'} \ {Pin}:	FL	FR	W	X	Y	Z
-30	{Sout}: L	0.850	0.000				
-30	R	0.000	0.850				
0	C			0.601	0.850	0	0.000
60	SC			0.601	-0.425	0	0.736
0	SL			0.601	-0.531	0.638	-0.184
0	SR			0.601	-0.531	-0.638	-0.184
min/max rec	Transformation {S'} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE

Mode *j' high tilted*

°Up*	{Sj'} \ {Pin}:	FL	FR	W	X	Y	Z
-30	{Sout}: L	0.850	0.000				
-30	R	0.000	0.850				
-60	C			0.601	0.425	0	-0.736
0	SC			0.601	-0.850	0	0.000
60	SL			0.601	-0.106	0.638	0.552
60	SR			0.601	-0.106	-0.638	0.552
min/max rec	Transformation {S'} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE

Mode *k' high tilted*

°Up*	{Sk'} \ {Pin}:	FL	FR	W	X	Y	Z
-30	{Sout}: L	0.850	0.000				
-30	R	0.000	0.850				
0	C			0.601	0.850	0	0.000
90	SC			0.601	0.000	0	0.850
30	SL			0.601	-0.368	0.736	0.213
30	SR			0.601	-0.368	-0.736	0.213
min/max rec	Transformation {S'} 4x4 or 6x6			W=0.707W	cosAcosE	sinAcosE	sinE

Table 11. 4x4 matrix {S'} (lightly shaded) transforms six PerAmbio 3D channels into one of three modes I', j', or k' for standard ITU 5.1/6.1 replay without a decoder (–30° tilt shown).

--ITU 6.1-to-PerAmbio reconstitution--> {B} noise degradation dB 1.41

{Pi'} \	{Sout}: L	R	C	SC	SL	SR	{Pout}
[Pout}: FL	1.176	0.000					1.00
FR	0.000	1.176					1.00
W			0.624	0.208	0.416	0.416	0.71
X			0.735	-0.147	-0.294	-0.294	0.50
Y			0	0	0.784	-0.784	0.50
Z			-0.085	1.104	-0.509	-0.509	0.50
Reconstitute {P}		min/max	-0.784 1.176		{Pout}-{Pin}= 0		{P}x{Sout}

--ITU 6.1-to-PerAmbio reconstitution--> {B} noise degradation dB 1.41

{Pj'} \	{Sout}: L	R	C	SC	SL	SR	{Pout}
[Pout}: FL	1.176	0.000					1.00
FR	0.000	1.176					1.00
W			0.624	0.208	0.416	0.416	0.71
X			0.441	-1.029	0.294	0.294	0.50
Y			0	0	0.784	-0.784	0.50
Z			-0.594	-0.425	0.509	0.509	0.50
Reconstitute {P}		min/max	-1.029 1.176		{Pout}-{Pin}= 0		{P}x{Sout}

--ITU 6.1-to-PerAmbio reconstitution--> {B} noise degradation dB 3.08

{Pk'} \	{Sout}: L	R	C	SC	SL	SR	{Pout}
[Pout}: FL	1.176	0.000					1.00
FR	0.000	1.176					1.00
W			0.609	-0.352	0.703	0.703	0.71
X			0.746	0.249	-0.497	-0.497	0.50
Y			0	0	0.679	-0.679	0.50
Z			-0.431	1.425	-0.497	-0.497	0.50
Reconstitute {P}		min/max	-0.679 1.425		{Pout}-{Pin}= 0		{P}x{Sout}

Table 12. 4x4 matrix {P'} reconstitutes six PerAmbio 6.1.10+ channels without loss and restores any tilt for periphonic 3D replay (+30° shown).

Other conversions for 5 or 8 media channels?

The discussion above suggests "PanAmbio 5.1" – a 2D hybrid of Ambiophonic and Ambisonic WXY-only transformed to the five full-range channels of AC-3, and reconstituted when a decoder is provided to the original Ambio+Ambisonic WXY-only. If the original recording included all four B-format signals, tilting would be possible prior to 5.1 mastering, but would permanently alter the ambience space, as upon replay no reconstituted Z is available for "untilting." The result is equivalent to sitting in the balcony where one might naturally tilt his/her head to hear the sound produced, although in this case without visual cue.

If 8 full-range channels are available (such as 8 channel Dolby E or 8 of the 48 channels specified in MPEG4/AAC), then 3D "Pan/PerAmbio 8.0" with extra channels can be realized: three in the 2D form above or two in 3D. Two or three extra channels might be used for L,R or L,C,R spot mics, three second order Ambisonic signals, second (third, fourth) language(s), additional effects speakers, etc.

Stability of center voices; the "Sweet Spot"

For 5.1/6.1 replay, superimposition of the Ambiophonic image with the Ambisonic contribution stabilizes important central voices by anchoring them with a hard Center channel derived from the B-format at 0° azimuth and tilted or not depending on the 2D transformation mode chosen. Unlike phantom images in 2-speaker stereo or surround mixing that ignores the advantage of a hard center channel, in the PerAmbio 2D transformation central soloists do not toggle to the nearer speaker as one moves around the listening space, but remain a stable central image.

