ART AND SCIENCE IN THE CONTROL ROOM

Floyd E. Toole   Harman International, Northridge, California, USA

1 INTRODUCTION

Audio engineering employs both art and science to capture and process the audio experiences of the entertainment industry. Storage and reproduction of the art are – or should be – scientific/technical exercises. Confusion of the two domains has created some colorful audio folklore, assisted by the willingness of the human brain to generate perceptions supporting much of what we want to hear. Our product is sound, and the premise upon which our industry is based is that customers will be able to hear close replicas of the sounds that were created in concert halls, jazz clubs, dubbing stages and recording studios. The art needs to be preserved. This is a profound challenge, since we know that monitor loudspeakers in control rooms certainly do not all sound alike, and consumer loudspeakers and rooms cover an enormous range of qualities. Evidence of variable recording quality and variable playback quality exists in abundance. It sounds like a hopeless task and, if professionals ignore the existing science, it is hopeless. However, thanks to advances in consumer audio, playback quality is improving and examples of genuine excellence can be found. Consequently, the old problem of trying to guess what a recording will sound like through a 'typical' playback system is less of a lottery. Science has given us the means to technically and subjectively identify truly good loudspeakers with high reliability. Equally important, it has given us the means to deliver reliably good sound in different rooms.

2 PRESERVING THE ART

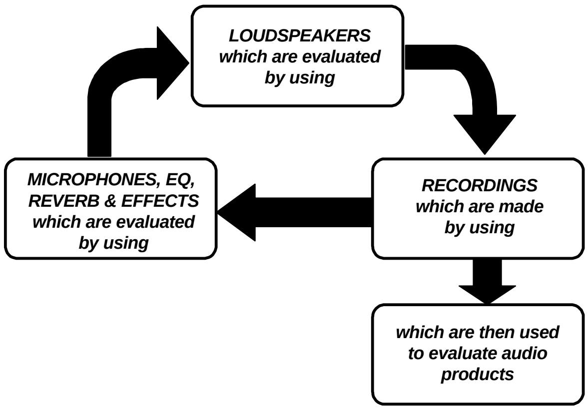
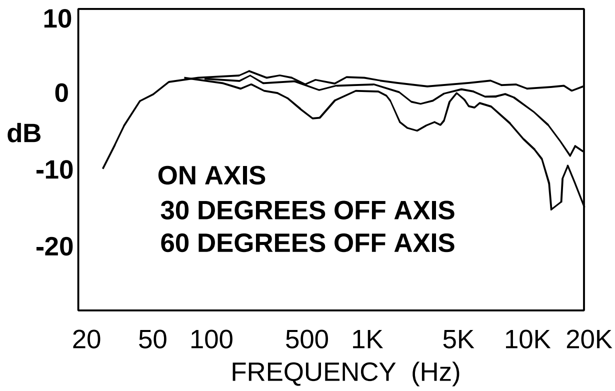


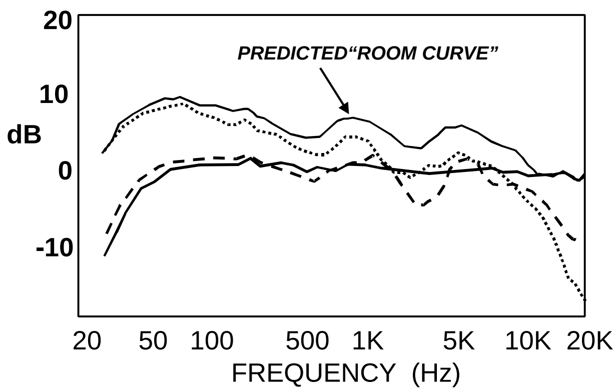
Figure 1 (left) is what I call the 'circle of confusion', illustrating that recordings inevitably must be influenced by the performance of a combination of the control room and the monitor loudspeakers in it. Since the industry lacks rigid standards for the performance of either, recordings end up being inconsistent in various ways. These recordings are then used to demonstrate and to evaluate audio products of all kinds. It is the equivalent of making a technical measurement with an uncalibrated instrument. This is not good engineering – either artistically or technically.

If we could rely on the studio monitor loudspeakers sounding like the reproducing loudspeakers used by consumers, then the 'circle of confusion' would be broken. This, then, must be our objective, and it includes controlling the combination of the loudspeaker and the room. Sounding alike is one thing, but sounding 'good' is another. What is our performance objective?

For decades we have expected ruler flat frequency responses from our electonic devices. Why, then, do we think that ± 3 dB, say, is an acceptable frequency-response specification for a loudspeaker? There are several explanations. Loudspeakers radiate 3-dimensional sound fields that defy simplistic descriptions. It is costly and difficult to get accurate, high resolution, acoustical data on loudspeakers, especially at low frequencies. Also, a listening test is still the final arbiter of quality, and that must be done in a room, with all of the uncertainties added by a complex acoustical space. The relationship between anechoic measurements and the results of listening tests has historically not been a happy one. Some subjectivists assert that we simply cannot measure what we hear. If the measurements are inaccurate or incomplete – a common failing – the assertion is most certainly correct. However, enough accurate measurements, and some computer processing of the data can change all of that.

