MUSIC, ROOMS AND LISTENERS
SCIENCE IN THE CREATION AND DELIVERY OF AUDIO ART

Floyd E. Toole
1301 King James Court
Oak Park, California

Could the "best" loudspeaker simply be the one perceived to be the "least bad?"

Sound sources and rooms are interactive systems. Concert halls and auditoriums are integral parts of live performances. Because architects thrive on distinctive designs, the venues are all different, making each combination of conductor, orchestra and hall a unique auditory event, never, perhaps, to be repeated again. Audiences expect and embrace the spatial and timbral idiosyncrasies and music is enjoyed. Generations of trial and error, and scientific research, have provided guidance about how to design halls that maximize pleasure while not exceeding the limits of listener adaptation. With care, the art —the performance—is satisfactorily delivered to audiences. The music may be relatively constant, but the auditory experience is not. This is sound *production*. It is what it is at the time, and it may never be again.

Elaborately illuminated and sound reinforced, large-venue popular music performances begin with microphones that sample the extreme near field of individual voices and instruments. Gigantic loudspeaker arrays make no effort to place the music into a natural acoustical context; in fact, they are designed to address the audience, avoiding the room boundaries. Much of the artistry is the responsibility of the "front-of-house" mixer, who sits at a console determining how much we hear from each of the musicians on stage, while manipulating signal-processing parameters that affect perceptions of timbre, space and dynamic range. This person can make or break a performance, regardless of how well the musicians perform, how excellent is the inherent design of the loudspeaker system, or the quality of the acoustical environment. This also is sound *production*. It is what it is at the time, and it may never be again.

As enjoyable as live performances are, the bulk of our music is mundanely delivered through loudspeakers in our homes, cars, in cinemas, or through headphones or ear-buds as we walk the dog or travel. The music itself has been captured through microphones that sample portions of the near and far fields of voices and musical instruments, with or without additional information from acoustical settings. These streams of data are manipulated in control rooms by recording engineers who decide precisely what we, the audience, will hear of those sounds. Voices and instruments are modified using any of the nearly countless electronic processing algorithms. This is done while monitoring the experience through specific loudspeakers in a specific room. Normally this is done in two channels—stereo. This is the creation of the art, the original performance; it is sound *production*. Unless the audience has playback—i.e. sound *reproduction*—capabilities that precisely duplicate this situation, this is the only time it will be heard. It is what it is at the time, and it may never be again. (Figure 1)

There are no standards for loudspeakers or rooms used in the music industry. Individual studio designers, owners and recording engineers have expectations of what they want to hear in control rooms. There are large differences among them, especially with the advent of home studios. Recording engineers attempt to anticipate what consumers are hearing, trying out their mixes in cars and over inexpensive systems in vogue at the time. Some choose to use monitor loudspeakers that they think portray the characteristics of "average" consumer playback systems. The problem with this approach is that it is not possible to standardize "bad sound." In reality, most playback systems, at all prices, aspire to be neutral. For a variety of reasons they may fail, and when they do they fail in infinite different ways. After nearly 40 years of examining

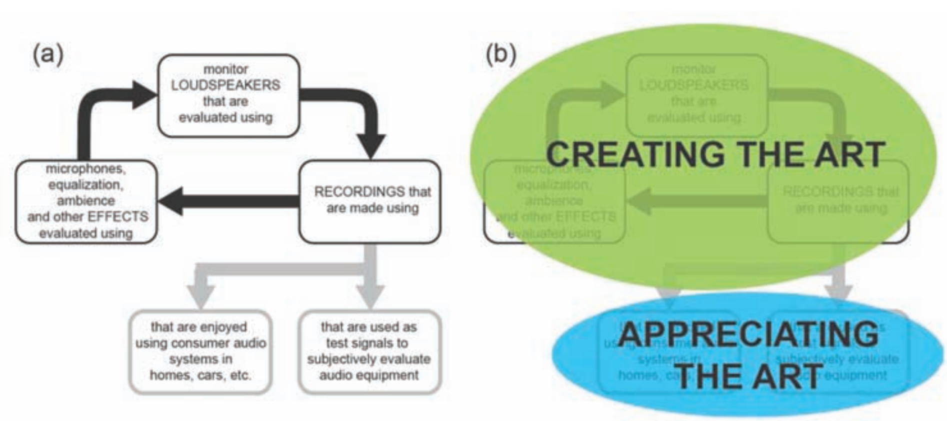

Fig.1 (a) The "circle of confusion" at the core of the audio industry and (b) the two domains that must exhibit fundamental similarities if listeners are to hear the art that was created. From Toole, 2008.

consumer playback devices, I will assert that the only *consistent* factors identifying small and/or inexpensive loudspeakers are a lack of low bass and an inability to play loud.

So, in our everyday music listening—sound reproduction—what can we expect? At the beginning of the process, microphones sampled the sound field radiating from voices and instruments. All of the sound that would reach our ears in a live performance is not captured. Therefore a perfect reproduction of a "live" event is precluded at the outset.