Reconstituted to PerAmbio 3D, the hybrid approach largely solves Ambiophonics' main disadvantage: By supplementing a large listening area, there is no longer the absolute need to sit on the median plane for enjoyable sound. In a modest home theater, six music listeners or movie viewers can be accommodated with very plausible surround sound, although the middle two listeners on the median plane will benefit from the frontal localization accuracy of the Ambio pair, as shown in Fig.8.

Ambisonic transformation in the horizontal plane is a regular hexagon, creating virtual speakers at ±60°, ±120, 0°, and 180°. L & R are virtualized wider than the standard ITU angle of 60° in order to match the 120°+ reproduction angle of Ambiophonics, with its inherent frontal image accuracy of ±5° [4,7].

Vertical ambience transformations are necessarily not coplanar, as described above. However, vertical acuity of human hearing relies on learned pinna response and is much less than horizontal acuity, which relies on HRTF level and time differences (ILD and ITD) of our two ears. Note that conventional recordings are often made with room and spot microphones placed far from the main microphone and likely not coplanar with it.

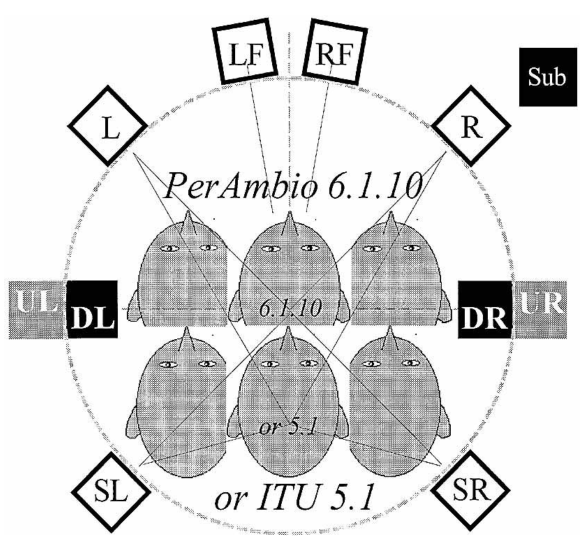

Fig.8 – Compatible PerAmbio 6.0.10 full sphere 3D surround layout can accommodate audiences up to six. For 5.1, viewers sit back 26% of the speaker diameter, where the angles meet ITU standards (with DSP changes in levels/delays). For 3D music appreciation, one or two listeners sit at the focus of 10 speakers, plus subwoofer(s).

ITU 6.1 TRANSFORMATION MODES: i j k i' j' k'

80 combinations (= $3^4 – 1$) were considered for transformation matrices to encode 3D directionality into 6 full range ITU media channels for reconstituting full periphony, but only about a dozen proved useful. If metadata permitted unlimited flags to command the user's processor, all 80 could be available to the recording engineer. Each conversion matrix is at most a 6x6 array of coefficients for each mode, accessible to the DSP by a table lookup. It is also possible for the user to download newly developed coefficients from the Internet to a decoder's FLASH memory.

This paper presents six "modes" for a six-channel main microphone array in six common applications for music recording, cinema ambience, and multichannel broadcasting (see Fig.6 & 7). Work is ongoing to refine these choices.

The three basic modes, i, j, & k, are so designated for mnemonics that describe their function. **"i"** has both C and SC "inclined" upward with respect to the four "corner" channels L, R, SL, & SR inclined downward. **"j"** "juxtaposes" C with LR and SC with SL & SR – and is the only letter designation with a descender, which reminds us that C tilts downward. **"k"** lying on its back represents that C and SC angle upward from the corner channels, which lie flat.

The three tilted variant modes rotate C, SC, SL, and SR with respect to L,R by any practical angle, e.g. –30°, in order to raise the microphone (suspended or on a high stand). The output of the sphere microphone does not vary with height incidence, but the baffled "ambiophone" does, so physical tilting may be appropriate for the FL, FR channels. The same applies were an ORTF microphone used for FL,FR.

The choice of mode may be made either during recording or post-production. If in post it is desired to change the basic mode or tilt, the original PerAmbio channels may be reconstituted and a new mode and tilt transformation made. A raised (suspended) microphone perspective is irreversible. During mastering, a flag is set in metadata of the dual-format recording in order for users' replay equipment to reconstitute 3D without loss.

From experience, most recording engineers can identify applications that spawned the modes below (cf. Fig.6, 7). Or, even without hearing the hall in Fig.9, which of the following modes would you choose (keeping in mind that you can change it in post)?

- **i** The microphone array is placed at source level (L, R), below acoustic shell reflections (C), e.g. an outdoor amphitheater event with audience low and behind (SL, SR) and raked upward (SC).
- **i'** The microphone array is on a high stand or hanging in opera house or orchestra hall with orchestra widely spaced in a pit or strings downstage (L, R), singers or winds upstage (C), and hall ambience back (SL, SR) and up (SC).
- **j** The array is more closely placed before a small ensemble at source level for direct sound and early floor and side wall reflections (L, R), higher direct solo and ceiling reflections (C), and hall ambience from back-up (SL, SR) and back-down (SC).
- **j'** The array is hanging closer to a proscenium to pickup downstage event sounds (L, R), upstage drama (C), high-back hall ambience (SL, SR), and additional audience pickup (SC).
- **k** The microphone array is in an arena with sports play-action or musical instruments at microphone level (L, R), and with good high-front (C) and back (SC) crowd sounds or ceiling ambience.
- **k'** The array is on a high stand or hanging in a cathedral with upstage choir (C) and front-of-church organ divisions and floor reflections (L, R), organ antiphonal and congregation in back (SL, SR), and organ trumpet directly overhead (SC).