3 MEASURING THE LOUDSPEAKER / ROOM SYSTEM





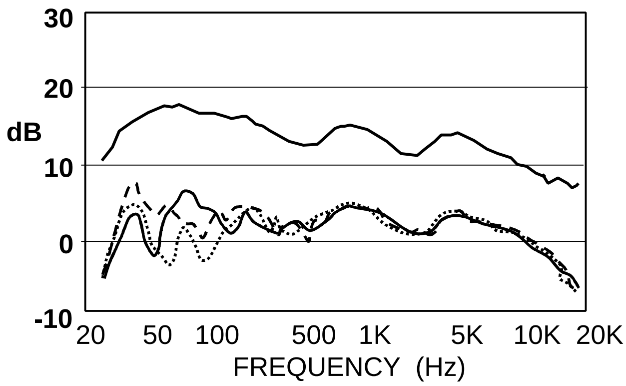


Figure 2 shows anechoic measurements on a loudspeaker with (top to bottom) a very flat on-axis response and increasingly non-flat behavior at increasing angles off axis[1], a result of crossover-frequency choices.

One approach to loudspeaker design considers the axial frequency response to be paramount. The consequences of this philosophy to listeners in normally reflective listening rooms are profound. Here, measurements made at many locations surrounding the loudspeaker, are used to compute the sounds that would arrive at a listener's location in a typical domestic room (propagation losses included).

Figure 3 shows the sequence of sound fields arriving at a listener's ears. The direct sound is represented by the on-axis frequency response (middle solid curve). The second sounds are early reflections from walls, floor and ceiling (the dashed curve), and the late multiple-bounce reflections are represented by a calculation based on total sound power (dotted). The top curve is the energy sum of the three – a prediction of what might be measured in a real room: the room curve[1]. At this point is it very clear that an on-axis frequency response is an inadequate predictor of what a listener hears in a room. In fact, so is any single curve. A true evaluation requires a complete set of data.

Figure 4 shows measurements made in a real room, with the loudspeaker moved among three realistic left/right locations within a two-foot (0.6 m) radius. The top curve is the calculated room curve from Figure 3, shifted upward by 10 dB for clarity[1].

Below about 400 Hz the measured room curves are clearly dominated by loudspeaker position and the degree of coupling with standing waves in the room. None of this was included in the prediction. Above 400 Hz the responses have settled down to a relatively stable shape that is very well predicted by the room curve synthesized from a set of anechoic measurements[1,2].

And now the inevitable question: can any of this be improved with equalization? Above 400 Hz the answer is no. The problem with this loudspeaker was its non-uniform directivity. Changing its frequency response cannot fix the problem. At low frequencies, the intelligent use of equalization can improve things, but only a better loudspeaker will correct the problems at middle and high frequencies.

4 AN IMPROVED FREQUENCY RESPONSE MEASUREMENT

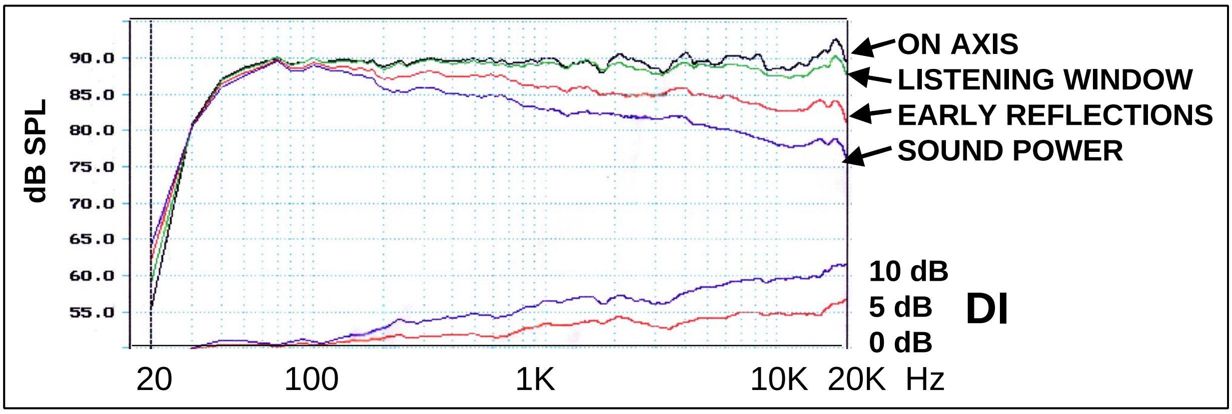


Figure 5. Anechoic data showing frequency responses on-axis, averaged over a listening window of ± 30º horizontal and ±10º vertical, single-bounce early reflections from room boundaries, and total sound power. At the bottom are directivity indices: the differences between listening-window and sound-power curves (top) and listening-window and early-reflections curves (bottom). The top curve is the classic DI.

