Ambiophonics Terminology

The language of stereo-and-5.1-compatible loudspeaker-binaural

Robert E. (Robin) Miller III, AES SMPTE BSEE, ©2009
Filmaker Technology ● www.filmaker.com

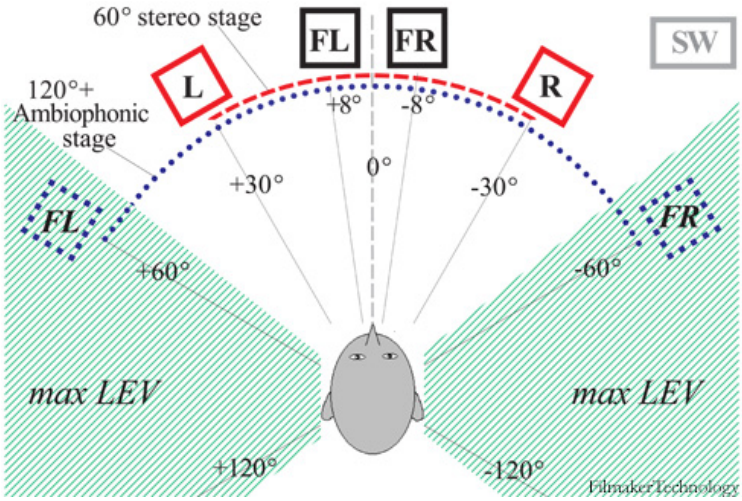

Fig 1: **Ambiophonics'** speakers FL & FR convey a stage equaling the original recording angle extending to virtual *FL* & *FR*, uncolored center voices, and maximal listener envelopment (LEV), reaching within the regions shown, all in contrast to that of conventional stereo speakers L & R.

A critical method of sound reproduction for entertainment (music, movies, games, rides), *Ambiophonics* has both its own lexicon, and shares terminology with related methods which are explored elsewhere. This paper provides useful definitions of unique Ambiophonics terms deemed essential for understanding by audiophiles and audio engineers, presented in the context of conventional *stereophonic* and *5.1* stereo-surround sound.

In sum, the root name *Ambiophonics* is given to several related audio technologies for improving *imaging/localization, spatiality* (envelopment, immersion), and *timbre (tone color)* of sound reproduction of music, movies, and games. Championed by Glasgal, Farina, and the author, Ambiophonics has been in development since 1995, with levels of perceived "reality" ranging from front-only presentation to life-like 3D, including:

a) Reproduction of 2-channel recordings intended for *stereo* (60° front-only) that extends the perceived reproduced front stage width to 120° or more (with or without the addition of *convolved* ambience) – known as *Ambiophonics (Fig1);*

b) Enveloping 360° (2D, horizontal plane) reproduction of multi-channel surround sound recordings, e.g. ITU-R775 standard 5.1 or 4.0 (implying directly-recorded ambience or sources around sides/back) – known as *PanAmbio (Fig2);*

c) Immersive periphonic 3D reproduction (with height, up to the full sphere of human hearing, directly recorded) using special 6-channel recordings – known as *PerAmbio, TriAmbio,* and *High Sonic Definition (HSD-3D, Fig3)*.*

Note: Not related to "Ambiophony," introduced in 1960 by Parkin as a method for electronically simulating acoustic reflections in concert halls that are lacking in reverberation, especially for symphonic music.

The Ambiophonics "family"… (terms in italics defined below)

Ambiophonics: (Fig1) loudspeaker-*binaural* reproduction of 2-channel LP/CD/MP3/SACD, etc., which processes a *stereo* pair of signals and presents them to a centered listener by an *Ambiodipole* (closely-spaced speaker pair) in front. Optionally, image width may be augmented by duplicating the signal pair by a rear *Ambiodipole*. Additionally, to restore a natural level of ambience that has been deliberately reduced in commercial stereo releases, two or more *Ambiostats* (surround speakers) and hall *impulse response convolution DSP* provide 2D or even 3D surround *reverberation* based on two media channels (including most legacy stereo recordings on CD, LP, games or broadcasts).