Fig.9 - shows the PerAmbio 6.1.10+main microphone array in a recital hall for experimental recordings to test dual format 2D/3D. Mode j' ("juxtaposed", high and tilted) was preferred for ITU6.1 transformation.

RECONSTITUTING FULL PERIPHONIC 3D

The year is not 1492; the world of sound is not flat! To implement "periphonic" full sphere 3D surround for an uncompromised experience, the author has determined experimentally [2] a minimum requirement of 10 speakers (plus subwoofer). From the six full-range media channels (plus optional LFE) can be derived feeds for any number of speakers, but there is diminishing return above 10, illustrated in Fig.10.

In order for consumers to recover PerAmbio periphonic 3D from the dual-format recording, a "smart decoder" with DSP is required, controlled by one of six (or more) reconstitution matrices, selected by detecting a flag in metadata. Then an ambience speaker feed matrix and at least 8 speakers if not 12 or 24 are required, as shown for 8 in Table 13, in addition to the Ambiophonic pair FL,FR, plus subwoofer(s).

AMBISONIC SPEAKER FEEDS		cosAcosE	sinAcosE	sinE	Speaker placements	
Gen Ambis8 Dec	*W=0.707W*	*X*	*Y*	*Z*	*Azim°CCW*	*Elev°Up*
L	1.000	0.707	0.707	0.000	45	0
R	1.000	0.707	-0.707	0.000	315	0
UL	1.000	0.000	0.707	0.707	90	45
UR	1.000	0.000	-0.707	0.707	270	45
BL	1.000	-0.707	0.707	0.000	135	0
BR	1.000	-0.707	-0.707	0.000	225	0
DL	1.000	0.000	0.707	-0.707	90	-45
DR	1.000	0.000	-0.707	-0.707	270	-45

Table 13. Matrix to feed an 8-speaker Ambisonic array, a regular cube tilted forward 45° for PerAmbio 6.1.10+ (Fig.10). W =1.000 because 0.707 (-3dB) normal at this stage was applied in transformation {S}.

PerAmbio 6.1.10+ speaker layout (3D)

The minimum PerAmbio layout is shown in Fig.10. It is described as a main Ambiophonic pair (FL, FR) plus a regular cubic Ambisonic array of eight speakers tilted forward 45° [2]. Additional speakers may be placed isotropically so the ambience array becomes a cuboctahedron (12 speakers) or a practical maximum of 24 speakers in a dodecahedral array.

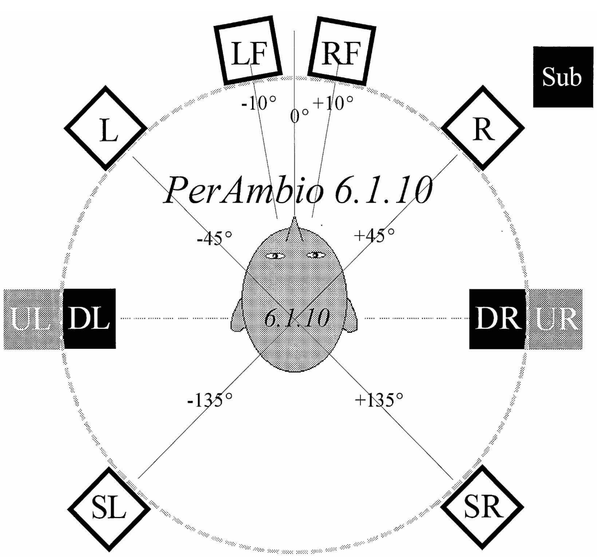

Fig.10 PerAmbio 6.1.10+ with 10 speakers plus subwoofer(s) – white at ear level, black on floor, gray at ceiling. Ambience can grow from 8 to 12 speakers (cuboctahedron), or to 24 (dodecahedron).

PerAmbio 6.1.10+ delivery media (3D)

Whether using the minimum of 10 or a greater array of 14 or 26 speakers, + subwoofer(s), the approach is named here "PerAmbio 6.1.10+" – see "Multichannel Numbering" below. Music can be delivered to users today using DVD-A, SACD, or DTS-ES Discrete. Current DVD-V can also transport PerAmbio using DTS 6.1 ES Discrete, but not using Dolby Digital 5.1 or DD-EX because it derives SC from a matrix.

PerAmbio 3D surround cannot be delivered to television audiences with current implementations for terrestrial Advanced Television (ATV) broadcast or cablecast, nor with DVD-V (except using DTS-ES 6.1 Discrete) [17]. These standards would need to be updated for 3D, or implemented e.g. using MPEG4 AAC for ATV [18]. The DVD-V standard provides for 8 channels that also have yet to be implemented. However, while DD/AC-3 in currently implemented ATV and DVD-V standards is incapable of PerAmbio 3D, DD/AC-3 can transport another Ambiophonic-based 2D surround reproduction technique, PanAmbio 4.1, which is compatible with ITU 5.1 and offers tools for 5.1 production, as explored in prior papers [2,4,7].