These curves attempt to anticipate performance of this loudspeaker in a 'typical' domestic listening room or home studio installation. The combinations of the raw data that contribute to the individual curves are explained in reference 2. Ideally, the axial response should be smooth and flat, and all other curves should deviate from it smoothly and gradually – at least for a traditional forward-firing loudspeaker design. The DI's would then also be relatively smooth and slowly changing.

Support for this objective has come from hundreds of double-blind listening tests, conducted over a 25-year period, in several different rooms[1,3,4]. The results have shed important light on the correlation between technical and subjective domains (psychoacoustics), and also on listeners, explaining why some of us don't always agree on what sounds good. If sufficient attention has been paid to eliminating biasing effects that have nothing to do with the sound of the loudspeakers themselves[5,6,7], most people, most of the time, agree on what sounds good[3]. Recent work has focused on selecting those listeners who have an aptitude for critical listening, and training them to be even better[8]. The result is a pool of listeners who can speedily generate reliable, repeatable opinions. Most importantly, they agree with other persons having widely differing levels and kinds of listening experience[4]. The bottom line is that humans can be remarkably trustworthy 'measuring instruments' if they are given the opportunity, and their opinions correlate very well with data of the kind shown in Figure 5. This is an excellent loudspeaker, a control-room monitor. It will need assistance from a good subwoofer if it is to reveal the very low-frequency sounds that are reproduced by the best home systems. Otherwise it is an example to be imitated. In a frequency response, relatively small deviations from flat and smooth, if they are caused by resonances, are audible[9]. The on-axis curve of Figure 5 shows some fluctuations and the question arises: are they resonances or are they acoustic interference effects? The latter will be different for every measuring location, so we conclude that irregularities that survive spatial averaging are resonances. Since all but the small bumps at 4 and 18 kHz are attenuated by even the limited amount of spatial averaging used to calculate the listening window, we conclude that the rest are relatively benign interference effects. And what of the resonances? Both are well damped, and both are below the known thresholds of detectability[9], so even they are not audible problems. This is an exceptionally 'neutral' loudspeaker, a transparent window into the art.

But, how can recording engineers use such an accurate, analytical, loudspeaker to determine how recordings will sound through the mini-systems, boom boxes, TV's and clock radios that populate average homes? Don't we need some 'bad' speakers to make that judgment?

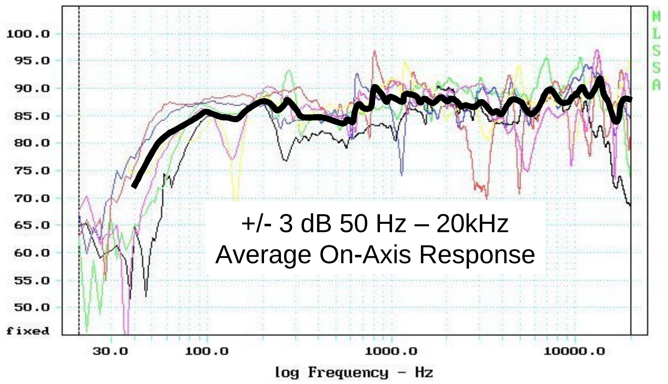


Figure 6. The on-axis frequency responses of six consumer mini-systems (combination CD/cassette/AM/FM/amplifier/speakers) ranging in price from $150 to $400. The darkest curve is the average of all six curves.

There are many ways to fail and few ways to succeed in making good sound. It would appear that each of these manufacturers had the same goal in mind – relatively flat and smooth – but they failed to achieve it in different ways. Any good small monitor loudspeaker would be a match for the average curve shown here. A large monitor with a high-pass filter inserted to attenuate the bass would do equally well. The only consistent failing of 'bad' loudspeakers is the lack of low bass. Otherwise, some are bright, some are dull, some are boomy, honky, hollow, and so on and on. There really is no need to clutter our control rooms, and offend our ears, with nasty little boxes that can provide only one example of the countless ways to be bad. If a recording engineer has a comfortable familiarity with a loudspeaker known not to be neutral, it is time to break the habit. A caution: picking a new loudspeaker that flatters one's old recordings may only perpetuate an error.

5 ROOMS WREAK HAVOC AT LOW FREQUENCIES

Figure 3 gave one example of what a room can do to a loudspeaker at low frequencies. A large-scale study of professional control-room monitor installations[10] paints an even worse picture. It showed that, while the overall average performance of the 250 installations was quite good, the individual systems had frequency responses with max/min differences of over 15 dB at 100 Hz, and worse at lower frequencies! Measurements were made at the prime listening position, and the systems used the same or similar loudspeakers. So, let us not be overly disparaging of consumer audio.

