THE AUDIO INDUSTRY : THE STATE OF OUR SCIENCE AND ART

Floyd E. Toole
Vice President Acoustical Engineering, Harman International Industries, Inc., 8500 Balboa Blvd.
Northridge, CA, 91329 U.S.A. ftoole@harman.com

1. INTRODUCTION

Music and movies are art. Audio is a science. Using science in the service of art is the essence of the audio industry. There is a substantial, and still growing, foundation of scientific knowledge behind most of the audio products with which we are familiar, in spite of some glaring exceptions in the marketplace. Some of the old myths persist. In understanding the psychoacoustic portion of the science, loudspeakers have been especially troublesome. These electromechanical devices are required to operate over 10 octaves, with a dynamic range of 105 dB or more, delivering sound to our ears through rooms that are completely unpredictable in their size, shape, layout and acoustical characteristics. It seems like an almost hopeless task yet, through a combination of factors, including good engineering, psychoacoustic knowledge, and human adaptation, we manage to derive substantial satisfaction from our audio systems. And they continue to get better.

In my years at the National Research Council, I worked on providing some of the answers to the underlying questions[1,2]. Others have contributed more data, to the point where, now, we can say that truly good sound reproduction is no longer a matter of chance.

2. SEPARATING THE VARIABLES

At this stage, electronic devices, including the better storage media, are – or can be - essentially transparent. Assuming that a perfect voltage waveform is delivered to the terminals of the loudspeaker from a low impedance source, the next challenge is to minimize all audible linear and non-linear distortions in the transduction process. Then we optimize form of the radiated sound field, bearing in mind the physical nature of the listening environment, and find ways of taming the prominent and lively resonances of small rooms. Let us look at some of the major variables in this complicated picture.

2.1 Non-linear distortion. Simultaneous masking by the audio signal itself prevents much of the perception of non-linear distortions. Not all, of course, and not all of it is bad. At low frequencies significant numbers of listeners react positively to a little added timbral 'richness'. Good conventional engineering can reduce non-linear distortions to acceptable levels. Exceptions are usually the result of compromises driven by cost.

2.2 Linear distortions. Conventional engineering principles encourage maintaining the integrity of both amplitude and phase in the complex transfer function. This preserves waveform information. However, abundant psychoacoustic evidence tells us that we humans are substantially 'phase deaf', especially when listening in normally reverberant spaces. Even in circumstances where one may hear a difference, allocating a preference is difficult. We encounter these situations regularly in normal listening, whenever the direct sound from a source is modified by the addition of one or more strong reflections. We routinely can recognize differences but, since we know the source is unchanged, we do not assign preferences. If waveform information were critical to sound fidelity, where would one place a microphone to capture the definitive waveform of a grand piano? There is no such waveform.

In contrast, humans are remarkably sensitive to very small changes in frequency response or spectrum. In terms of overall loudness, changes of program level less than about 1 dB are normally inaudible. However, the threshold of detection for a spectral tilt is about 0.1 dB/oct., a $Q=1$ resonance can be detected when it adds about 0.3 dB to an otherwise flat frequency response. Narrower bandwidth, higher Q, spectral changes are less easily heard, with $Q=50$ spikes reaching 10 dB before arousing our conscious reactions with certain kinds of music. Ringing resonance decays tend to be audible only at very low frequencies. If this all seems countintuitive, consider also that our sensitivity to medium- and low-Q resonances is lowest in anechoic conditions, and increases when we listen in a reverberant space! A concert in the park is timbrally enriched when rain drives the orchestra into the community hall. Loudspeakers sound less colored in anechoic chambers than they do in normal rooms [3].

In terms of technical measurements, all of this argues against gross $\pm x$ dB tolerances on frequency response curves, unless the tolerance is very small : e.g. $\pm$ 0.5 dB. A perceptual criterion requires that the allowable tolerance be related to the bandwidth of the deviation. Oh yes, and 1/3-octave resolution is woefully inadequate when it comes to describing what might or might not be audible as a timbral difference. Critical bands apply to loudness summation, not the perception of timbre.