Multichannel Numbering in the form *r.lfe.s*

Time was the number of production channels equaled the number of media channels, which in turn equaled the number of speakers. In the new order, the number of production channels, number of media channels, and number of speakers can all be different!

Consumer marketers have blurred the standard designations 5.1, 6.1 etc. to mean "number of speakers," not discrete production or media channels. Reputable manufacturers avoid a numbers game: Dolby officially does not designate DD-EX (matrix surround center) as 6.1, as it is really still 5.1 channels. DTS-ES 6.1 Discrete is a true 6.1 channel system.

To clarify that fewer media channels "r" are required than speakers "s" (when their feeds can be derived during replay), the author proposes [2] use of MCN - multichannel numbering, in the form *"r.lfe.s"* – which simply appends to the usual number of media channels, r (mnemonic "radioed"), a second decimal point and number, s, to indicate unique speaker feeds.

Examples are: 5.1.5 for ITU standard 5.1; 5.1.6 for Dolby EX (SC matrixed from LR, RS); 6.1.6 for DTS-ES 6.1 Discrete, etc. PerAmbio 3D is 6.0.10+

varying with the user's number of ambience speakers and feeds (the recording is Ambisonically encoded for any number). Subwoofer(s) are always implied depending on the user's installation.

BEYOND 5.1: PERIPHONIC 3D SURROUND?

With the exception of IMAX, "multichannel surround sound" today means surround in the horizontal plane. Spawned by cinema, only left and right boundaries of the picture frame are extended by the sound "image" to full circle; vertical sound is contributed only by reverberation of the theater (significant if large) or home theater (ranging from not much to awful if untreated). This definition of "surround" was predicated on the oversimplification that humans have but two ears on a horizontal axis.

However, human ears also have pinna. These personalized appendages are direction finders for confirming fine horizontal localization to about 1°, but also cueing height above and below the plane of the ears. So future music recording, Advanced Television, and cinema sound need to expand the definition of "surround" to include height. Indeed, IMAX's five-story screen hides a single height speaker near its top. One need only experience full sphere surround with height, as the author demonstrates daily (Fig.11), to recognize that its contribution is as important to the ear as a jump in picture resolution of HDTV is to the eye.

Why height?

When we experience live events, our eyes+brain system resolves an image with "pixels" measuring half a minute of arc, exceeding high definition television at normal viewing distances. And we are immersed in a bubble of pulsing air, where, at for example a sporting event, we expect to hear the crowd elevated and the action below. Or the inverse – an air show above and the crowd below, or a lofty pipe organ with choir below. In a church, we would surely notice if organ and choir were squashed together at the horizon. Or if at the opera, suddenly the ceiling and floor were on the same plane, so that all the reflections we've come to expect to be strongly and spherically directional would flatten, erasing much of the acoustic signature of the hall, along with our envelopment within and enjoyment of it. We expect in real life our ears+brain system to resolve a sonic "image" with acuity of one degree horizontal and several degrees vertical – due to changes in comb filtering by the pinna with varying elevation. In real life, we would be astonished, were 3D sound suddenly absent!

Conversely, since we can not yet expect this level of realism from our home theater, we are just as astonished when we hear our first reproduction of full periphony – 3D sound that blows away the walls of the listening room, coming at us from everywhere in the *sphere* – not just the circle – wherein *we* are the psychoacoustic center. If not direct sources above and below (front as well as back), directional room reflections affect our perception of "natural" tonal color *of all sources.* Many hearing the author's demonstration of 3D drop their jaw, go wide-eyed, and express a desire to have its realism.

While for listeners who have heard them compared, 5.1/6.1 is spacious and enveloping; however hybrid "PerAmbio 6.1.10" – at a "cost" of one decoder and 4 or 5 speakers – transports the imagination. For some, given the choice, these improvements will justify their added cost; for others, they might not. Most say that the listening area is large – almost anywhere within the speaker sphere except near a speaker – and is enjoyable-sounding even *outside* the Ambisonic array! At the Ambiophonic focus, the accurate front stage image compels a "belief" that *you are there.*

Fig.11 - Multi-format control room monitoring (showing 7 of 18 speakers +2 subwoofers) is switchable between PerAmbio 6.1.10 and ITU 5.1/6.1 (no need to change listening position as in Fig.1). The common "Sweet spot" accommodates two within the control room's reflection free zone (RFZ). In addition to Periphonic 3D surround demonstrations, this facility produces compatible 5.1 and PanAmbio 4.1 mixes, including evaluation CDs available to audio engineers at www.filmaker.com.

Listening results – transformed ITU 5.1/6.1

During PerAmbio 6.1.10+ recording or post-production, the recording engineer has chosen one of six encoding modes in order to optimize 2D reproduction in ITU 5.1/6.1. This choice has no affect on ultimate 3D full sphere replay, where decoding mirrors the chosen mode. No decoding is required for 2D 5.1/6.1 replay and, from informal listening tests, the results are uncompromised in 2D in terms of perceived quality in comparison to many commercial recordings made or remixed for 2D-only.