Storage and playback through two channels has been the industry norm for decades, and it may be convenient, but it is incapable of delivering the timbral nuances, directional effects and spatial envelopment of live performances. Instead we get two "real" sources of sound, the left and right loudspeakers, and some number of panned phantom images between the loudspeakers, assuming that we have the discipline to sit in the symmetrical sweet spot. The phantom images suffer from acoustic crosstalk—the sound from both loudspeakers reaches both ears—and both the timbre and spatial representations are unnatural (Toole, 2008, Figure 8.4 and Section 9.1.3). The spectral corruption of the important phantom center image—often the featured artist—is such that even speech intelligibility is degraded (Shirley et al., 2007). With the best of intentions, and unlimited financial investment, when listening to stereo recordings what we hear cannot be the same as a live acoustical experience.

Playing stereo recordings through headphones generates a totally different experience, and one not anticipated by a production process using loudspeakers. It is what it is, and whatever it is, it is not what was intended by the creators of the art. Multichannel audio moves us significantly closer to a desirable objective, but sadly, other than for movies, it has not been commercially viable.

Therefore, in sound *reproduction*, just as in concert hall situations, the "music" may be relatively constant, but our auditory experiences are not. It is what it is at the time, but because it is reproduced sound, we can play recordings again, and again. However, only if our personal playback equipment shares important qualities with that used to create the art, can we be assured of who or what takes the credit or blame for what we hear. We need to disrupt the "circle of confusion" by making the two domains shown on the right in Fig. 1 as similar as possible.

How is it that we find ourselves deriving pleasure from this grossly flawed system? It is because human listeners are remarkably adaptable, and not a little bit susceptible. Over 100 years ago Edison, in his "tone tests," was able to persuade normally intelligent people that his first generation phonograph was indistinguishable from real voices and instruments. He and others mounted live vs. reproduced tests in concert halls. They were all successful (Toole, 2008, Section 2.1). But wait, perhaps listeners were responding to the excellent acoustics of the halls (the recordings were "dead," without reverberation). As several studies have shown, envelopment is a critical quality of a good hall, and therefore of anything produced—or reproduced—within it. If this is not a factor, we are forced to consider that there has been no consequential improvement in reproduced sound in the past century.

Apparently it is not necessary to deliver sounds to the ears that are identical to the "real thing" for listeners to think that they are hearing something resembling, even closely resembling, the real thing. If the basic clues are there, the brain can fill in a lot of blanks. The boundary between reality and perception is a blurry one. Perhaps the most perfect sound reproduction systems are those that provide the most, and the most persuasive, perceptual "hooks" without exhibiting flaws that go beyond the limits of human adaptation.

But expectation also plays into this. There are examples of people hearing things that simply cannot be there. In high-end audio there have been numerous examples of tweaks and gadgets that defy both common sense and physical laws, all of which found a following. If you believe something, there is a chance that you will hear it. All of this can be entertaining so long as it does not encroach on the basics of a family budget.

And then there is the scientific approach.

The literature on concert hall and large-space acoustics and psychoacoustics is extensive, and it has contributed much to understanding sound reproduction in small rooms (Toole, 2006, and 2008, Chapters 4 – 11). However, recording control rooms, mastering rooms, domestic homes and cars are all small spaces. They are subject to enormous variations due to room modes that add low-frequency coloration, and the associated standing waves dictating that no two people in a room will hear exactly the same bass. At higher frequencies, the small dimensions would seem to be contrary to attempts at creating impressions of being in large spaces. But these are understandable phenomena, responsive to mathematical analysis and psychoacoustic experimentation. The problem is that relatively little scientific effort has been invested in trying to understand the acoustical factors underlying the recorded music and film industries. Is this scientific elitism? As a result, folklore, misinformation and simple ignorance compromise what is achieved in these industries. Without some trustworthy technical and acoustical guidance, the circle of confusion will never be broken. At some time, measurements of the right kind need to be trusted to describe what could be considered to be a "reference" sound quality, one that could be the target performance for both production and reproduction. The question is: what are those measurements?

Identifying the right quantitative measures

The familiar claim that "we cannot measure what we hear" stems from observations that curves may look the same but the sound is different. In the early years of audio this was certainly true. In 2013 it mostly relates to situations where the measured data are inadequate in quantity and quality, or are of the wrong kind, or that post processing has not been applied for more effective interpretation.

An omnidirectional microphone at head height at listening locations has long been employed as a basic method of evaluating sound systems in rooms. Traditionally these have been 1/3-octave filtered steady-state amplitude responses. There is a superficial logic to this, but it is not reasonable to assume that a simple omnidirectional microphone, however technically excellent, coupled to a real-time or other analyz-

