

Evaluating Loudspeakers

Subjectivity using non-sentimental sound-makers

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



Left: Test vase, shells, triangle, “thunder machine,” egg shaker. Right: Monitor-quality loudspeakers compared using recordings v. live sources. (The pair moved toward the center was eliminated.) A high-end audiophile pair was also evaluated. While the informal study was not blind and the participating audio engineers had prior knowledge of the units’ reputations and prices (\$1,800 to \$25,000 per pair), results were surprising.

Budget and aesthetics aside, savvy consumers would prefer that their home loudspeakers perform similarly to professional monitors – the quality category used to base material decisions during recording, mixing, and mastering as to the “sound” of the music, movies, and games that users enjoy. Alas, even if less expensive than high-end audiophile speakers, monitors (actual, not claimed-by-marketing!) are often too utilitarian-looking. However, some manufacturers have equivalents designed for consumer tastes. Whether for consumers or professional audio engineers, some means for intelligent selection is needed.

Accordingly four practicing, professional audio engineers including the author conducted a unique speaker “shoot-out.” Four pairs of monitors were assembled in an acoustically fine control room. They ranged in price from \$1,800 to \$7,000 US per pair. A high-end audiophile loudspeaker costing \$25,000 per pair was auditioned separately. The four were self-powered, fed by a stereo pair of analog signals from a routing switcher. Speakers were clustered so it was not obvious which pair was playing. However instead of auditioning music familiar to the subjects, a special assortment of noise-makers was assembled, recorded, and made portable in a gym bag so they could be conveniently made available during subjective testing, where reproduction could be compared to the live sources. This paper explores the theory, methodology, and surprising results.

“Subjective Testing” by audio professionals or consumers

Typically, shoot-outs are organized informally, with participants bringing favorite source materials. Often musical selections, movie scenes, or gaming environments are drawn from commercial releases that the subjects like, not material they have produced or had first-hand experience in monitoring during production. These after-the-fact evaluations are similar to consumers evaluating audio equipment in the varied (poor?) acoustics of dealers’ showrooms based on a “reference CD” brought to the audition in order to be “objective.” (Certainly their prerogative, it might be sufficient for consumers, in the

absence of standards to guide them, to simply stop searching at any sound they like, but at home their new speakers likely won’t sound the same because the acoustics have changed.¹)

Any scientist would find fault with this methodology. They call “subjective Testing” an oxymoron: because instead of a single variable, a single device under test (DUT), there are many. Interaction and inconsistency reign. And emotion – *sentiment* – gets in the way of meaningful, repeatable results.

What is heard during audio reproduction is the end of a complex chain of events and devices, from the original sources, recording acoustics, recording microphones and technique, recording electronics and storage format, mixing techniques including processing, mastering techniques including more processing, distribution media, users’ reproduction electronics, speakers varying in positioning, and the listening acoustics.

If loudspeakers in a room were perfect reproducers, they would all sound alike – indistinguishably neutral, neither adding nor detracting from life-like sound. In imperfect reality, they mostly sound very different, varying in ability to reproduce all frequencies equally, all volumes equally, and all directions equally. Often costing half or more of a user system’s budget, they compromise these ideals, and color the sound.

Conditioning and emotionality of professionals or consumers

All of the various interplays among all the links in the audio chain are too large a subject for this paper, and have been observed and debated for decades of stereophonic tail-chasing by audio practitioners. Unfortunately, most people have been

¹ Furthermore, the next recording the consumer hears, having been mixed using different speakers and under different acoustic conditions, could establish a different reference point, and therefore sound better to the consumer on another choice of speakers. Such a moving target suggests that, in the extreme, each recording require different speakers. A more consistent approach would have both mixing engineers and consumers selecting speakers similar enough to establish a common reference – a standard “sound” for good results in any case.

subjected to so much questionably recorded material that they are conditioned to imperfection. We've become inured to over-processed, unnatural, and conflicting psycho-acoustic cues. So much so that it redefines "transparent." Paradoxically when we are exposed to naturally life-like sound, it is often perceived in the redefined context of "audio," and we think it is wrong.

