Expand your *space*; Put *color* in your life? – Robin Miller AES/SMPTE/BSEE/Filmaker Technology

presented to audio engineers at AES2007, Parma, Italy – Recording Workshop #2, November 25, 2007 *rev.10/7/2010*

Human perception of sources of sound, as characterized by their tone color (timbre) and the space they are in (spatiality), is the end result of a chain of events, described in the five familiar steps that follow:

Step 1. Energy from striking blows (drumsticks, piano hammers), friction (violin bows), breath (voices, horns), guitar amps, etc. causes objects to resonate, usually at multiple frequencies and amplitudes that vary over time;

Step 2. Vibrating sources cause the air around them to pulsate, radiating waves of sound that propagate away from the source, varying from spherical (omni-directional) to favoring one or two directions; [Quite understandable so far for most recording engineers, but read on to be engaged in less understood concepts.]

(Step 2.5, when applicable, is the stage where audio recording and reproduction is "imposed;")

Step 3. Waves of air pressure variations travel different distances, undergo environmental changes, and arrive at individual listeners from all directions, either directly from sources or indirectly from reflections[1];

Step 4. A listener's ears gather and respond to these air pressure variations that impinge upon them from all directions in 3-space, and that at any instant can be very different at each ear;

Step 5. A listeners' brain (or an on-stage performer's brain) continuously evaluates these ear signals, processes time, level, and spectral differences between them, and does "table-lookups" from remembered experiences to map them into *auditory events* located in a *perceptual sphere,* with the listener's consciousness at the center.

Spoiling things by degrees, step 2.5 is where audio reproduction occurs. How can we "keep it real?"

Keeping sound reproduction real; what could possibly go wrong?

When audio recording and reproduction are inserted as Step 2.5, the result by step 5 will have deteriorated by degrees. Pertaining to step 1, the excitation dynamics and frequency ranges are often much greater than the capabilities of reproduction. (Of course, some audio manufacturers purposely limit expense for market driven reasons, such discounting a mid-bass octave between 90~200Hz in order to market tiny satellite speakers.)

In step 2, it is often assumed that sound sources are only those we *see* – those within 60° in front (stereo), or a horizontal circle (ITU 5.1) – and that more directionally-dependent reflected sound, being unseen, doesn't exist or matter much. This is a 2-dimensional view, not 3-dimensional life-like hearing.[2]

In step 3, it is similarly assumed that loudspeakers only need to replicate the 60° pie-shape of stereo or the 2-dimensional circle of 5.1 – ignoring the rest of the 3-dimensional sound sphere[3] – and that the acoustically modifying properties of the listening room the loud speakers and listener are in don't exist or matter.[4]

In step 4, it is often assumed that monaural signals panned among two or more speakers (at the price of harsh comb-filtering) will create realistic inter-aural difference. Or that spatiality can be added by simulating "reverb," or (more successfully) by convolving with room impulse responses (IR) even though measured at only two assumed source locations, at stage left and stage right, when every point on stage has its own 3-D room IR.[5]

In step 5, it again is often assumed that a listener's processed, now conscious "soundscape map" coincides with vision, and therefore is narrowly presented within an arc including only the players on-stage, and not a full sphere with dimensions including elevation as well as azimuth and radial distance.

Inserting step 2.5, no wonder by step 5 – every listener's brain – hope is gone of truly compelling audio enjoyment. How then might audio reproduction achieve realistic spatiality & timbre? And monetize its value?

---

[1] Everywhere beyond a room's critical radius the energy of indirect (reflected) sounds exceeds that of direct sound.
[2] 2-speaker stereo crosstalk creates bogus early reflections, and the phantom center harshly comb-filters soloists.
[3] Hearing is more acute frontally and less vertically, but still the full sound sphere is integrated in the brain.
[4] Marketers may tout 2D reproduction as "3D," but vertical cues are false, due entirely to the listening room.
[5] What of the other 98 voices, each with a unique 3D room IR? Or movie/game effects moving in 3-space!?

Implicit in the issues above are clues that audio engineers as well as marketers need to address in order to satisfy increasing demand for greater realism in sound reproduction. To step backward from conventional wisdom to the point before each shortcoming occurs, then proceed forward again with greater knowledge. The objective for step 2.5 (recording & replaying sound) is to capture precisely for each ear the sphere of sounds, seen *and unseen,* and to preserve these signals and their perceived direction until they arrive at listeners' ears.