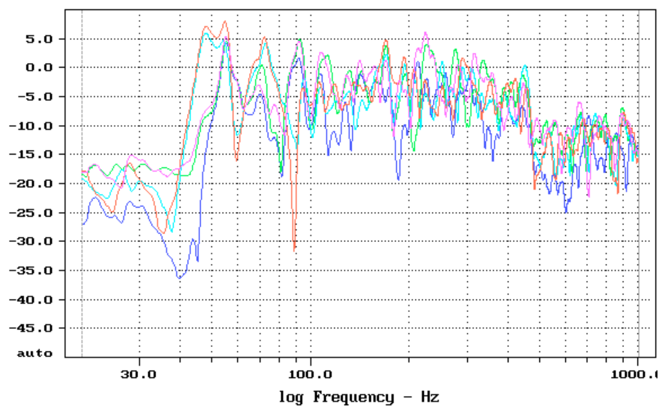


Figure 7. Low-frequency measurements made from each of the five loudspeaker locations commonly specified for multichannel music monitoring (0º, ±30º, and ±110º) to the center console location, in a 20' x 24' (6 x 7.3m) room.

The notion of having five identical full range loudspeakers for multichannel music monitoring has an appealing simplicity. However, the consequence to bass response is appalling. The range of bass levels in Fig. 7 is about 40 dB, all of the large deviations occurring below about 100 Hz. If ever there were an argument to use bass management and subwoofers, this is it. Another is that virtually all multichannel installations in homes employ that configuration. It is not a cost-saving measure; it is a superior way to reproduce bass in rooms. But there is more. Getting consistent bass for each of the five channels is a good thing, but now we need to be able to get similar bass performance in different locations in the same room, and similar performance in different rooms – ideally the same in the control room and in every listener's room or car. This is a crucial element in the 'circle of confusion'.

6 TAKING CONTROL OF ROOM RESONANCES

In 1990 the author published an example of using selective acoustical coupling of two subwoofers to eliminate an annoying room resonance (Fig. 14 in Ref. 11). Other exercises followed, culminating in a study by Welti[12] addressing the possibility of finding optimal arrangements of multiple subwoofers to attenuate room resonances and expand the satisfactory listening area to most of the room. In 'small' rooms – control rooms and home listening spaces – there really are only a few audibly significant modes, so the prospects appeared to be good. The initial investigation limited itself to perfectly rectangular rooms, with no large shape irregularities, nor any openings into other spaces.

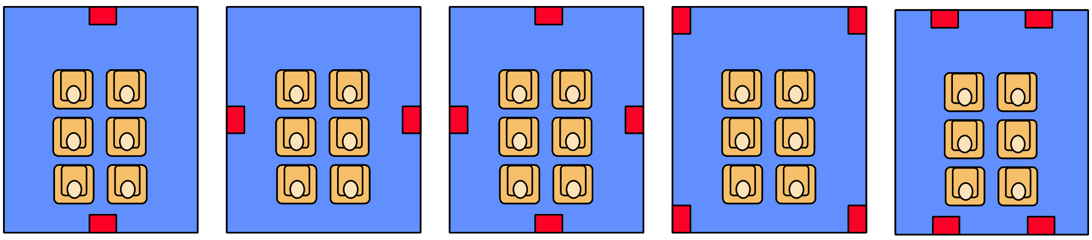


Figure 8. Some of the optimum arrangements of two and four subwoofers in a rectangular room[12].

The objective of this study was to achieve maximum similarity in bass performance over a large listening area, as in a home theater. Of course this also applies to a control room in which there are multiple listening locations. If the objectives are more restricted, as in achieving consistent bass across the width of a console, there are more specific solutions: in this case, two subwoofers at the front of the room located at the 25% positions, as in the right example in Fig. 8. Two subwoofers in these locations will seriously attenuate width modes, leaving the length modes untouched[13].

A subsequent investigation addressed the more common circumstance of rooms that are not perfectly rectangular, and that may have openings to adjacent spaces. It also considered the realities that all prescribed subwoofer locations may not be useable in a real room, and that we may be more interested in maximizing bass performance at a few specific locations, rather than over a general area. This involves measurements and signal processing for each of the multiple subwoofers[14]. An example:

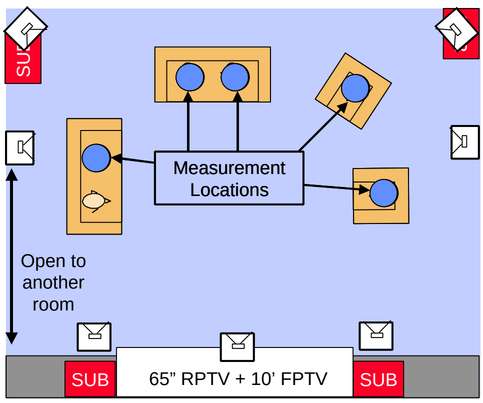


Figure 9. An entertainment room with seven channels, four subwoofer locations, and five prime listening locations, all well separated in the listening space. For practical reasons,

front corner locations were not available for subwoofers. The room has one wall substantially open to an adjacent room and an asymmetrical cathedral ceiling.