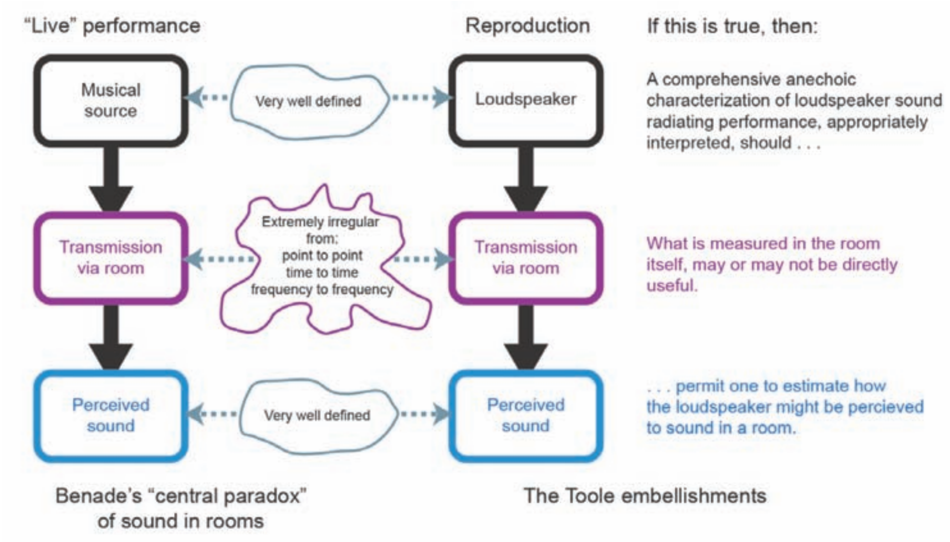

Fig. 2. A concept presented by Benade, (1984) extended by Toole.

er, is a substitute for two ears and a brain.

- To begin with, even though 1/3-octave filters approximate the critical bands/equivalent rectangular bandwidth (ERBs) over some of the frequency range, timbral cues in the form of beats and roughness originate within each of those bands—we need higher resolution if we are to have an adequate predictor of perceived timbre from sound reproducing devices (Toole, 2008 pp. 450-451).
- The common ± 3 dB tolerance is extremely generous, especially because there is no bandwidth associated with it. Humans respond to localized spectral variations at much lower amplitudes (Toole and Olive, 1988).
- The measurements include the room, and the associated non-minimum-phase reflections. Humans treat these very differently than measuring devices, because they arrive at different times and amplitudes, and from directions different from that of the direct sound. What may be perceived as innocent, indeed pleasurable, spaciousness to a human may be interpreted as a bump or dip in a measured curve that suggests a need for equalization. Evidence of non-minimum-phase phenomena should not be equalized, or what is thought of as a remedial measure has the potential to create audible problems. I suspect that this misuse of equalization is responsible for much of the criticism of it.
- The "room curve" may fluctuate because of amplitude response flaws in the loudspeaker or because of frequency-dependent directivity. Equalization can compensate for the former, but not the latter. Neither can it compensate for all fluctuations caused by frequency selective absorption at room boundaries. Significant understanding of underlying causes is required before deciding on remedial actions.
- Finally, there is indecision about the target curve to which a sound system is equalized. There is broad agreement that a flat steady-state "room curve" sounds too bright. So, depending on the venue and the program, different installers/consultants/industries employ different forms of high-frequency rolloff and/or downward spectral tilts. This is usually done with no knowledge of, or requirements for, the loudspeakers or the rooms, and yet what is measured embraces both. Such practices cannot be generalized.

This incomplete list of issues refers to common practice within the audio industry. But, more seriously, some or all of them are embodied in international standards purporting to set objectives for sound quality within the broadcast and film industries.

Can we do better? Almost certainly. I can think of no better way to introduce the viewpoint than with Fig. 2, beginning on the left with the basic observation from Arthur Benade, including my embellishments to bring it into the present context.

This is a significant change in perspective, yet it aligns with everyday experience. We can track a voice as a conversation moves from one room, down a corridor to another room. There are huge, complex, changes to the sounds arriving at our ears, and yet subconsciously we know that the sound of that voice remains essentially constant. It is perceived as a voice in changing acoustical contexts. Some of us have experienced moving around within a space, listening to a repeated passage of music, noting that what we hear is more stable than the varying details in "room curves" measured where we are located. When we stop moving, and adapt to the acoustical circumstances, rooms tend to become contexts. To a physicist, a room adds an impossibly complicated distortion of the transfer function between a sound source and a listener. To a listener, a good room embellishes the music. Understanding how reflections are perceived is important (Olive and Toole, 1989).

Adaptation—adjusting to life in an ever-changing (acoustical) world

Chapter 9 in my book discusses adaptation, beginning with:

> "In the contexts of precedence effect (angular localization), distance perception and spectral compensation

(timbre), humans can track complex reflective patterns in rooms and adjust our processes to compensate for much that they might otherwise disrupt in our perceptions of where sounds come from, and of the true timbral signature of sound sources. In fact, out of the complexity of reflected sounds we extract useful information about the listening space, and apply it to sounds we will hear in the future. We are able, it seems, to separate acoustical aspects of a reproduced musical or theatrical performance from those of the room within which the reproduction takes place. This appears to be achieved at the cognitive level of perception – the result of data acquisition, processing and decision making, involving notions of what is or is not plausible. All of it indicates a longstanding human familiarity with listening in reflective spaces and a natural predisposition to adjusting to the changing patterns of reflections we live in and with. The inevitable conclusion is that all aspects of room acoustics are not targets for "treatment". It would seem to be a case of identifying those aspects that we can, even should, leave alone, and focusing our attention on those aspects that most directly interact with important aspects of sound reproduction—reducing unwanted interference on the one hand or, on the other hand, enhancing desirable aspects of the spatial and timbral panoramas." (Toole, 2008, p.171).