Ambio*:* colloquial usage, as *stereo* is for *stereophonics.*

Panorambiophonics (PanAmbio 4.0/4.1, Fig2 ***):*** Uses independent front and back *Ambiodipole* pairs of speakers to reproduce 2D surround recordings from 4.0/5.1 media. Records and replays two Ambiophonics stages – a front stage typically of direct and frontal reflected sounds, and a stage behind the listener for reflected and ambience sounds, along with any direct sources around the sides and in back. In contrast to *5.1*, PanAmbio is isotropic (equal in all directions) in the horizontal plane, it renders accurate 360° localization ±5°, and therefore avoids 5.1's spectral tearing of *phantom* images at each side.

Perambiophonics (PerAmbio 6.0/6.1): Efforts toward 3D reproduction which is spherically isotropic, as with live, natural human hearing. *TriAmbio* adds a third, elevated Ambiodipole for reproducing height using special 6.0/6.1 recordings. *High Sonic Definition (HSD-3D)** achieves full-sphere *immersion* using 6-channel recordings and 10 or more loudspeakers (Fig3).

Ambiophonics' hardware…

Ambiopole: A loudspeaker optimized for use with digital *crosstalk cancellation* processing, having controlled dispersion and phase-consistent performance in any crossover regions.

Ambiodipole: A closely-spaced (~16°) pair of speaker *Ambiopoles* placed in front of and optionally behind listeners.

Ambiophone: two-channel main microphone used to make stereo-compatible two channel recordings for critical Ambiophonic reproduction according to HRTF principles.

Panambiophone: main microphone array used to make stereo-compatible two channel recordings with ambience control, or 5.1-compatible *PanAmbio* surround recordings.*

Ambiostats: surround speakers used to radiate concert-hall *reverberation* by *convolution* from hall impulse responses.

Ambiovolver – PC with DSP software for both crosstalk cancellation and hall ambience *convolution*. Two instances of *crosstalk cancellation* are required for 2D *panambio* surround.

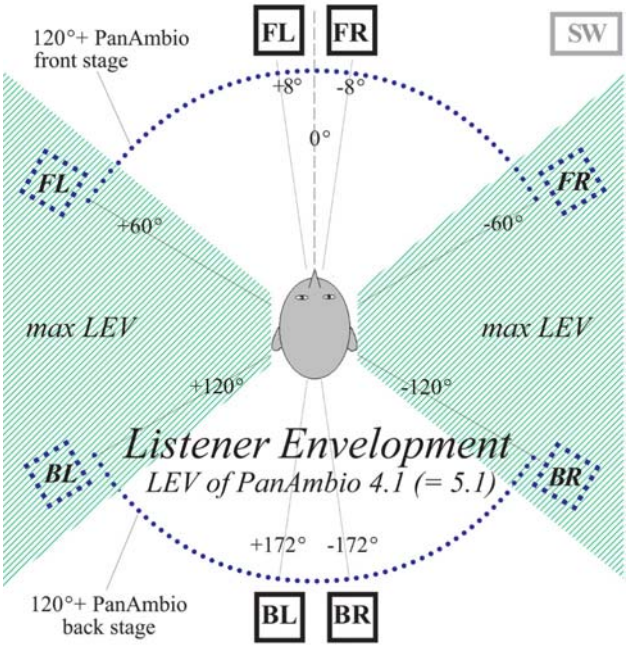

Fig 2: ***PanAmbio*** surround adds back speakers BL & BR, imaging as *BL* & *BR* – four "speakers" (double 5.1/7.1's two) within regions where listener envelopment (LEV) is maximum. (Play 5.1 in PanAmbio by setting the player to "no center" to mix the C-channel to FL and FR.)

Related audio formats & scientific terminology…

Stereo, stereophonic – 2-channel audio reproduction attributable to work in the 1930s at Bell Labs and EMI.

"5.1" – international surround standard ITU-R775 for 5 to 7 speaker channels plus Low Frequency Effect (LFE, aka "0.1") for home theaters, evolved from multi-channel cinema sound since the 1939 movie "Fantasia." *PanAmbio* is 5.1-compatible; *PerAmbio, TriAmbio,* and *HSD-3D* are 6.1/7.1-compatible.