In demonstrations at the author's studios using experimental recordings, made both in concert halls and studio, results in standard ITU 6.1 (discrete Surround-Center speaker) are gratifying. Played on common studio and consumer speakers (see Fig.11), several reported observations follow, first when played in ITU 6.1, then when the recordings are reconstituted in 3D. The test recordings are simply excerpts with no EQ, level compression, or reverberation added:

Excerpt #1: Pipe organ with Trumpet rank directly overhead and Antiphonal division directly back

This excerpt is a staple of the author's PanAmbio 4.1 demonstrations, where the 32 foot pedal (down to 16Hz) is redirected to the ".1" LFE channel. While giving a significant improvement over stereo in a more spacious, enveloping, and integrated room impression front-to-back, 6.1 also reveals that the antiphonal division is behind you. The Trumpet that was overhead is merely all-around (cf. 3D below).

Excerpt #2: Harp concerto with small orchestra (ten players) in 500 seat hall – microphone array at greater than critical distance, raised and tilted –30°.

As the author has come to expect, initially this recording was observed by another recording engineer who was present during the session as "over-ambient, seeming to cause some loss of front focus" unless the three surround speakers were reduced –8dB and the room impression was "correct." After a week's time, he found the ambience "borderline too much" unless reduced just –3dB. He rationalized it was his long-time conditioning to 2-speaker stereo. (See different reaction in 3D below.)

Excerpt #3: Guitar concerto with small orchestra (ten players) in 500 seat hall – array at less than critical distance, at player level and not tilted (horizontal).

This example is usually played in informal listening tests beginning in 2-channel stereo, then switching mid-excerpt to 5.1/6.1. The reaction is nearly universal – even "untrained" listeners notice an obvious difference, describing it as "sound wrapping around" them or words similar. (A stronger reaction is universally reported in 3D, below.)

Excerpt #4: Oboe, piano, & trumpet in 500m3 studio with array at critical distance and level with players.

This experimental recording, played first in 2-speaker stereo, was received by the musicians who performed it as "acceptable" as a demo for promoting their ensemble. However, in 6.1 it garnered more interest, the reported effect being of replacing the listening room with the studio where they recorded. They became interested in what was required for a home theater system to play 2D surround. (See more dramatic reaction below in PerAmbio 3D.)

PerAmbio 6.1.10+ 3D informal listening results

The same recordings were then played in PerAmbio 6.1.10+ using the minimum layout of 10 speakers, plus two subwoofers. Reconstituted from dual-format recordings, the listening results in 3D are identical with the original PerAmbio 6.1.10+ replay – i.e. the system is lossless – so priceless musical performances recorded using PerAmbio 3D are "future-proof" when released in this 2D/3D system.

PerAmbio 6.1.10+ uses 10 or more speakers to envelop the listener in natural sound. The author and his colleague Angelo Farina have demonstrated this hybrid layout in Parma, Italy and Bethlehem, PA USA (Fig.11) using modest speakers for ambience. For nearly all 3D listeners, hearing sound from overhead and underfoot from a recording is astonishing, as though the ceiling, floor, as well as walls have been yanked away! Musicians express that PerAmbio is just what they heard during in the concert hall or studio and so is just like "normal hearing." The leap from 2D 5.1/6.1 to PerAmbio 3D reproduction of otherwise identical recordings is like going from stereo to 5.1.

Again, the 3D recordings have no EQ, level compression, or artificial ambience added. Engineers involved in the sessions and listening tests say the natural impression makes these conventions less needed. They also remark that PerAmbio seems to obviate most need for spot and room microphones, along with their complexity and possibly conflicting imaging. Here are some of the observations (compare with the 2D comments above):

Excerpt #1: Pipe organ with Trumpet rank directly overhead and Antiphonal division directly back

While giving a "vague" sense in 2D (above) that the antiphonal division was behind, in 3D it is clearly in back. Surprisingly, the ceiling-mounted Trumpet is now directly overhead! The organist – a Julliard-trained and experienced recording artist – upon hearing the recording went visibly slack-jawed in wide-eyed amazement. (He has expressed interest in recording an album in PerAmbio 3D.)

Excerpt #2: Harp concerto with small orchestra (ten players) in 500 seat hall – microphone array at greater than critical distance, raised and tilted –30°.

Paradoxically, this recording, observed by another recording engineer as overly ambient in 2D, seemed less so once height cues were added in 3D. He rationalized that the more "natural" the reproduction (3D), the more acceptable surround is. One conductor who participated in listening tests expressed the strong desire to have a system for his own home theater – and pressed the author for an estimate of "when?"

Excerpt #3: Guitar concerto with small orchestra (ten players) in 500 seat hall – array at less than critical distance, at player level and not tilted (horizontal).

This example is demonstrated in informal listening tests beginning in 2-channel stereo, then switching mid-excerpt to PerAmbio 3D. The universal reaction is elevated with respect to 2D above – both "trained" and "untrained" listeners [19] expressing variously that they felt if they closed their eyes, the listening room had been "erased" and replaced by the concert hall. One audio engineer who was present during the recording called it "a sound holograph."

Excerpt #4: Oboe, piano, & trumpet in 500m3 studio with array at critical distance and level with players.