2.3 Frequency response and directivity. So, we can hear very small differences in amplitude response. What, then, is the target curve from which deviations are assessed? Is 'flat' the ideal? For electronic devices it clearly is. For loudspeakers, it depends on what is being measured. Loudspeakers radiate a three-dimensional sound field. In rooms all of this sound reaches listeners, most of it after one or more reflections from room boundaries and furnishings. To evaluate the performance of a loudspeaker it is necessary to collect enough data to be able to reconstruct the major features of the sounds arriving at a listener in a room. Figures 1 thru 3 illustrate the essence of the loudspeaker/room interface problem.

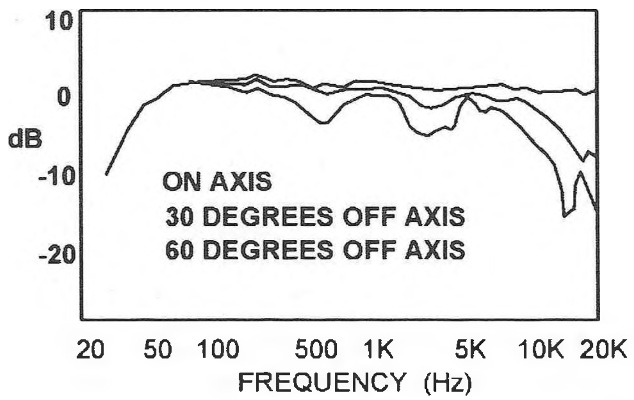

Fig.1 Anechoic frequency responses of a loudspeaker showing (top to bottom) a very smooth, flat, axial response and progressive deterioration off axis.

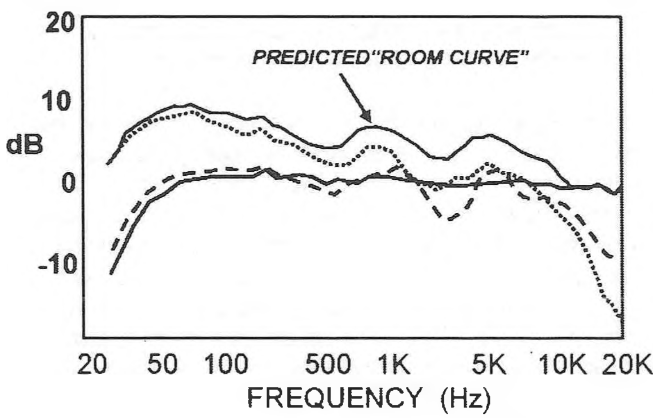

Fig. 2 The sequence of sounds arriving at a listener in a room. The first arrival is the direct sound (the solid line around 0 dB). Second is the sum of adjacent boundary reflections (dashed). Third is the reverberation, represented by total sound power (dotted). The solid curve plotted over them all is the energy sum of all three – a prediction of what might be measured in a real room.

It is evident from this that sound power is the dominant factor at low frequencies, and that the direct sound is dominant at the very highest frequencies. In between, over most of the frequency range, everything contributes. So, if the purpose of the measurements is to be able to anticipate loudspeaker performance in a room, it is necessary to measure everything. The on-axis response, by itself, is merely a start. The sound power is also incomplete evidence. All of it must be viewed as an ensemble.