Binaural audio – lifelike reproduction recorded with a surrogate "dummy head" with microphones at its ear positions, then played using headphones. When reproduced directly into the listener's ear canals such as with intra-aural ear receivers ("ear-buds"), the dummy head has generic *pinna* that might not work with an individual listener. *Pinna*-less microphones such as the P*anAmbiophone*, reproduced over loudspeakers, involve listeners' own *pinna*, and so satisfy life-like perception for all.

ILD, ITD – interaural level & time (phase) difference in binaural ear signals from a sound source, registered by auditory sub-brain structures, which contribute to the perceptual qualities of *localization, spatiality*, *timbre*. About equally splitting the audible spectrum, *ITD* is the controlling mechanism for sounds with frequencies in the five octaves below 700Hz and *ILD* for the five octaves above 700Hz (depends on listener's head size).

Localization (imaging) – the ability of humans (using *binaural* cues) to identify within a sphere of possible arrival angles the direction of origin of sounds, both those that arrive along a line from the source (direct sounds) or of temporally distinguishable reflected (indirect) sounds. Acuity within an ellipse in front of the listener is about ±1° horizontally and ±10° vertically. Evolved from survival needs (charging tiger, etc.), localization is an important basis for perceiving *spatiality* and, due to directionally-dependent *pinna* filtering, integrating tone color (timbre) that together figure in the enjoyment by our sonic sense of live or recorded speech, music, environmental sounds.

Spatiality – perception of the space containing both the listener and a source of sound, via recording or live hearing. The sense has been integrated in the auditory cortex of the brain from neural signals processed in sub-cortex brain structures as *ILD* and *ITD* of ear signals that have picked-up direct and indirect (reflected) sounds generated in the space. Even with eyes closed, humans perceive a 3-dimensional "image" of the space within a potentially infinite spherical frame of reference.

Timbre (tone color) – character of a sound, as determined by qualities of an instrument, technique of a performer, relative position and properties of the receiver (listener or microphone), and acoustics of the space containing all these. The holy grail of hearing enjoyment, timbre describes each instrument or voice along with its extension, the reflective space containing it (see *spatiality*). Physically tone color is a time-varying composite of a sound, integrated in the conscious brain from dynamically combining the direct sound and multiple, variably delayed arrivals of reflected after-sounds from a sphere of directions (see *localization*). Each arrival is filtered by pinna and, as it arrives at the other ear, is subject to ILD and ITD comparison.

LEV (listener envelopment) – a sense of being surrounded in 2-dimensions (or better yet *immersed* in the full-sphere 3D of human hearing); essential to human perception of the space containing the listener and sources of sound. This quality of *spatiality* is largely missing in conventional stereo reproduction, but is much improved using Ambiophonics, where the ambience sounds especially may be localized within direction-dependent regions of maximum LEV (see Fig1). PanAmbio surround is capable of regenerating more LEV than 5.1 surround (see Fig2).

Immersion – by extension along the vertical axis, a sense of being engulfed in sounds in 3-dimensions – the full-sphere of human hearing; essential to life-like perception of the 3-space containing the listener and sources of sound. This ultimate *spatiality* is noticeably lacking in conventional 2D surround reproduction (*5.1* or *PanAmbio*), but by degrees approaches "life-like" perception using 3D ambience convolution or directly recording/reproducing 3-space, as with *PerAmbio, TriAmbio,* and *High Sonic Definition (HSD-3D)*, where sounds may be localized at any spherical angle (see Fig3).

Spatial resolution – degree, in contrast to either front-only stereo or to 5.1, to which 2D *ambio, panambio* surround sound, or full-sphere 3D preserve individual voices amid the ensemble (in the frequency domain), and do not lose or muddle the stream of "auditory events" and silent "air" (in the time domain), but allow listeners to perceive separation of voices and events, as in live hearing. Heroics by stereo mixers to maintain audible "layers" (singer discernable from instruments, etc.) are needed far less given the greater spatial resolution of surround sound.