There is a caution to be noted here. It is that adaptation takes time. When we are moving around we hear things that may gradually disappear when we sit down, or which may not be identified at all if one is seated when the sound begins

A dramatic example of the power of this adaptation is described in Section 11.3.1 (ibid), where three very good loudspeakers were subjectively compared to each other in four different rooms. In addition to live (listener in the room) double-blind, randomized, comparisons, binaural recordings were made for subsequent comparisons using insert earphones. It turned out that when the comparisons were organized in the manner of the live tests, one room at a time, the binaural test results were essentially the same as the live results. Statistically, the variable "loudspeaker" was highly significant ($p = 0.05$) and "room" was not a significant factor. Then those same binaural recordings were presented in a different sequence, allowing each loudspeaker in each room to be compared to each other. The results were very different: "room" was a highly significant variable ($p = 0.001$) and loudspeaker was not a significant factor. The sound of the room had merged with the sound of the loudspeaker and could not be separated because listeners had no opportunity to adapt. In this version of the test, the sounds of the different rooms were more distinctive than the sounds of the different loudspeakers. Among other things this is a caution to observe when performing binaurally recorded subjective comparisons.

Characterizing the sound source: collecting the data

Describing the three-dimensional sound fields emanating from voices and musical instruments could be one of those endless tasks because they exist in infinite variations. However, describing loudspeakers intended to *reproduce* voices and musical instruments is entirely feasible, indeed desirable, if one expects to reproduce those sounds without degradation. Ideally, we would look for indications of transparency, or "neutrality". Because we listen in reflective rooms, it is necessary to make many measurements.

Beginning in the early 1980s I collected data on loudspeakers over full horizontal and vertical orbits. It was very revealing of what listeners were responding to when judging sound quality (Toole, 1985, 1986). A smooth and flat on-axis frequency response was a starting point. As loudspeakers improved, it became clear that the loudspeakers awarded the highest subjective ratings also had relatively smooth sound power—i.e. relatively constant directivity vs. frequency. It was also shown that these anechoic data were capable of closely predicting steady-state room curves measured at the listening positions in a small room (Toole, 1986, Figures 18 – 20). This provided the basis for taking the technique to a higher level by combining measurements made at different angles to estimate the sounds arriving at a listener's ears in more generalized listening rooms (Devantier, 2002).

What we now call the "spinorama" consists of 70 anechoic measurements made at 2 m at 10° increments on horizontal and vertical orbits, frequency resolution 2 Hz (1/20-octave smoothed). These data are then processed to reveal:

- The on-axis curve: important to design engineers and solo listeners.

- Listening window: the average frequency response within a ± 30° horizontal and ± 10° vertical window. This describes the direct sound for an audience.
- Early reflected sound: the average frequency response at angles estimated for first reflections in typical domestic listening rooms.
- Sound power: average of all 70 frequency responses, each one weighted according to the proportional area of the sphere it represents.
- Total sound power directivity index: the difference between the listening window and sound power curves (a unique definition).
- First reflections directivity index: the difference between the listening window and early-reflected sound curves (a unique definition). This is included because first reflections account for much of what is measured, and heard, in rooms.

See Fig. 3. The increasing spatial averaging that occurs in the progression from on-axis through to sound power allows for the separation of acoustical interference effects (not very audible) from resonances (easily audible). The example loudspeaker is exemplary in all respects: flat and smooth axial frequency response, well-behaved, relatively constant, directivity, no evidence of audible resonances. This professional monitor loudspeaker should ensure that the recording engineer is making artistic decisions while listening to sound that is about as good as it gets. However, as noted earlier, it also represents the performance target for the majority of loudspeakers at any price. The important consistent factor, the limited low-frequency extension of small inexpensive loudspeakers, can be imitated with a variable high-pass filter in the signal path.

But what about consumers? To disrupt the circle of confusion shown in Fig. 1, consumer loudspeakers must be similar in performance to professional loudspeakers. Figure 4 shows that this is possible, even at moderate prices.

Consumers listening to these loudspeakers will hear spatial and timbral aspects of the art very much as the creators did. Most of the irregularities in the curves are close to or below the thresholds of detectability, and are not likely to seriously detract from the experience (Toole and Olive, 1988). Very low bass output is somewhat lacking, and this small cone/dome system will not play as loud as the monitor. Bass management and subwoofers would address both problems. However, larger, more expensive, domestic loudspeakers can perform in a manner that closely emulates the monitor. Sadly, price is not a reliable indicator of sound quality, and most manufacturers are reluctant to reveal useful specifications on their products, leaving consumers in an unfortunate circumstance. At present there is a standards group working on implementing the spinorama as a basis for loudspeaker specifications. However, it will be a voluntary standard.

The intent of this measurement scheme was to be able to anticipate how loudspeakers would behave in rooms both subjectively and objectively. As was shown in Toole, 1986, sound power is the dominant factor at low-to-middle frequencies, and the direct sound is the dominant factor at the highest frequencies. Taking this simple approach, Fig. 5 shows a comparison for a loudspeaker having both frequency response and directivity problems, measured in a typical domestic listening room.