Seeking to elevate part of the situation, this informal study attempted to deal with perhaps the most critical hurdle – the fact that different genres of music elicit different emotional reactions in different people. Some like jazz but not rap; heavy metal tastes clash with classical, etc. So a kit of sound-producing instruments was assembled that avoided listener sentimentality. Some are un-identifiable. Their sonic signals test the limits of reproducibility. Not distracted by preconceived notions, subjects could concentrate on perceived differences while comparing live to reproduced sounds where the only the variable was the monitor speakers under test. While not perfect, this resulted in a practical approach to more precisely evaluating the very means of monitoring that in turn leads to better decisions and artistic results in the work of the engineers involved, and by extension, to greater satisfaction among their customers.

Gotta Take a Non-sentimental Journey

Recording engineers have jangled keys or walked around recording venues rattling maracas with a similar goal – exciting transient and tonal sounds, but avoiding emotion in the equation. In academic studies, researchers such as at McGill University statistically analyzed responses of subjects to bells and bamboo chimes. The writer e-mailed fifty fellow engineers around the globe, requesting nominees for the *non-sentimental analogs* kit that were practical and portable, or easily replicable.

Ideal test sounds ought to represent singly, for clearer discernment, individual characteristics similar to those that will be found in vocal, instrumental, or sound effects. They should fully exercise the capabilities of transducers and electronics in the recording chain. And they should be portable enough to be carried in a bag or shipped for use by others in their evaluation sessions, so results can be compared widely. Since the most critical testing begins and ends, as does most recording and reproduction, with analog sound waves, the instruments chosen are all analog wave-makers. (It is acknowledged that electronic instruments are important, but in the absence of the actual instrument, we have no reference for how they sound live.)

With these criteria in mind, along with a budget of \$100 US, the writer assembled an initial kit of five "non-sentimental analogs" in a gym bag. These were recorded in the writer's studio using a spatial microphone technique (i.e. not a panned monaural microphone) so that an uncorrelated two-channel recording was made to avoid comb filtering during speaker-stereo reproduction. Microphones and converters (by Schoeps and Grace Designs) were of high quality. Taken to the shootout for live comparison with speaker replay were both the 2-channel recording in CD form and the portable kit of five sounds:

1. An orchestral triangle (having no harmonically related overtones, so that any integer-multiple overtones would be harmonic distortion produced in recording or reproduction);
2. A bunch of about 20 shells with dried seeds emitting a cackle of transient and low- and mid-range tonal resonances;
3. A "thunder machine" with a dangling spring to excite low-mid fundamental frequencies in the drum membrane;

4. A German earthenware vase 24 inches high and nine inches in diameter that, rapped by a knuckle, emitted a bell-like tone, followed by a tremulous sustain;

5. A common percussionist's plastic egg shaker that makes a lot of high-frequency transients.

No source emitted significant energy below about 200Hz, avoiding variables in the bass region which might have been introduced by subwoofer implementations and room modes.

Shootout: The Last Monitor Standing

A sentimentally-contaminated process serves no productive purpose. For example, one might assume, even before hearing it, that the most exotic (expensive or cool-looking) apparatus would have to be best. Or that if one now hears qualities s/he prefers to any prior hearing of a familiar recording, then the speaker is *magic*. (Tantamount to second-guessing, it simply reveals that whatever it's doing "better" should have been modeled during production.) Or that rather than believing one's ears, that advertisements or marketing talk couldn't possibly be misleading, could they?

So four candidate speaker pairs were obtained, set up avoiding reflections from nearby walls and console surfaces, and had their SPL outputs calibrated. Several hours later a consensus was reached that surprised everyone:

- Pairs #1 & 3 were most neutral when directly comparing the recorded to the live triangle – possibly the most telling source. Pair #2 added overtones that were musical enough to reveal they were harmonic distortion. Pair #4 was off-putting in tone color, and so was eliminated from further consideration.
- Pair #1 was thinner sounding compared to the live shells, which had tonal characteristics emphasizing 2~300Hz. Pair #3 was nearly as full as the real thing. With the shells' high crest-factor transients approaching full scale, pair #2 distorted badly.
- The thunder machine, with a low-mid frequency tonal resonance, reproduced equally naturally on pairs #1, #2, & #3.
- Like the triangle, the earthenware vase also separated the men from the boys. Pair #3 was able to reproduce the long sustain along with the subtle but beautiful tremolo heard live in the instrument, where Pairs #1 & #2 abbreviated this sustain.
- The egg shaker, with multiple very high frequency transients, was reproduced without exaggeration on pair #2 but was boosted in highs by pairs #1 & #3.