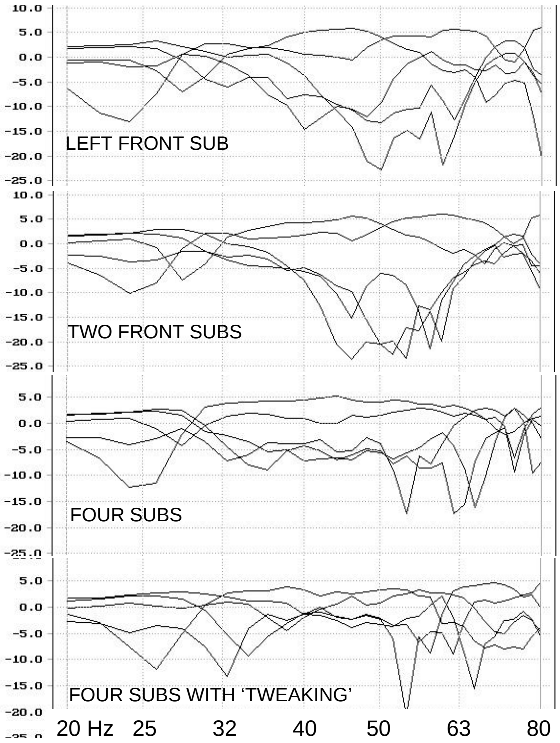


Figure 10. Frequency-response measurements made at each of the five listening locations (2 Hz resolution) for each of four subwoofer configurations.

One subwoofer obviously provides poor seat-to-seat uniformity. The max/min deviation is 28 dB! Two front subs exercise some control over the width modes and we see front and back 'row' clusters of curves. The two front seats are doing quite well. Four subs flatten the back row curves but the 'rows' are 6 to 10 dB apart in overall level. Processing the individual sub channels brings all of the curves closer together. Now they are within about ± 3dB over most of the range. Global equalization has been applied to all of the curves to arrive at a flat average performance. However, the curve for one of the seats could be selectively equalized, and the rest would follow with the same tolerance. This would be the likely approach in a control room.

If all this seems expensive or complicated, just think of the expensive real estate that historically has been dedicated to gigantic 'bass traps' in control rooms, many of which didn't eliminate the problems. This room has no dedicated low-frequency absorbers.

7 INTELLIGENT EQUALIZATION

Once uniform bass has been achieved at the listening locations, equalization is the final operation. In this we are significantly aided by the fact that low-frequency room resonances behave as minimum-phase systems, i.e. their phase characteristics, and therefore their transient behavior, are linked irrevocably to the amplitude response. If there is a bump in a high-resolution (e.g. 1/20-octave) room curve, it may be a resonance and, if so, it will ring. If the bump is attenuated, the ringing diminishes.

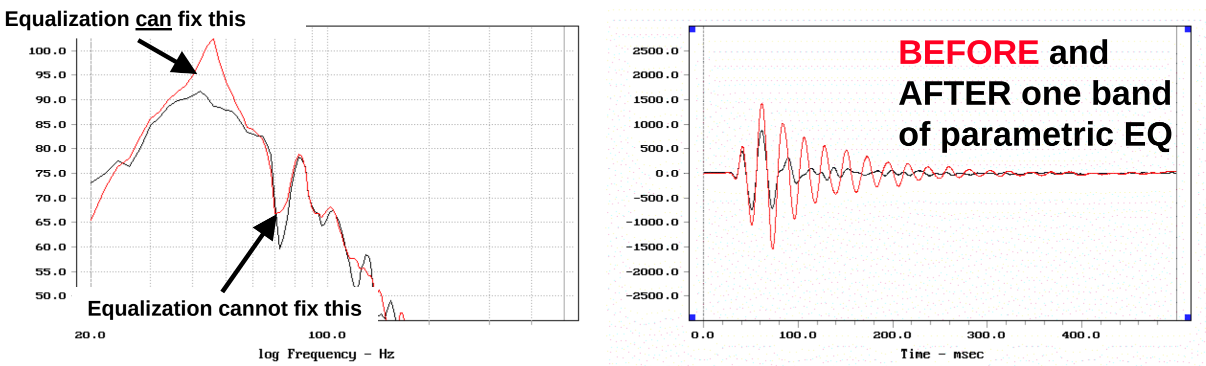


Figure 11. A subwoofer, measured at a single location, exhibiting a strong resonance at 47 Hz, and an interference dip around 75 Hz. Red curves: before EQ, black curves: after EQ.

Here, frequency and time domain measurements show how improvements in one domain are paralleled by improvements in the other. Needless to say, the boomy, flabby bass exhibited by the original system is smooth and tight after equalization; an electronic fix for an acoustical problem – but only at one seat, unless the mode-control measures described above have been implemented.