Obviously there is a close relationship among these three curves. The sound power includes effects of both the axial frequency response and directivity, the inverted Directivity Index (DI) relates only to directivity. The fact that a major cause of the unfortunate shape of the room curve is off-axis radiated sound, i.e. reflections, means that equalization may not be an appropriate corrective measure. Replacement of this $10K loudspeaker seems like a better option. The shortfall at the highest frequencies indicates that the direct sound is the dominant factor in that range. Disagreement at fre-

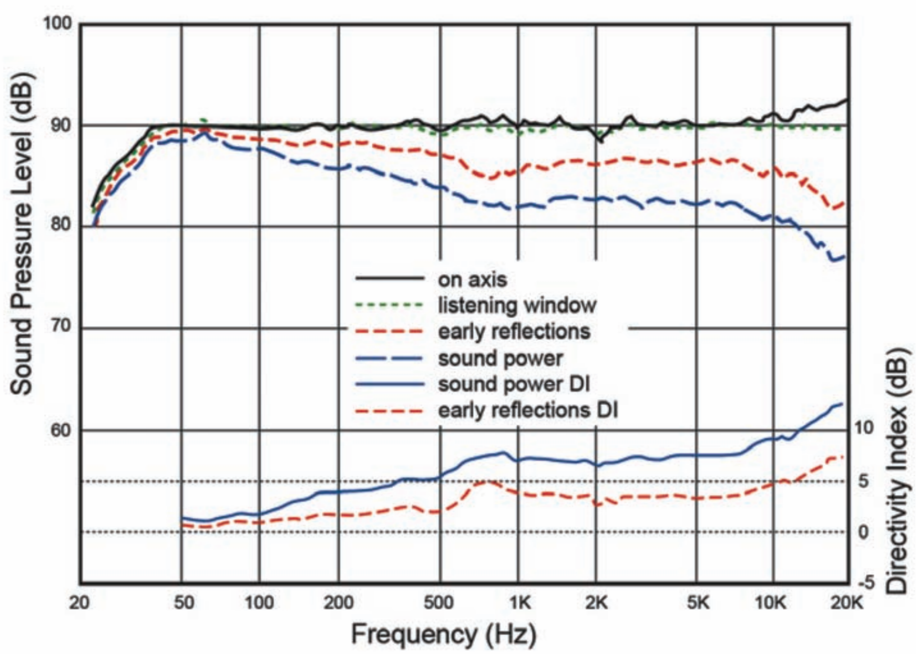

Fig. 3 The spinorama for a high power cone/horn professional monitor loudspeaker. Data: Harman International.

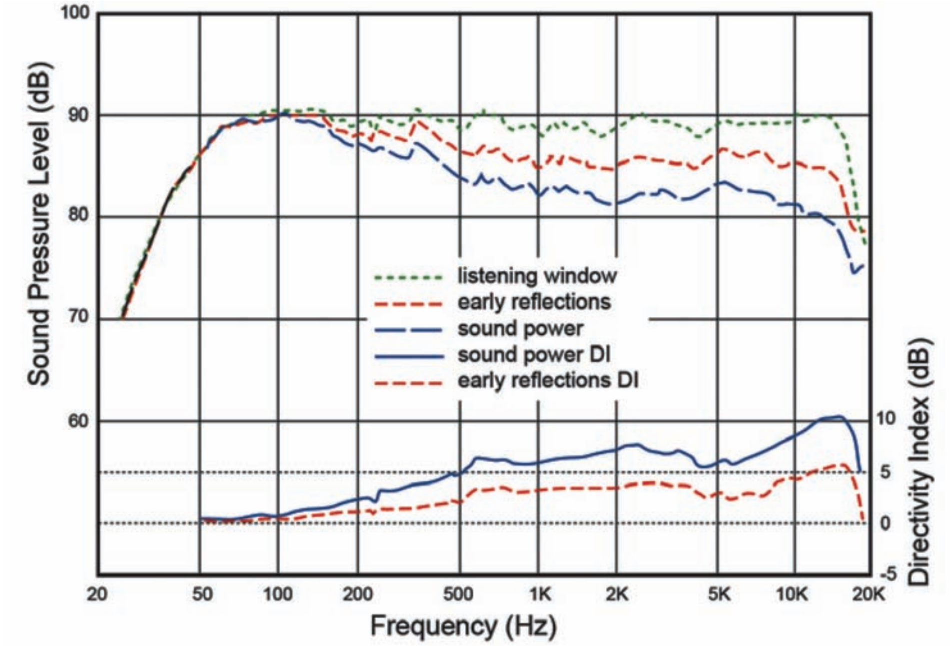

Fig. 4 The spinorama for a well-designed floor standing domestic loudspeaker with a retail price of $329 at time of writing. Data: Harman International.

quencies below about 200 Hz is the result of standing waves in the small room. Clearly an on-axis measurement alone is insufficient to describe events in small listening rooms.

Recently I presented evidence that this basic relationship holds for loudspeakers in rooms ranging in volume from small domestic rooms to large auditoriums (a volume ratio exceeding 100:1). In large venues the effects of air absorption must be incorporated and, of course, there are the inescapable effects of low-frequency modes in small rooms (Toole, 2012). Notwithstanding the assurance that comprehensive anechoic data on loudspeakers can be used to estimate acoustic measurements in rooms, the real question is: how does all of this relate to subjective evaluations of sound quality? Can we look at a set of anechoic measurements and anticipate how a loudspeaker will sound in a room?