Back to basics, then forward to life-like audio

Several technologies have been developed by Filmaker Technology by going back to basics, then moving forward again: 1) Recursive Ambiophonic Crosstalk Elimination, or RACE); 2) Pan-Ambiophonic surround (Ambiophonics both front & back); 3) High Sonic Definition 3D (hybrid Ambiophonics plus Ambisonics); and 4) Binaural Bass Management (stereo bass); 5) The PanAmbiophone microphone array ("zooming" up to 3D). Details of these solutions and their application are at www.filmaker.com in scientific papers by the author.

Among other qualities, the approaches mentioned address *listener envelopment* (LEV) in 2-dimensional systems (stereo and 5.1 surround), or *immersion* in the deeper waters of full-sphere 3D. As ever, each listener's unique head-related transfer function (HRTF) will be encountered – generic approaches have been proved to work for only 25% of the population. In the model for life-like realism, listening in 3-space, whether live or recorded, each pinna filters every sound as it arrives in time from a different direction, including from above and below. If these directions are limited, such as reproduction confined to 60° or only in the horizontal plane – the narrow pie-shape of stereo, or even the 360° flat circle of surround 5.1/7.1 – or that is absent any elevation that would complete the sphere of hearing, then pinna-filtering action is correspondingly limited. This *spatial confining* results in lacking or confused pinnae contributions that are necessary for direction-coding of many sonic arrivals, including indirect sounds contributed by the space (early reflected & late reverberant). When all are integrated in the brain, they are perceived as a sound's distance, localization, spatiality, and tone color.

The brain relies on its memory of auditory events, cataloguing its experiences of sonic "signatures" of sources and spaces. Whenever the listener is introduced into new sources and spaces, the brain must adapt to new information until sounds becomes familiar, and thereby, it becomes able to interpret and perceive source distance, timbre, localization, and spatiality. At the onset of this learning timeframe, the new information comes as a surprise, and auditory events are often misperceived. For example, after being acclimated to a live orchestra in a concert hall, Jens Blauert found that suddenly introducing a monaural sound close to subjects and on their median planes (so that their ear signals were highly correlated), causes the subjects to be confused, perceiving the new sound as inside their heads. Since all other recent stimuli were uncorrelated ear signals, typical in listening in acoustic spaces including concert halls and identifiable movie & game settings, the alien correlated signals must be inside the brain! (Not always "wrong," this can be sparingly used for artistic effect.)

Listeners using headphones learn that identical in-phase ear signals, such as conventionally panned-to-center (monaural) vocal tracks, while not naturally occurring – they wouldn't in spatial recordings – are OK as a pretender for front & center, and so are able to resist locating them inside their heads. An opposite "unnatural" response occurs if otherwise identical ear signals are coherent but out of phase, such as when, in relatively reflection-free environs, stereo channels are phase-reversed, and intended phantom center images seem to jump around. In non-reflection-free zones, this conditions is described as "diffuse," "phasey," or "lacking focus." Naturally occurring random-phase of spatial recordings, such as those made by an Ambiophone or other head-spaced main microphone array in acoustic spaces, will not localize in-the head, nor be phasey, but capture uncorrelated early reflections that allow listeners to determine position, distance, and space. (Demonstration recordings in stereo, RACE, and PanAmbio surround are at www.filmaker.com and www.ambiophonics.org.)

Conditioned by a lifetime of hearing experience, the brain expects uncorrelated ear signals (neither totally identical nor totally out of phase). The brain refers to its memory of prior, familiar auditory events, but gathers new information quickly and adapts, such as to different listening spaces and loudspeakers in them. Otherwise, the brain is "surprised" – either confused, or becomes party to an intentional effect, that can be appreciated if judiciously applied. However, the brain is most welcoming of ear signals that are most "natural" – that represent direct and indirect arrivals that have been preserved in direction, and are identifiable and pleasing in timbre and spatiality. Sources so captured and reproduced will be perceived as life-like rendering, in spaces of gripping interest, whether the eyes are closed or open and delivering visual confirmation. – *ROBIN MILLER*