On the basis of the non-sentimental analogs kit live v. recorded, the consensus was that pair #3 was the clear winner, with pair #1 either tied or a very close second. Surprising was that pair #3 was the least expensive and pair #1 the most expensive by a factor of 3½. Pair #2 came in a clear 3rd despite a price 20% less than pair #1. Pair #4 failed to make the cut.

Music, Music, Music

After the relative sterility of the analogs, it was decided after all to listen to music to see if any correlation could be determined. Acoustic music included orchestral, big band, and voice & piano trio. The first two, produced by the writer, were pronounced by the others as "natural-sounding" on pair #3, a bit too bright on pair #1, and "boxy" on pair #2, with cabinet resonances reacting to timpani. The voice & piano trio was a commercial recording that, although no one present had

experienced its production, was preferred on pair #1 despite emphasizing high frequency harshness on sibilants – probably artifacts of compression. (No attempt was made to EQ highs driving pair #3 to match pair #1, simulating that emphasis being made in mixing/mastering rather than altered by the speakers.)

This last point is the dilemma of circular logic that affects strictly subjective shootouts, where material familiar to subjects only as heard in various replay situations, not from having produced it, takes on a new preference. This retrospective view is not helpful to recording engineers who seek improvement, or at least consistency of results. When choices are being made during production or in post, loudspeakers proving to reproduce sounds most like the real signals should be preferred as the reference. Make the sound whatever you will using these monitors, confident there will be fewer surprises later. *Finding a sound that is preferable after-the-fact for one isolated replay situation is irrelevant or even wrong for every other.*

Discussion & Conclusions

Using “non-sentimental analogs” as test signals, the group was able to reach consensus about speaker selection: that hype, whether suggested by high-end pricing or sales talk, is no substitute for rigorously selecting a well-engineered monitor loudspeaker. Is it a coincidence that the top-rated monitor is the only one of the four evaluated whose manufacturer publishes measured performance data and curves for this and many of its other products, and has earned a reputation for honest sound? Such information is meaningful, especially regarding off-axis dispersion response, which tells more about how the speaker will couple to the room, with which it forms a system. If speaker selection is not rigorous, mixing decisions and results will be tainted, adding to acoustic changes in timbre, imaging, envelopment, and speech intelligibility due to room reflections.

All four participants agreed that evaluating the non-sentimental analogs prior affected how they listened to and evaluated the music that followed. While they might have succumbed to an emotional bias had they listened only to their “reference music,” the non-sentimental sources set the stage for more informed listening-based evaluation. If knowledge is power, perhaps more finely-honed perception is too.

The transducers in the recording/reproduction chain are the most challenged components. Unlike electronics’ relatively easy tasks, microphones and loudspeakers need to change one form of energy into another. The closer to ideal a microphone is, the more likely it will sound like another close-to-ideal microphone, neither contributing artifacts of its own. The same should be true of loudspeakers, however the dimensions and power levels involved imply greater, clearly audible colorations.

Of course, another circular logic is that real microphones were involved in recording the analogs used to evaluate the loudspeakers: did something about the microphones favor one speaker over another? Microphones, preamplifiers, and A-D converters were a necessity in the process of course, but most would accept the equipment chosen (Schoeps microphones, Grace Designs preamp/AD, no EQ or compression).

Even if surprising, the writer, until he knows a better method, is confident of these findings using non-sentimental analogs. Coincidentally, the orchestral recordings used were recorded with the same microphones used to record the analogs, and mixed on the same speakers evaluated as pair #3, validating

that no compression or equalization was needed – choices which were confirmed when the recordings were reproduced only slightly brighter in the close-second placing pair #1.