Resonances can be equalized with parametric filters addressing the frequency, Q and attenuation needed to mirror the bumps. 1/3-octave resolution in measurements and equalization simply is not adequate. Separating the bumps due to resonances from those that are associated with acoustical interference is the trick. In this we are aided, again, by spatial averaging, or multiple-location measurements of the kind seen in Fig. 10 – bumps that persist are probably resonances. Dips are caused by acoustical interference, they are highly position dependent, and they cannot be equalized. Acoustical remedies, perhaps only some positional changes, are required to address these problems.

So, below about 100 Hz room resonances can be controlled. Above about 300 Hz the loudspeaker is king, to the point where, if one has comprehensive and trustworthy loudspeaker measurements, there may be nothing that can be done in the room to make it sound better. However, between about 100 and 300 Hz mounting geometry, adjacent boundaries, and strong solitary reflections can cause problems that defy simple electronic cures, and most probably need to be dealt with acoustically.

8 REFLECTION CONTROL AND SMALL-ROOM ACOUSTICS

Listening to a multichannel reproducing system in a small, acoustically well-damped room is very different from being one of hundreds or thousands in a large, reverberant concert hall. The science of concert halls is mature, although there are still areas of debate. The science of small rooms for sound reproduction is still developing and it changes with every significant change in recording/playback technology. Listening-room designs that evolved in the age of stereo need to be reconsidered for multichannel audio. Then there is the fundamental disparity between film sound and music, relating almost entirely to the number, placement and directivity of the surround loudspeakers, which makes the issue more suitable for debate than standardization.

A diffuse sound field does not exist in small rooms, making large-room concepts like reverberation time, critical distance and Schroeder frequency of limited use. There is a kind of parallel reality in small rooms, but not at all in the statistical acoustic sense of large halls. The difference is heightened by the use of several moderately directional loudspeakers for sound reproduction systems, in contrast with the classic omnidirectional loudspeaker or starter's pistol used to imitate live musicians for measurements in concert halls. One could argue that a perfect multichannel recording should contain all of the necessary directional and spatial information, and that the role of the listening space is simply not to get in the way. If so, a large truckload of glass fiber should provide the ideal acoustical room treatment.

The direct sound dominates localization, and initiates a precedence (forward-temporal-masking) interval, during which some of the later arriving sounds are perceptually suppressed in some respects (e.g. localization) but not in all respects (e.g. timbre). Reflections and reverberation add to timbral richness[9], which helps explain why dead rooms sound so unpleasant. The key issue, then, is how loud these later reflections should be so as to embellish 'live' sounds and at what level they degrade the directional and spatial effects in recordings. In fact, should not recordings be monitored in rooms with some 'normal' reflections, since they exist in all customers' homes, even custom home theaters.

A provocative observation: Meyer derived a measure for diffusion in a room, according to which, a model of a bare room yielded a diffusivity of 69%. Making the floor absorbent dropped it to 46%. Using the same absorbing material, and placing pieces of it so as to absorb the first reflections between the sound source and microphone reduced the diffusivity to only 26%. Throughout, the reverberation time remained approximately the same[15]. However, one can imagine the huge differences in perceived sound and spatial qualities in a stereo or multichannel system auditioned in those circumstances. Many current control-room concepts tend towards low diffusivity sound fields in which RT_{60} and related measures are simply not relevant. However, something measurable must be, so what is it?

Looking at the effects of just a single reflection, it is very clear that, in small rooms, it is probably not practical to eliminate all early reflections that are audible, especially those from transient sounds.

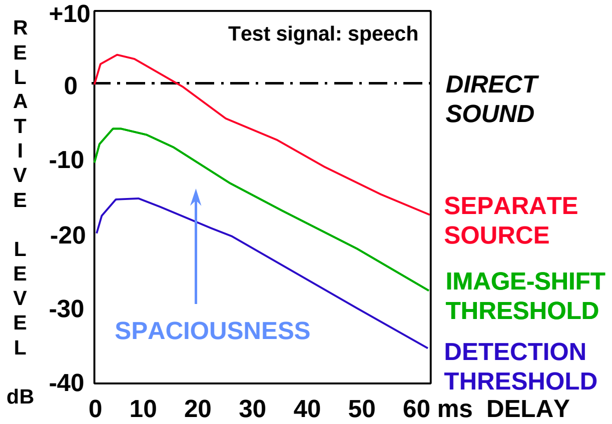


Figure 12. Relative to the direct sound, the levels at which a single lateral reflection creates different audible effects, as a function of delay. The tests were done in an anechoic chamber using speech as a signal[16].

Here is the progression of audible effects for a reflection, beginning with the point at which it is detected. The common description of the perception is spaciousness. The next significant perceptual effect appears about 10 dB higher, when listeners perceived a change in the size and/or position of the principal image. About 10 dB higher than this, the reflection itself becomes audible as a second image. In the figure, the reflection can be louder than the direct sound since the entire array of signals was artificially created.