In the 1970s, when I began my research into this topic, it was clear to me that without good subjective data progress would be impossible. I conducted well-controlled, double-blind, multiple-comparison (four at a time) subjective evaluations of loudspeakers, using many listeners, and many musical selections. It was all randomized, not automated, and it took a lot of time. The results showed that certain features of loudspeaker performance, as evidenced in the anechoic data, appeared to be strongly related to subjective ratings of sound quality. Others were not (Toole, 1982, 1985, 1986a, 1986b, most of this is summarized in Toole, 2008). It was found that listeners with even relatively small hearing losses exhibited measurable degradation in rating consistency. Nowadays, listeners are selected and trained (Olive, 1994, 2001), resulting in more efficient tests, while not affecting the ability of those ratings to reflect opinions of consumers in general (Olive 2003).

It has been a consistent observation that the relative merits of loudspeakers are basically stable across different rooms. It may have been advantageous that these were multiple-comparison tests, as opposed to single stimulus or simple A vs. B tests. Having more comparison sounds could have assisted listeners in separating, or streaming (Bregman, 1990), the sounds of the three or four loudspeakers as distinct from the relatively constant timbre contributed by the room. As Bregman says, “It seems likely that the auditory system… has developed principles for “betting” on which parts of a

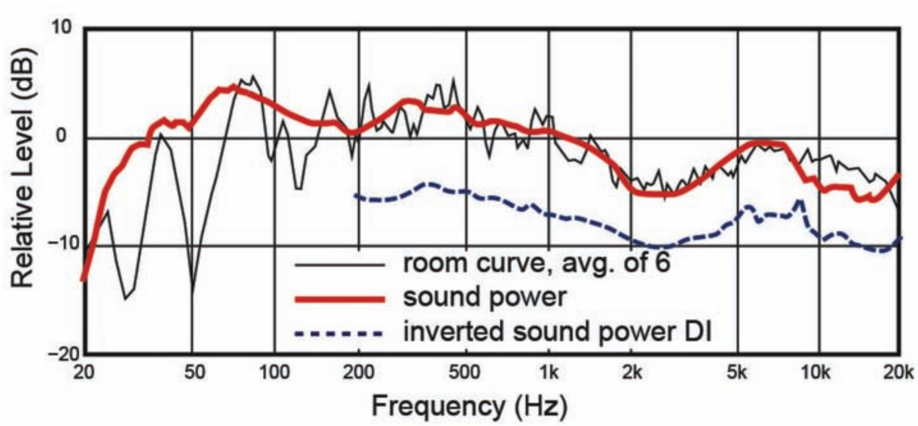

Fig. 5 The average of six in-room measurements is compared to the sound power, and the inverted sound power directivity index (DI) for the same loudspeaker. Acoustically the room is a typical, furnished, domestic space. Data: Harman International.

sequence of sensory inputs have arisen from the same source. Such betting principles could take advantage of properties of sounds that had a reasonably high probability of indicating that the sounds had a common origin." (ibid, p.24).

Of comparable interest is the fact that listeners formed remarkably similar and consistent sound quality ratings when listening to widely different kinds of music, most of which was created in recording studios. Listeners in the tests never heard it in the control room and thereby had no true reference of excellence—there was only one's internal generic sense of what it might have sounded like. But, without a certain mental image of perfection, one may be able to recognize imperfections—aberrations, colorations or distortions that are not part of *any* natural sounds. Evidence that this might be so could be seen in the descriptive reports prepared by the listeners. Detailed essays incorporating sometimes colorful language described unpleasant attributes associated with low scores, while high scores were justified with few words of flattery. Could the "best" loudspeaker simply be the one perceived to be "least bad"?

Summarizing what these listeners seem to have done, they first separated the timbral contributions of the loudspeakers from those of the rooms in which they were evaluated. This was done while listening to program material for which they could not have had a "live" reference experience. And, finally, they were able to identify, and rate, loudspeakers according to the degree of imperfection. All highly rated loudspeakers conformed to very simple "motherhood" objectives: smooth flat axial frequency response, relatively constant directivity, and low distortion. This is remarkable, yet my colleagues and I have conducted hundreds of such tests over about 35 years, and there have been no surprises.

Closing the loop

Until recently the observed relationships between measured data and listening test results have been entirely subjective. That we had no numerical correlations didn't mean that we couldn't see what good loudspeakers "looked like" in spinoramas. Nevertheless, ultimately, the objective was to devise an algorithm for processing measured anechoic data that yielded predictions of subjective ratings in rooms. In 2004, Olive (Olive 2004a, 2004b) assembled data on 70 loudspeakers that had been used in competitive analysis of products at Harman International. They ran the gamut from large expensive floor standing units to small bookshelf units. For each of them he had the results of double-blind listening tests in a room, and anechoic spinorama data. Based on years of observation, and psychoacoustic research data, he created metrics and exercised them in a multiple regression model. The result was a correlation coefficient of 0.86 between a rating predicted from anechoic data, and the results of listening tests conducted in a normal listening room. Clearly this is not guesswork. This is benchmark research, but it is not a complete answer. In these tests the listening room was a constant factor, meaning that all room mode and adjacent boundary issues were fixed. And, this was a domestic/control-room-size room. Comparably competent listening tests and correlations have yet to be done in large venues.

About 30% of the factor weightings leading to sound quality ratings related to bass performance. Therefore, in calibrating systems in different venues, in addition to spinorama data, we will need some in-room measurements, and possibly some room-specific adjustments at low to mid frequencies.