The surprising result is that the least expensive DUT came out on par with the most expensive using this approach. This seeming paradox may be due to the DSP correction inherent in #3’s design, whereby driver nonlinearities are corrected by inverse digital filters, rather than by more expensive physical solutions (explored at the AES 32nd International Conference on DSP in Loudspeakers, summarized in the JAES March 2008).

Excerpt wave files of the non-sentimental analogs are available for download, along with additional information, at Filmaker Technology, see www.filmaker.com.

*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 – cf. www.filmaker.com*

APPENDIX: Units Evaluated

Microphone – Schoeps CCM3 (2), CCM8 (2), 18cm sphere

Preamp/Converter – Grace Designs M802

Pair #1 – Barefoot MM27

Pair #2 – Focale Twin6 BE

Pair #3 – JBL LSR4328P*

Pair #4 – Quested S-7

Pair #5 – SoundLab Majestic (see below).

* Note: While this unit’s automatic Room Mode Correction system was utilized during evaluation, the sources had insignificant energy below 200Hz where these corrections might have had effect.

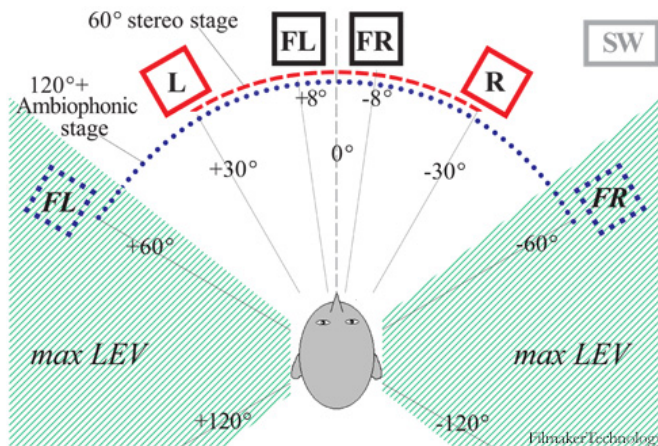


Pair #5 above are SoundLab *Majestics*, full-range electrostatic speakers (bi-directional) in the foreground, pictured at the Ambiophonics Institute, that compared very favorably to all five live sources in the experiment.

SIDEBAR: Speaker-binaural Ambiphonics

In a separate session using high-end full-range electrostatic speakers (pair #5), the writer compared the non-sentimental analogs kit, live v. recorded, using Ambiphonic principles and crosstalk cancellation. No matter what the price of the speaker, successful crosstalk cancellation requires good acoustics, and consistency between a pair of loudspeakers, especially of phase and frequency response through the crossover region. (Of course full-range electrostatics have no crossover issues.)

The microphone technique used in recording the analog sound makers is compatible with quasi-binaural Ambiphonic reproduction. In fact, for critical listening by one or two seated precisely on the median between Ambiphonic's closely-spaced speakers, many if not most legacy stereo recordings, whether from LP or CD sources, benefit from Ambiphonics.



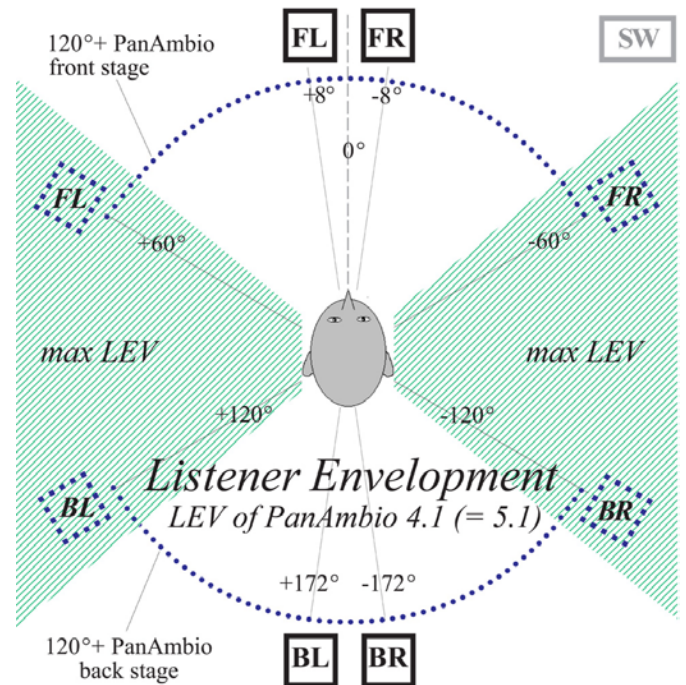
Ambiphonics speakers FL & FR in front convey a stage equaling the original recording angle extending to virtual FL & FR, uncolored center images, and more listener envelopment (LEV), reaching the maximum regions shown, all in contrast to conventional stereo speakers L & R.