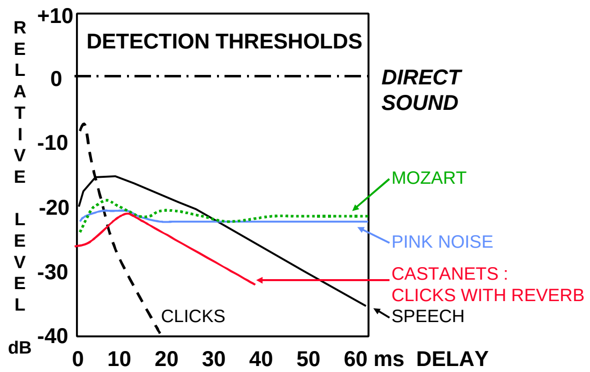


Figure 13. The detection thresholds for a single reflection determined with different sounds[16].

Sounds that are 'continuous', like pink noise, or almost so, like Mozart (with long reverberation), have almost horizontal detection curves. More discontinuous sounds, speech, show a tilt. The tilt increases with the degree of discontinuity, culminating in a very vertical curve for "dry" clicks. The requirements for room design and treatment to eliminate audible transient reflections will need to be very much more aggressive than that currently specified in popular standards (-15 dB up to 20 ms), which is about right for speech. But, as has been hinted at by Walker[16], how much is really necessary? Some things that look bad in measurements just don't seem to sound that bad. This is where speculation and opinions take over, and the need for psychoacoustic data derived from listener preferences in real-world situations becomes painfully evident. We live with room acoustics all of our lives, and it is a fact that we adapt, very quickly, to some things, and not to others. We need to know the distinctions, and the acceptable limits of the parameters that are problems.

What do we do with the surround channels? In music recording, the surround loudspeakers are equal to the fronts, just in different locations. With movies, things are different. In home theaters emulating the cinema array of multiple surround loudspeakers, the surround loudspeakers may be multidirectional units (dipole, bipole, etc.), designed to generate an active reflected sound field. Conceived as an aid for the single surround channel in early matrix decoders for movie sound tracks, they continue to be installed in general-purpose multichannel systems. The reflective environment that allows such loudspeakers to function as intended is common in untreated domestic rooms, but control rooms? While such loudspeakers may work well for music recordings in which the surrounds provide only ambience, there is a conflict with recordings that direct specific sounds to specific surround channels. Now, systems with three and four surround channels exist. What loudspeaker directivity is now optimal? Do we treat each of these to the degree of reflection control that we sometimes apply to the front channels? Or, acknowledging that much of what surrounds do is in the category of ambience or ambiguous localization, do we design in some judicious reflections and/or selective scattering to increase the interaural decorrelation at the listeners' ears?

In the consumer world, forward-facing wide-dispersion loudspeakers dominate the marketplace. Good examples are praised for the sense of space, depth and acoustical setting they provide for many kinds of recordings. Reflected sounds are part of this picture. A recording engineer listening in an environment cleansed of reflections hears something very different from what these customers hear. The need for an acoustical 'magnifying glass' for recording is understandable, so it seems that recording facilities may need to provide adjustable acoustics, or a second, normal, acoustical environment for playback evaluations. This need takes on another dimension when one appreciates that, these days, not only expensive home theaters, but affordable A/V receivers are equipped for 6 and 7 channel playback, and many people listen to stereo through multichannel conversion algorithms. It might be a good idea to see how a recording might sound to these customers too. Oh yes, several high-end cars now have 5- or 7- channel audio.

9 DISCUSSION

Whether one is on the technical or artistic side of audio engineering, and whether the product one is involved with is hardware or software, the result that counts is sound, and it only counts when it reaches the ears of a listener in his or her chosen environment.

In terms of sound quality, fidelity, the capability exists to make consistently neutral, transparent loudspeakers, which can display similar timbral excellence to professionals and consumers alike.