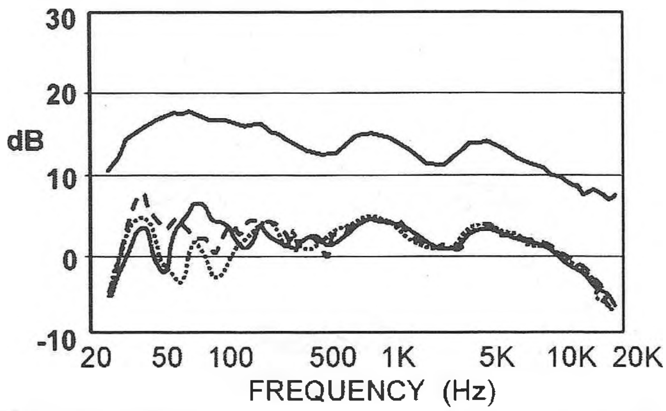

Fig. 3 The three bottom curves show steady-state measurements of the example loudspeaker in three typical locations in a normal room. The top curve is the predicted curve from Fig.2, raised 10 dB for clarity.

Fig.3 shows that, at frequencies below about 300-500 Hz, room resonances increasingly dominate what we hear. The differences are not subtle and the situation can be evaluated only by measurements in the room itself. However, above 300-500 Hz it is possible to predict with good accuracy what it delivered to the listening position from a collection of anechoic measurements that have been appropriately processed. Steady-state measurements in a room are reliable at low frequencies, but at middle and high frequencies they are useful only in conjunction with comprehensive anechoic data. In this example, the undulations at middle and high frequencies are in response to the frequency-dependent directivity of the loudspeaker, so it means that any attempts to change the shape of the room curve by equalization will, in fact, not correct the problem. The only solution to this kind of problem is a loudspeaker designed with directivity that is constant, or relatively so, over most of the frequency range. Only then will the direct, early reflected, and reverberant sounds convey similar timbral messages to the listeners [4].

It is a nice story, but do listeners agree? Yes. Hundreds of double-blind subjective evaluations, conducted over the past 20+ years, confirm that these are the loudspeakers that listeners award with the highest ratings. To get consistent opinions from listeners, however, it is necessary to deal with the huge variability at low-frequencies caused by standing waves and loudspeaker and listener locations. For the purposes of achieving consistency in listening tests, standardizing the room and locations is a practical solution. We currently use a pneumatic 'speaker shuffler' to achieve a consistent location for the active loudspeakers in listening tests [5]. Delivering consistently good sound to listeners in their homes is a much more formidable challenge.

3. AN AUDIO INDUSTRY PROBLEM

Audio enthusiasts tend to take for granted that everything upstream of the playback device is under control. The sad fact is that it isn't. Many factors in the sequence of events leading to a music recording or a film sound track contribute to systematic and random variations in the final product. Not the least of these are the humans involved in the process, but physical factors also have a say.

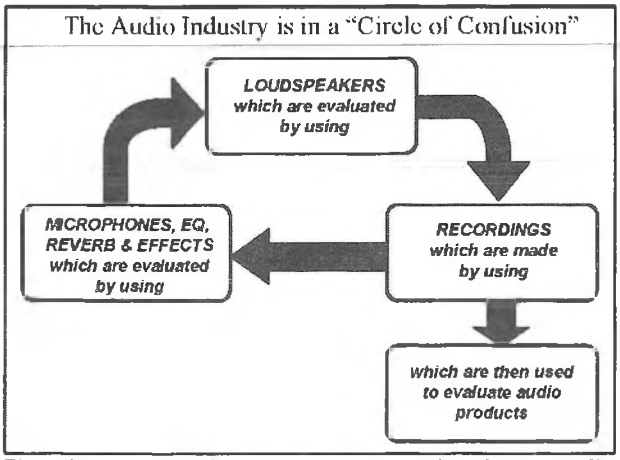

Fig. 4 Loudspeakers in rooms are the means by which recordings are judged while being made. They are the 'window' through which the art is viewed. If the window is colored or distorted, the art will be adjusted to compensate. The compensated art is then used for enjoyment (through different loudspeakers and rooms) or, worse, used as a basis for evaluations of other audio products.