While hearing this recording in 2-0 stereo the musicians thought it just average and in 5.1/6.1 a bit more enthusiastically as sounding "around you" (see above), now in PerAmbio 3D, they discussed with great enthusiasm and at length how natural and uncolored their instruments sounded. The oboist immediately exclaimed "Finally, it's my sound after all these years." (The author takes this to mean "This replicates the sound I was making while performing.") For their leader, minor "mistakes [in performance] now seemed trivial," transcended by having captured the underlying feeling of their performance that was now "more intimately conveyed," having "engulfed" them.

ITU 5.1 recordings compatible with PerAmbio

The PerAmbio 6.1.10+ layout enables users to enjoy legacy 5.1 recordings and movies. By simply moving back in the listening space by 26% of the speaker diameter, the user is able to hear ITU 5.1 with speakers at the proper angles specified in the standard, as illustrated in Fig.12. Side speakers may be used for a form of 7.1 listening, as shown in Fig.13.

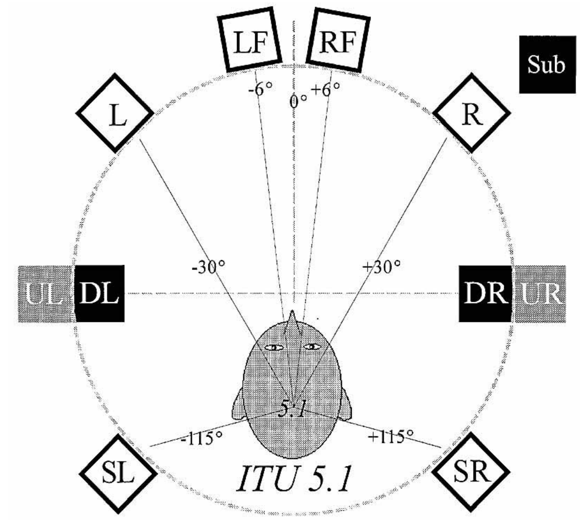

Fig.12. For 5.1, listeners sit back 26% of the speaker diameter, where angles are at ITU standards (requires DSP changes in levels/delays).

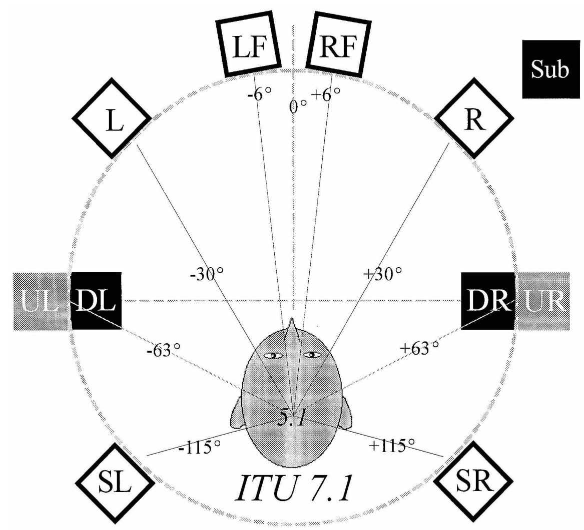

Fig.13. The four side speakers provide for a form of 7.1 listening.

Stereo recordings compatible with PerAmbio

PerAmbio 6.1.10+ embraces Ambiophonics championed by Glasgal [13,14] whereby 2-channel (stereo) acquires 3D ambience by convolution with the 3D impulse responses of concert halls – in MCN terms 2.0.10+. In essence, it is unnecessary to waste mics, channels and bandwidth to "record the hall," when its acoustic imprint is reproducible given its IR.

Reproduced Ambiophonically, many existing stereo releases sound better because Ambio solves problems inherent to 2-speaker stereo of coloration, pinna confusion, and toggling to the nearer speaker for center images. These "errors" are not necessarily endemic to the recording – they are mainly a function of the two 60°-spaced speakers upon replay – and can be eliminated by HRTF-based Ambiophonics. Refer to www.ambiophonics.org.

Additionally, in 2-speaker stereo, all sounds including ambience come from the front 60° between the speakers – only 1/6 of the sonic panorama. Since many stereo recordings have been made acoustically somewhat "dry" to reduce muddling the front stage with ambience, they are ripe for convolution that both restores natural ambience and places it all around – including height if reproduction in 3D. For serious listening, an ultimate sonic experience may outweigh the inconvenience of needing to be in the "sweet spot" – implying only one or two listeners. Single-person listening such as auto sound, PC-based gaming and VR, and audio-centric amusement rides are other viable fixed-position applications where the highest priority is an uncompromised aural experience.

WORK IN PERIPHONIC REPRODUCTION

6.1.10+ and 2.0.10 (Ambio with 8-channel convolved surround) are relatively economical yet effective means of full 3D periphonic reproduction. Objectives for further work include:

- Optimize 3D reproduction alternative(s) for both home and studio;
- Develop DSP for crosstalk cancellation and ambience using 3D hall impulse responses;
- Develop automobile PanAmbio reproduction (demonstrated by the author and his colleagues at 111th AES in New York, Dec 2001);
- Encourage implementation in distribution of 6 full-range audio channels in any medium;
- Develop a main microphone array that is more camera-friendly e.g. for live event broadcasts;
- Introduce artists, producers, and engineers to PerAmbio to "add a dimension" to their creativity;
- First PanAmbio 4.1 (4.1.4) broadcast, Webcast, and DVD (-V, -A, SACD, or DTS-CD) release;
- First PerAmbio 6.1 (6.1.10+) broadcast, Webcast, and DVD (-V, -A, SACD, or DTS-CD) release.
- ANOVA analysis to rank simultaneously made recordings by trained auditioners [19];
- Make recordings with prototype microphones to test new approaches and explore compatibilities;

Recordings for professional evaluation and education are available (www.filmmaker.com).