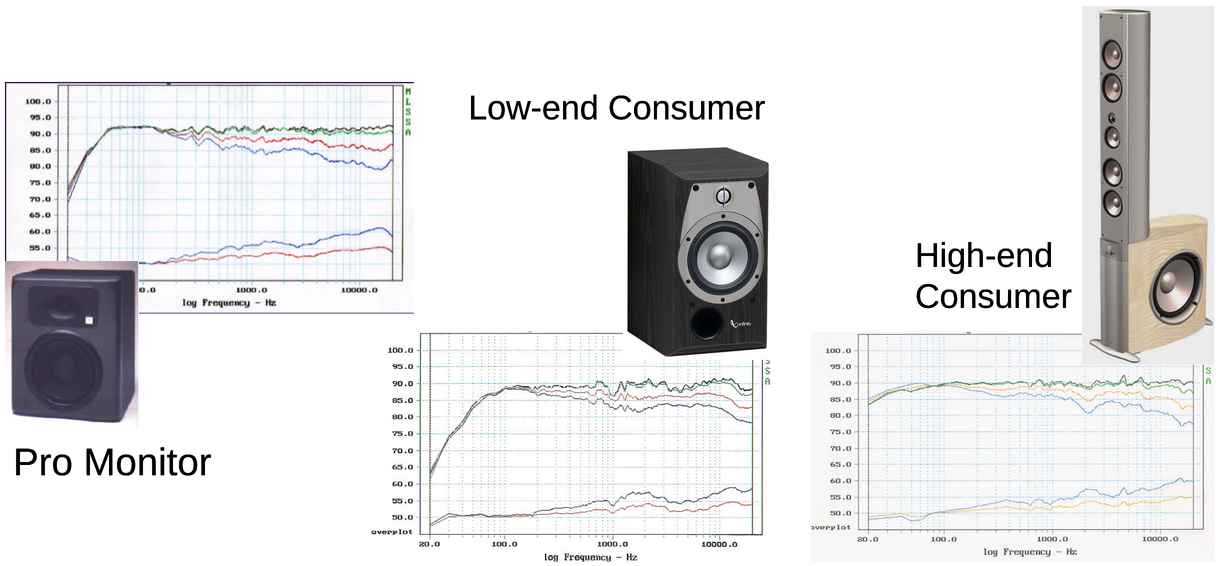


Figure 14. Anechoic data on three examples of loudspeakers that exhibit compatible levels of timbral accuracy, at very different prices and aimed at very different customers.

The traditional problems with inconsistent bass performance in rooms can now be addressed with multiple subwoofers and signal processing that can deliver excellent bass to one or several listeners, again similarly well in different rooms.

The fact that consumers are treated to better sound than ever before, even in their cars, relieves the recording industry of the need to monitor mediocrity. Hotel room clock radios, however, remain a stubborn reminder of how bad audio can be.

Nevertheless, preserving the art in audio recordings is a possibility. We need now to increase the probability that it will happen. That will require a moderate amount of time, effort and money to apply the science we have to establish quality monitoring circumstances. It will also require the courage to confront the past for what it is and, perhaps, to break some bad habits.

10 REFERENCES

1. F.E. Toole, "Loudspeaker Measurements and Their Relationship to Listener Preferences", J. Audio Eng, Soc., vol. 34, pt.1 pp.227-235 (1986 April), pt. 2, pp. 323-348 (1986 May).
2. A. Devantier, "Characterizing the Amplitude Response of Loudspeaker Systems", 113th Convention of the Audio Eng. Soc., Preprint No. 5638 (2002 Oct.).
3. F.E. Toole, "Subjective Measurements of Loudspeaker Sound Quality and Listener Performance", J. Audio Eng. Soc., vol 33, pp. 2-32 (1985 January/February).
4. S.E. Olive, "Differences in Performance and Preference of Trained versus Untrained Listeners in Loudspeaker Tests: A Case Study", vol.51, pp. 806-825 (2003 Sept.).
5. F.E. Toole, "Listening Tests, Turning Opinion Into Fact", J. Audio Eng. Soc., vol. 30, pp. 431-445 (1982 June).
6. F.E. Toole and S.E. Olive, "Subjective Evaluation", in J. Borwick, ed. "Loudspeaker and Headphone Handbook - Third Edition", (Focal Press, London, 2001).
7. F.E. Toole, "Listening Tests - Identifying and Controlling the Variables", Proceedings of the 8th International Conference, Audio Eng, Soc. (1990 May).
8. S. E. Olive, "A Method for Training of Listeners and Selecting Program Material for Listening Tests", 97th Convention, Audio Eng. Soc., Preprint No. 3893 (1994 November).
9. F.E. Toole and S.E. Olive, "The Modification of Timbre by Resonances: Perception and Measurement", J. Audio Eng, Soc., vol. 36, pp. 122-142 (1988 March).
10. Mäkivirta and Anet, "The Quality of Professional Surround Audio Reproduction – A Survey Study", 19th Conference, Audio Eng. Soc. (2001 June).
11. F.E. Toole, "Loudspeakers and Rooms for Stereophonic Sound Reproduction", Proceedings of the 8th International Conference, Audio Eng, Soc. (1990 May).
12. T. S. Welti, "How Many Subwoofers are Enough?", 112th Convention, Audio Eng. Soc., Preprint No. 5602 (2002 May). Also at www.harman.com under 'white papers': "Subwoofers: Optimum Number and Locations".
13. F.E. Toole, "Part 3 - Getting the Bass Right", under 'white papers' at www.harman.com.
14. T.S. Welti and A. Devantier, "In-Room Low Frequency Optimization", 115th Convention, Audio Eng. Soc., Preprint No. (not issued yet), (2003 October).
15. S.E. Olive and F.E. Toole, "The Detection of Reflections in Typical Rooms", J. Audio Eng, Soc., vol. 37, pp. 539-553 (1989 July/August).
16. R. Walker, "A Controlled-reflection Listening Room for Multi-Channel Sound", 104th Convention Audio Eng. Soc. Preprint 4645 (1998 May).