Reverberation – Physically, the acoustical *impulse response* of a reflective room or space. The more reflective or larger the space, the longer lasting the reverberation. Example: a simple tom-tom is a physical *impulse;* timbre is that *impulse* plus the *impulse responses* of the room at each ear. A single bass viol performing with vibrato (varying its pitch) becomes a chorus with itself, arriving at the receiver simultaneously at different frequencies. A Gregorian chanter can "sing" a chord!

"Reverb" – jargon connoting artificial reverberation, sometimes acceptable, but often phony-sounding. Realistic results are either directly recorded in a reverberant space, or carefully convolved using a 2D or 3D hall *impulse response*.

Anechoic – devoid of reflected after-sounds; the absence of *spatiality* cues by which a listener creates in his/her mind an "image" of the space containing the listener and a source of sound; direct sounds alone. Considered an unusual condition, a deeply snow-covered field imperfectly approaches anechoic conditions. (Useful for acoustic testing, an "anechoic chamber" is not enjoyable for humans to occupy for any length of time.)

Impulse; Impulse Response (IR) – momentary non-zero event (e.g. hand-clap, gun-shot, balloon burst, or electronic "Dirac" pulse); which characterizes electro-acoustic systems – its Impulse Response ("IR", e.g. *reverberation*). Each sample of a digital recording can be considered a Dirac impulse, which spawns its own IR by *convolution* (e.g. hall ambience by DSP).

Convolution – using a digital recording and applying *DSP* signal processing to "imprint" an *impulse response* on a source signal. For example, multiplying and adding to the samples of an anechoic instrument or voice the acoustical impulse response of a hall as though the source had been recorded in that hall.

DSP – digital signal processing. Used in *Ambiophonics* for *crosstalk cancellation* and optional ambience *convolution*. DSP for "*XTC*" include single delayed inversion (Wareing), inverse *impulse response* (Farina), band-splitting (Choueiri), and *RACE* (Miller). DSP for generating ambience convolves 2-channel stereo recordings with 2D or 3D hall *impulse responses*.

Pinnae; pinna confusion – the outer ear, whose fleshy convolutions imprint filtration depending upon the 3D angle of arrival of a sound. Together with ear signal difference *ILD* and *ITD*, *pinna* filtration encodes direction for the conscious brain to perceive *localization, spatiality, & tone color. Pinna confusion* results when reproduction systems (notably conventional *stereo* and 5.1 surround) create phantom images the brain interprets as coming from a fuzzily-perceived direction despite being *pinna*-encoded as coming from their actual direction, e.g. a speaker.

Phantom image – *localization* of a sound, the perceived direction of which is not real but the result of "conspiring" cues, such as two loudspeakers playing the same signal, perceived as coming from a point between the speakers. In conventional *stereo* reproduction, the common practice of "panning" a single microphone of a soloist equally to both channels/speakers is determined by the brain as coming from the center, although not convincingly (due to *pinna confusion*) and exhibiting a raspy *comb-filtering*. Phantom images slightly to one side of center are drawn in non-linear fashion toward one speaker or the other.

Comb filtering – distortion of a sound's spectral character (frequency response) due to wave cancellation alternating with reinforcement when the identical signal arrives simultaneously from two points (e.g. speakers). Comb filtering due to *crosstalk* in stereo affects the frequency response at each ear of important central voices (e.g. soloist, dialogue), and unfortunately causes mixing engineers to compensate by boosting equalization of the isolated central sound around 2kHz that permanently scars the recording for headphone or *Ambiophonic* replay. (The center "C" channel/loudspeaker solves this condition for 5.1 surround.)

Crosstalk – unintended leakage of reproduced sound from any speaker to the ear on the opposite side. Avoided by using headphones, speaker crosstalk at either ear is accompanied, for important central voices, by duplicate sounds, delayed in time, that interfere with earlier sound from the speaker on the same side, resulting in raspy-sounding *comb-filtering* and *pinna-confusion* as to localizing the sound and coloring its timbre. Crosstalk is the principal impetus for *Ambiophonics* solutions.