Fig. 4 shows that our audio industry is trapped in a "circle of confusion" that can only be broken if there is a reliable similarity between the loudspeakers and rooms used in the production of the art, and those used during playback for our entertainment. Such consistency requires accurate technical measurements that show good correlation with listener preferences. These now exist.

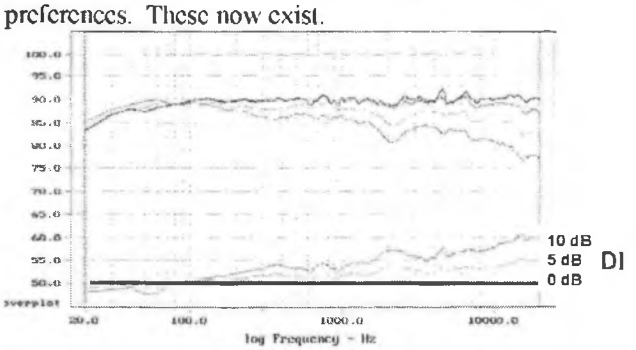

Fig. 5 Seventy-two anechoic measurements, made on horizontal and vertical orbits around a loudspeaker, processed to show (top to bottom): on-axis (direct sound), Average response within a ± 10° vertical, ± 30° horizontal listening window, estimated energy sum of the first six reflected sounds in an average room, and total sound power. At the bottom are the traditional directivity index (top) and an invented one for early reflections only. The measurements have 1/20-octave resolution.

Fig. 5 shows a form of measurements that has been found to correlate well with listener opinions as expressed in double-blind evaluations in normal rooms. The product described here is a good representation of the state of the art in loudspeakers today. It costs $10K/pr. (USD). In subjective evaluations of the best loudspeakers, it is common for the largest sources of judgment variation to be the recordings themselves, and individual differences among listeners.

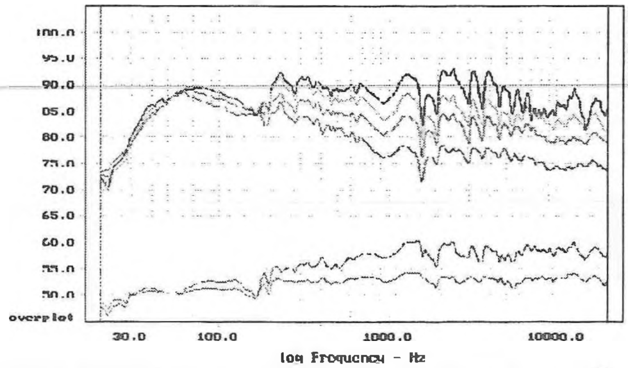

Fig. 6 A $11K/pr. product that was compared to the one in Fig. 5.

However, individual preferences are a factor only when the contests are very close. The products of Figs. 5 and 6 were evaluated by 124 unselected, untrained, listeners. All but one put this loudspeaker in a strong second place, and this person's preference was not statistically significant. There are two important lessons here: (1) most people DO agree on what is good if they are given an unbiased opportunity to judge, and (2) price is an unreliable indicator of sound quality (so, also, are the reviewers who raved about the product in Fig. 6). Our routine listening evaluations use persons selected for normal hearing and put through training to identify those with the necessary aptitudes (most people) and to increase their skills in detecting common faults and articulately describing what they hear. They become remarkably stable 'measuring instruments'[6].

The example shown in Fig. 6 is not uncommonly bad, but the good news is that more and more loudspeakers are emulating the performance of Fig. 5, even at affordable prices. Sacrifices at lower prices include the lowest bass frequencies, the ability to play cleanly at high sound levels, and visual aesthetics. The other good news is that there are a few professional studio monitor loudspeakers that bear comparison with Fig. 5. The standards within the audio industry, both consumer and professional, are rising.

Still, recordings remain more variable than we would like. In examining the loudspeakers used as professional monitors, one finds a range of sound quality only slightly less than that in the consumer domain. This is regrettable. However, some of the variable art comes from sources using the same good loudspeakers, so what is wrong? A recent investigation surveyed a large number of recording studios that used the same family of loudspeakers. Measurements were made at the head location of the recording engineers, and the data were compiled. The results were frightening.