The roles of room acoustics, acousticians and psychoacousticians

In auditoriums for sound production, the room is part of the performance, and therefore it matters greatly. The science applicable to this is very well documented, and research is ongoing. However, the small rooms in which we are entertained at home, and control rooms in which music recordings and many components of film soundtracks are created, are very different matters.

The room is the dominant factor at low frequencies—standing waves, and the manner in which sources and listeners interact with them are the central issues. Room dimensions, acoustical absorption and its placement, locations of sound sources and listeners, are prime determinants of the spectral and temporal quality of bass that is heard. Massive amounts of low frequency damping helps, but is costly and/or bulky—not compatible with common notions of interior décor. Because all modes are not equally energized by woofers and not equally heard by listeners, the traditional "ideal room" investigations do not yield generalizable solutions. The supposedly advantageous dimensional ratios apply only to predetermined arrangements of sound sources and listeners within the room boundaries. With multiple sound sources operating independently (i.e. connected to separate channels) the acoustical coupling to the room modes is simply not predictable. However, if the multiple sources of low frequency energy are driven by the same signal (bass management in surround processors), it is possible to employ strategies of constructive and destructive interference among the low-frequency room modes to control the modes in a perfectly rectangular space that are and are not energized. This allows the placement of multiple listeners in regions where the bass may be more uniform and more similar.

Taking this to a higher level, one that includes rooms of arbitrary shape and allowing for more flexible arrangements of listeners and subwoofers, it is possible to process the signals supplied to each subwoofer, manipulating the room modes so that the result is a more uniform bass performance at several listening locations, and a superior bass performance in all locations. Interestingly this can be very successful with no low-frequency absorption other than that naturally occurring in the room boundaries. Adding absorption simply makes it easier. All of this is discussed in detail in Toole, 2008, Chapter 13, and references therein.

Above what I call the transition frequency (called the Schroeder crossover frequency in large auditoriums), around 200-300 Hz in domestic-size rooms, the direct sound and first reflections dominate what is measured and heard, meaning that loudspeaker directivity is a major factor, as well as the frequency-dependent absorption at the reflection points. Evidence suggests that listeners prefer loudspeakers radiating similar spectra in all directions (i.e. relatively constant directivity). Consequently, it is logical that reflections of those sounds should not be spectrally altered by reflecting or scattering surfaces of room boundaries. In practical terms, this argues for areas of either full reflection or full absorption. At present I know of no spectrally neutral sound-attenuating device, although scattering/diffusing devices can approach this, but with other consequences. The widely used (and recommended in some standards) 1-inch absorbing panels are ill advised, certainly at first-reflection locations. At very high frequencies direct sound dominates, simplifying acoustic concerns.

The reverberation time (RT) target for home entertainment spaces, based primarily on speech intelligibility, is easy to hit: ≤ 0.5 s. This number applies also to cinemas and film production facilities (dialogue again), but music recording control rooms tend to aim for lower RTs, sometimes much lower. Even at 0.5 s, with relatively directional sound sources, there is nothing resembling a diffuse sound field, meaning that random incidence absorption coefficients are of limited use. The importance of first-reflected sounds suggests that it might be advantageous to know the angle-specific frequency-dependent absorption and scattering/diffusing properties of acoustical materials and devices.

There is evidence that the precedence effect deteriorates when the spectra of the direct and delayed sounds differ. It is plausible to think that similar effects extend to other aspects of perception, including spaciousness and timbre. Chapters 5 thru 10 in my book provide an overview of some of the factors, but it is clear that we need more data elaborating the progression of perceptual effects for level and spectral variations within isolated and multiple reflections. These data would ideally come from psychoacoustic experiments incorporating delay and directional variables associated with realistic listening circumstances. The result would be solid evidence supporting performance targets and tolerances for the off-axis performance of loudspeakers and the reflecting/absorbing/diffusing surfaces at which first reflections occur in room. This could be an interesting collaboration between scientists with acoustical and psychoacoustical expertise. A global industry awaits guidance.

Subjectively it has been found that the effect of the room is greatest with a single loudspeaker (channel) with its effects diminishing as the active channel count increases. However, with a very high proportion of movie and TV sound emerging from the front-center channel (a mono signal) the room cannot be ignored.

The inevitable question is: What constitutes an "ideal" listening room? Right now we don't know, and given the ability of humans to adapt to differing rooms, it may matter less than some people would like us to think. However, there is a limit to what we can adapt to, and adaptation very likely utilizes a portion of our neural "horsepower" (causing fatigue?). So, perhaps that feeling of exquisite relaxation I get when I listen to a superb sound system is real, not a figment of my imagination. If so, there is motivation for research by acoustical scientists, and work to do by competent acoustical engineers.

Looking Ahead

Because of the science we have, and the abundance of affordable measurement tools, the standards of sound reproduction in general have been elevated in homes and recording facilities. However, problems remain, in the form of loudspeakers that are less good than they could have been, flawed acoustical treatment practices and misguided attempts to "equalize" rooms. Right now, there is no assurance that reproduced sound closely resembles what was heard at the time the art was created. This is a pity, because it is not possible to confidently attribute credit or blame for what we hear.