CONCLUSIONS

Listening in ITU 5.1 is a more enjoyable experience than 2-speaker stereo and is driving home theater purchases by mainstream consumers. This acceptance obviates the "chicken & egg" dilemma common to any introduction of new entertainment technology – i.e. consumer surround systems for music are already being installed.

If an audience to has the means for surround sound, will they demand 2D surround music recordings of higher quality? Will they want 3D in the future? They might if led by producers and engineers who provide them access to 3D via extensible recordings, delivery standards, and replay systems. The choices available could add value for users, entertainment providers, and equipment manufacturers of dual-format DVD-As, SACDs, DTS-CDs for music-only, movies, virtual reality, amusement rides – and in the future for 3D ATV broadcasts, movies, and Internet Webcasts. Issues include relearning by practitioners, providing access by updating delivery standards, and avoiding onerous complexity for users – although some will feel that "10 speakers must be twice as good as 5!"

The hybrid approach PerAmbio 6.1.10+ described offers backward and forward compatibility, playing with no special decoding on standard 5.1/6.1 horizontal 2D surround systems today and with a decoder on full periphonic 3D systems tomorrow.

ACKNOWLEDGEMENTS

The author is grateful to Ralph Glasgal for his inspiration and wisdom and to the Ambiophonics Institute for support of the author's work. Additional support is by the University of Parma and colleagues Angelo Farina, Enrico Armelloni, and Anders Torger; Jörg Wuttke and Christian Langen of Schoeps GMBH, and Buzz Turner and Scott Boland of Redding Audio. Geoff Houser and Brad Frikkers fabricated microphone prototypes; Howard Moscovitz and Craig Greenwood assisted with recording; hundreds of musicians performed for the evaluation CDs; and DTS encoded them. Trademarks are those of their owners: DTS of Digital Theater Systems, Dolby Digital of Dolby Laboratories, SACD of Philips and Sony, IMAX of IMAX Corp., Soundfield of Soundfield.

Appendix A: Requirements for PerAmbio 3D

To fully realize 3D playback requires stringent listening acoustics [20]. Distribution formats can be DTS-CD, DVD-V, DVD-A, SACD, or multi-channel broadcast using AAC of the MPEG4 standard.

Requirements for 6.1.10 playback are:

- Acoustically treated room with RT less than recording venue and symmetrical layout of the speaker sphere with listeners seated at the center at less than the acoustic room radius;
- Two high quality speakers (especially identical phase response) FL & FR plus eight satellite-grade for total of 10 speakers plus subwoofer(s);
- Universal DVD/CD player or ATV receiver with 6 full range analog channels or DTS bitstream;
- A "smart" decoder that detects a flag in metadata to select automatically the transformation mode, obviating user inconvenience and error;
- Crosstalk-canceller, ideally based on impulse response of speakers used;
- Bass manager accommodating 10+ speaker feeds;
- 2 higher and 8 lower power amplifiers for total of 10 channels of power amplification (note crosstalk cancellation requires higher than normal power);
- Calibration of channels at listening position within ½dB using an SPL meter and filtered pink noise.

Any number of listeners not near any one speaker can enjoy the sense of enveloping 3D, although only one or two on the median plane will hear accurate front localization. Mixes intended for ITU 5.1 to 7.1 can be enjoyed by simply moving back 26% of the speaker diameter. 2-channel stereo can be enjoyed by feeding the 8 ambience channels convolved with hall impulse response. All 6.1.10+, 5.1, and 2-0 programs will reproduce up to 120° wide front stage localization for one or two listeners on the "ambiophonic plane."

Appendix B: Ambisonics very briefly

In 1985, Gerzon [12] approached wavefield synthesis mathematically using an omni (W) and three coincident bi-directional microphones aimed forward, leftward, and upward (X, Y, & Z). The array can be realized using individual microphones, or packaged as in a SoundField microphone. Four recording channels (B-format, first order) supply a unique hypercardioid signal (see Fig.14) to each of any number of speakers

$$s_{A,E} = 0.5W + \cos A_n \cos E_n \bullet X + \sin A_n \cos E_n \bullet Y + \sin E_n \bullet Z$$

Ambisonics can be created by assignment, say of monaural dialogue or effects using a 3 dimensional pan pot much as 5.1 can by a 2-dimensional one. B-format (4 channel) stems of music or ambient sound effects can be combined to a 4 channel Ambisonic mix.

Ambisonics' theory fails in many situations where wavefronts are still spherical – reverberation in smaller spaces and direct sound closer to sources; farther and direct sound is swamped in ambience.

For more distant plane waves and pan-pot simulation where the pressure channel W is not uncorrelated with velocity components X, Y, & Z, Ambisonics suffers from lack of spaciousness, similar to coincident M-S. Good spatial impression seems to require spaced microphones, such as the Ambiophone (baffled, pinna-less sphere), which when played Ambiophonically imparts wide, accurate front stage localization. Combined, PerAmbio exceeds in 3D envelopment + frontal imaging + spatial impression.