Briefly, replay over conventionally placed stereo speakers, positioned in an equilateral stereo triangle with the listener, limits auditory imagery, including ambience originating around and above the listener, to the frontal 60° span of the speakers. Lost are these details, captured in spatially rich recordings containing binaural cues with inter-aural time difference (ITD) important in natural hearing. In contrast to spatial recording techniques that capture ITD, many content producers typically pan monaural microphones or direct sources between channels to create phantom images with only inter-aural level difference (ILD). ILD produces correlated speaker signals accompanied by unintended, delayed arrivals at the ears because of acoustic crosstalk, inherent in conventional stereo speaker placement. The resulting comb filtering distorts tone color.

ITD is important for headphone listening, but also for *speaker-binaural*, enabled by crosstalk cancellation, such as used in Ambiphonics. By moving the speakers close together in front and using crosstalk cancellation, Ambiphonics creates images up to 150° wide and avoids comb filtering and pinna confusion for central voices that, with phantom imaging in conventional stereo, come in fact from speakers at the sides. The disadvantages of Ambiphonics are the need to listen on a “sweet line” median to the speakers, and the possibility of audible artifacts from the crosstalk cancellation DSP algorithm due to acoustics, mismatched speaker levels, or listening off-axis. Hence Ambiphonics is for listening by one or two seated in fixed positions such as at a workstation or gaming console.

In Filmmaker Technology's Listening Lab C, the analogs kit recordings produced images perceived up to 150° wide using pair #3 except closely spaced in front and using the writer's Recursive Ambiphonic Crosstalk Elimination (RACE) algorithm. The recording angle had been 180°, producing the maximum 640µs interaural time difference (ITD) inherent in the ear-spaced microphone used, along with head-shadow-like ILD.

At the Ambiphonics Institute outside New York City, full-range electrostatic speakers (pair #5) also reproduced images over a perceived angle of 150°. Unlike what might result using hybrid cone-electrostatic speakers that exhibit omni-directional dispersion below the crossover frequency (−6dB/double-distance, *inverse square* law of a point source) but −3dB/dd for the bi-directional cylindrical source above crossover, the full-range electrostatics compared very favorably to the live sources. Recordings made using the author's *PanAmbiophone* also reproduced images of both instruments on stage and important early side wall reflections up to 150° in width. Not possible to auralize these images with conventional 60° speaker stereo, this extra width allows legacy stereo recordings to approach the natural spatiality – maximizing *listener envelopment* (LEV, see illustrations) – whereby enveloping reflections are perceived at wider angles than staged instruments, as they would be hearing them live. Yet it requires only two channels and two speakers without surround speaker layouts and multi-channel recordings. (Ambiphonics also supports increasing envelopment by adding hall response convolution signals fed to surround speakers.)



PanAmbio surround adds back speakers BL & BR, imaging as BL & BR – four “speakers” (double 5.1/.1's two) within regions where listener envelopment (LEV) is maximum. (Note: Play 5.1 in PanAmbio by setting the player to “no center” to mix the C channel to the front speaker pair.)

5.1-compatible surround sound for movies, gaming, or multi-channel music is accomplished by adding a second Ambio pair in back – termed *PanAmbio* (see illustration above). Detailed papers and DIY tools are available for download free, along with listings and reviews of commercial products incorporating Ambiphonics, at www.ambiophonics.org.