In the end, consumers, audio professionals and acoustical consultants need to be able to anticipate whether a playback facility is likely to deliver a reasonable facsimile of an original performance, without exceeding the tolerances of normal adaptation. We certainly need more and better specifications on loudspeakers, and manufacturers with the courage to publish them. Between here and there are many opportunities for challenging applied research projects, generating new knowledge, and interminable committee meetings for those willing to undertake the standards work.

The technology to do much better exists. In the meantime, there are countless personal opinions to sort through, and a lot of adapting to get on with so that we can enjoy the abundance of music out there. Fortunately music is what it is, in spite of the acoustical variations and abuses we heap upon it.

References

Benade, A.H. (**1984**). "Wind instruments in the concert hall." Text of an oral presentation at Parc de la Villette, Paris; part of a series of lectures entitled "Acoustique, Musique, Espaces", 15 May 1984 (personal communication).

Bregman, A.S. (**1990**), "Auditory Scene Analysis, The Perceptual Organization of Sound," MIT Press, Cambridge, MA.

Devantier, A. (**2002**). "Characterizing the Amplitude Response of Loudspeaker Systems," 113th Convention, Audio Engineering Society, Preprint 5638.

Olive, S. E. and Toole, F. E. (**1989**). "The Detection of Reflections in Typical Rooms," Journal of the Audio Engineering Society, 37, pp. 539-553. Available at: <http://www.harman.com/en-us/our-company/innovation/pages/scientificpublications.aspx>.

Olive, S. (**2001**). "A New Listener Training Software Application," 110th Convention, Audio Engineering Society, Preprint No. 5384.

Olive, S. (**2003**). "Difference in Performance and Preference of Trained versus Untrained Listeners in Loudspeaker Tests: A Case Study," Journal of the Audio Engineering Society, **51**, pp. 806-825. Available at: <http://www.harman.com/en-us/ourcompany/innovation/pages/scientificpublications.aspx>..

Olive, S. (**2004a**). "A multiple regression model for predicting loudspeaker preference using objective measurements: part 1 – listening test results," 116th Convention, Audio Engineering Society, Preprint 6113.

Olive, S. (**2004b**). "A multiple regression model for predicting loudspeaker preference using objective measurements: part 2 – development of the model," 117th Convention, Audio Engineering Society, Preprint 6190.

Shirley, B.G., Kendrick, P. and Churchill, C. (**2007**). "The effect of stereo crosstalk on intelligibility: comparison of a phantom stereo image and a central loudspeaker source," Journal of the Audio Engineering Society, **55**, pp. 852-863.

Toole, F. E. (**1982**). "Listening tests – turning opinion into fact," Journal of the Audio Engineering Society, **30**, pp. 431-445.
Toole, F. E. (**1985**). "Subjective measurements of loudspeaker sound quality and listener preferences," Journal of the Audio Engineering Society, **33**. pp. 2-31.
Toole, F. E. (**1986**). "Loudspeaker measurements and their relationship to listener preferences," Journal of the Audio Engineering Society, **34**, pt.1, pp. 227-235, pt. 2, pp. 323-348. Available at: <http://www.harman.com/en-us/ourcompany/innovation/pages/scientificpublications.aspx>.
Toole, F.E. and Olive, S.E. (**1988**). "The modification of timbre by resonances: perception and measurement," Journal of the Audio Engineering Society, **36**, pp. 122-142. Available at: <http://www.harman.com/en-us/ourcompany/innovation/pages/scientificpublications.aspx>..
Toole, F. E. (**2006**). "Loudspeakers and rooms for sound reproduction – a scientific review," Journal of the Audio Engineering Society, vol.54, pp. 451-476. Available at: <http://www.harman.com/en-us/ourcompany/innovation/pages/scientificpublications.aspx>..
Toole, F.E. (**2008**). "Sound Reproduction, The Acoustics and Psychoacoustics of Loudspeakers and Rooms," Focal Press, Oxford.
Toole, F.E. (**2012**). "Home Theaters to Cinemas: Sound Reproduction in Small and Large Rooms," presented at a Sound for Pictures workshop, 133rd AES Convention, San Francisco. PowerPoint available for a limited time at: https://dl.dropbox.com/u/108963424/Toole-%20San%20Fran%20AES%20with%20added%20notes.pptx

Floyd E. Toole studied electrical engineering at the University of New Brunswick, and at the Imperial College of Science and Technology, University of London, where he received a Ph.D. In 1965 he joined the National Research Council of Canada, where he reached the position of Senior Research Officer in the Acoustics and Signal Processing Group. In 1991, he joined Harman International Industries, Inc. as Corporate Vice President—Acoustical Engineering. In this position he worked with all Harman International companies, and directed the Harman Research and Development Group, a central resource for technology development and subjective measurements, retiring in 2007.

His research focused on the acoustics and psychoacoustics of sound reproduction in small rooms, directed to improving engineering measurements, objectives for loudspeaker design and evaluation, and techniques for reducing variability at the loudspeaker / room / listener interface. For papers on these subjects he has received two Audio Engineering Society (AES) Publications Awards and the AES Silver Medal. He is a Fellow and Past President of the AES, a Fellow of the Acoustical Society of America, and a Fellow of CEDIA (Custom Design and Installation Association). He has been awarded Lifetime Achievement awards by CEDIA and ALMA (Association of Loudspeaker Manufacturing & Acoustics International).