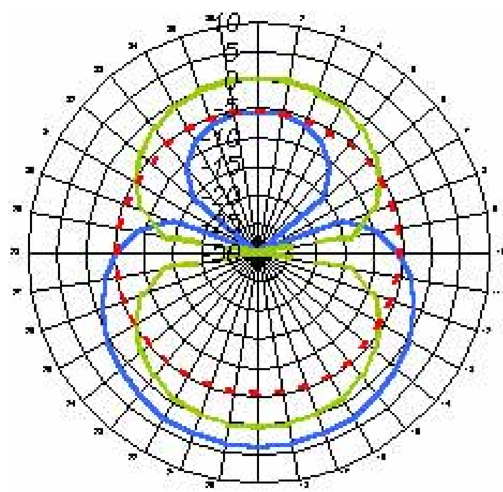

Fig.14. Ambisonics in essence has "aimed" in the direction of each speaker a hypercardioid microphone (directivity=0.6667), derived from coincident omni and bi-directional microphones. The "push-pull" of out of phase information in opposing (isotropic) speaker pairs synthesizes the wavefield for plane waves from distant sources. The experimentally determined minimum for first order 3D Ambisonics is eight speakers, with incremental improvement with 12 and 24 speakers.

REFERENCES

1. ITU-R BS.775.1: "Multichannel Stereophonic Sound System With and Without Accompanying Picture" (Geneva, 1992-4).
2. R Miller, "Compatible PanAmbiophonic 4.1 and PerAmbiophonic 6.1 Surround Sound for Advanced

Television – Beyond ITU 5.1," SMPTE 144th Technical Conf. & Exhibition, Pasadena CA USA, October 2002

3. Fabio Bozzoli, Enrico Armelloni, Emanuele Ugolotti, Angelo Farina, "Effects of the Background Noise on the Perceived Quality of Car Audio Systems," Proceedings of AES 112th International Conv., Munich, Germany 2002.
4. R Miller, "Contrasting ITU 5.1 and Panor-ambiophonic 4.1 Surround Sound Recording Using OCT and Sphere Microphones," Proceedings of AES 112th International Convention, Munich, Germany 2002, preprint #5577.
5. F. Rumsey, T McCormick, "Sound and Recording," Focal Press, ISBN 0-240-51680-X, pp 382-388
6. G. Theile, "Natural 5.1 Music Recording Based on Psychoacoustic Principles," Proceedings of the AES 19th International Conf., Schloss Elmau, Germany, June 2001.
7. R. Miller, "PerAmbiolating 360°: Panorambiophonic 4.0" multichannel evaluation DTS-CD, "PerAmbiolating 360°: Companion ITU 5.0," multichannel evaluation DTS-CD, with print insert, "Surround" pages, www.filmaker.com.
8. J. Wuttke, "General Considerations on Audio Multichannel Recording," Proceedings of AES 19th International Conf., Schloss Elmau, Germany, 2001.
9. H. Wittek, O. Neumann and M. Schäffler, C. Millet, "Studies on Main and Room Microphone Optimization," Proceedings of AES 19th International Conf., Schloss Elmau, Germany, 2001.
10. D. Griesinger, "The Psychoacoustics of Listening Area, Depth, and Envelopment in Surround Recordings and Their Relationship to Microphone Techniques," Proceedings of AES 19th International Conf., Schloss Elmau, Germany, 2001.
11. G. Theile, "On the Naturalness of Two-Channel Stereo Sound." J. Audio Eng. Society, Oct. 1991.
12. M. Gerzon, "Ambisonics in Multichannel Broadcasting and Video," JAES Vol. 33 No.11, 1985, pp 859 – 871T.
13. R. Glasgal, "Ambiophonics: Achieving Physiological Realism in Music Recording and Reproduction," Proceedings of AES 111th Convention, preprint 5426.
14. Farina, R. Glasgal, E. Armelloni, A. Torger, "Ambiophonic Principles for the Recording and Reproduction of Surround Sound for Music," Proceedings of AES 19th International Conf., Schloss Elmau, Germany, 2001.
15. R. Miller, "Ambiophonic Demonstration: AES Bavaria 2001" 2-chan. CD with print insert, and PS online Ambio demonstration, "Surround" pages, www.filmaker.com.
16. M. Gerzon, "Optimal Reproduction Matrices for Multispeaker Stereo," AES 91st Conv. 1991, New York, preprint 3180.
17. T. Holman, "5.1 Up and Running," Focal Press, ISBN 0-240-80383-3.
18. ATSC Document A/52A "ATSC standard: Digital Audio Compression (AC-3), Rev. A" 8/21/2001, www.atsc.org/
19. B. Blesser, "An Interdisciplinary Integration of Reverberation," AES 111th Conv. NYC 2001, preprint 5468 as presented.
20. "Multichannel surround sound systems and operations" – AES Technical Document ESTD1001.0.01-05

Author

Robin Miller is a member of AES and SMPTE, musician and orchestrator, broadcast engineer, teacher, and 53 award-winning producer. As an audio engineer, he has more than 40 years experience in music recording and mixing 300 films and television specials plus countless live broadcasts.