Crosstalk cancellation (XTC, CTX) – *Ambiophonics* solutions address the problem of *crosstalk:* if it can't be avoided it in the first place, cancel it. Approaches to XTC all attempt to segregate speaker sounds to the intended ear. Methods over the years include a physical barrier projecting from the listener's nose, adding speakers recessed and therefore delayed in time, and digital signal processing to introduce cancellation signals into the audio path, as with *Ambiophonics' RACE.*

RACE – (Recursive Ambiophonic Crosstalk Eliminator) method developed for creating using DSP properly timed and attenuated recursive *crosstalk cancellation* signals, with user control of the tradeoff between stage width and coloration artifacts. Crosstalk is cancelled acoustically at listeners' ears.

MIDI – musical instrument digital interface, used in PC implementations of *RACE* for user control of parameters.

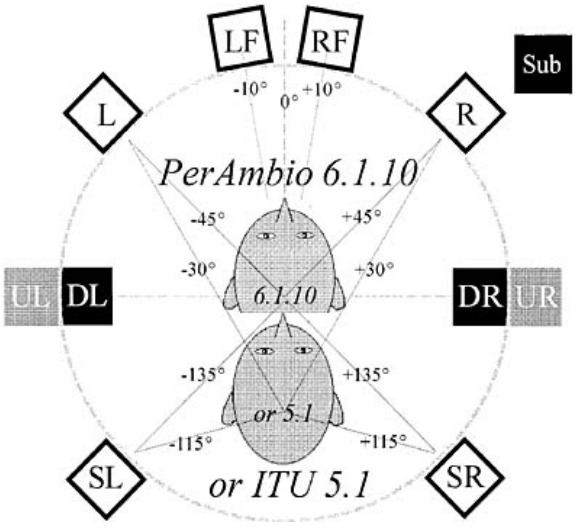

Fig 3: ***PerAmbio 6-channel*** (aka High Sonic Definition 3D) combines an Ambiophonics front stage, for the accuracy in front where humans need it, with a minimum of 8 speakers for directly recorded ambience. Specially-made recordings result in life-like full-sphere reproduction.*

Ambiophonics contributors…

Ralph Glasgal – creator through the Ambiophonics Institute of the name *Ambiophonics,* and prime mover of its development, progressing from a physical barrier through DSP for *crosstalk cancellation* and ambience *convolution.*

Robert E. (Robin) Miller III – engineer/pres. of Filmaker Technology who developed RACE DSP, the *Panambiophone,* and other *Ambiophonics* innovations and related demonstration 2D and 3D recordings, scientific papers, and website content.

Angelo Farina – *Ambiophonics* researcher at the University of Parma, Italy, who used measured inverse impulse responses of speaker pairs using a dummy head to convolve XTC, and of world-renown halls for ambience *convolution*.

Enrico Armelloni –student of Farina at the University of Parma, Italy, whose doctoral thesis was developed while an intern at Filmaker Technology during development of HSD-3D.

Edgar Choueiri – Princeton space physicist and *XTC* developer who used sophisticated band-splitting to optimize (minimize) coloration due to *crosstalk cancellation* impulses.

David Wareing – early *XTC* developer who modeled *crosstalk cancellation* as a single correction impulse based on the speed of sound and the geometry of the average head.

Anders Torger – Swedish developer of "AlmusVCU" *Ambiovolver* software (*XTC*, ambience *convolution*, bass management, speaker position timing) running on Linux PCs.

"The shoulders of giants" applies, upon which stand those above who have advanced the art termed Ambiophonics. While this list would eclipse this writing, it would include the work of the ISVR, Jerry Bauck, Günther Theile, and the auturs of commercial implementations by Carver and Yamaha, etc.

For more info, see www.ambiophonics.org.

---

*Internationally recognized engineer and Peabody award-winning film producer **Robin Miller** has presented papers and demonstrations on 2D and 3D audio to the Audio Engineering Society, Society of Motion Picture & Television Engineers, Acoustical Soc. of America, Canadian Acoustical Assn., and German Tonmeisters. His company, Filmaker Technology, does applied science research, systems design & integration, surround recording, and has patented a system of full-sphere 3D recording & reproduction – www.filmaker.com*

*Subject of *U.S.Patent #7,558,393*