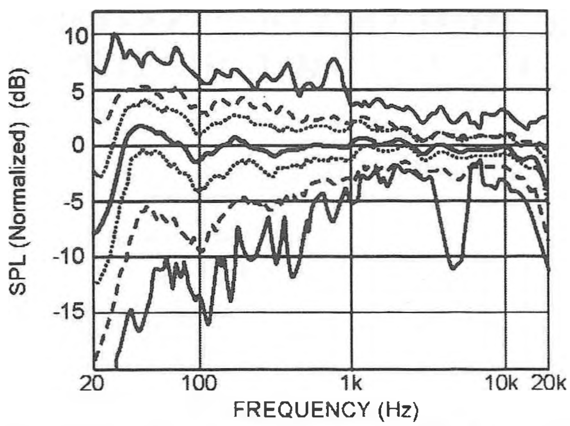

Fig. 8 The solid curve approximating the 0 dB line is the median of 250 measurements. This is the good news. The solid curves at the top and bottom represent the max/min limits of all of the measurements. The dashed curves show the upper and lower limits of 90 % of the measurements, and the dotted curves show the same for 50% of the measurements [7].

The picturesque character of the 250-curve median displays the ability of statistics to shield us from the truth. Only when we see the huge variations extant in individual studios do we see reasons why artists can be deceived about what is going into their master tapes and discs. Looking further, the largest variations are in the low frequency range where the room is in control. The majority of installations show relatively good performance in the middle and upper frequencies where the loudspeaker dominates. Correcting this situation, and the parallel situations in all of our homes, means coming to grips with room resonances, methods of measurement and equalization, and some new ways to employ multiple subwoofers in acoustic mode-canceling arrays.

A casual glance at a high-resolution curve measured in a small room suggests that perfection may be forever elusive. However, at low frequencies the dimensions of the standing waves are generous, the events are less numerous, the identification of individual resonant modes is possible, and solutions begin to present themselves. Damping with mechanically- or acoustically-tuned absorbers (or flexible walls) reduces the standing wave peak-to-trough ratio, producing more uniform bass over larger areas of the room. Equalization for specific listening locations is possible, so long as one addresses only the peaks of the resonant modes. Low-frequency room modes behave as minimum-phase systems and flattening the frequency response also eliminates the time-domain ringing. Doing this successfully requires high-resolution measurements to show the true center frequency and Q of the resonances, and parametric equalizers to match the shapes. Traditional 1/3-octave measurements and equalizers are not adequate.

Now that multichannel audio allows us to treat the frequency range below 80 Hz with dedicated subwoofers, they can be located to optimize bass performance, and it is possible to use multiple subwoofers to destructively drive modes or drive them at their pressure minima. For example, two woofers located at the 25% and 75% points across the 20-foot width of a room will seriously attenuate all width modes below 80 Hz. Many of us have probably done this accidentally in stereo setups, and it works because most low bass is monophonic.

This principle can be extended to a set of general solutions for rectangular rooms, where the objective is to minimize the variations in low-frequency performance over the central portion of a room, where several persons can enjoy multichannel performances of music or movies. It turns out that more than one subwoofer is needed, but more than four are not advantageous. The best of the practical arrangements are four subwoofers in the corners, in the mid-wall locations, or the 25% and 75% locations on front and back walls. Two in opposite mid wall locations are almost as good. Once relatively uniform performance is achieved over an area, intelligent equalization can be applied [8].

So, if we diligently apply the existing science, it can yield more consistently good recordings, and it can ensure that the finer qualities of the audio arts can be reliably delivered to our homes. Science, truly, can be used in the service of art.

REFERENCES

1. F.E. Toole, "The Acoustics and Psychoacoustics of Loudspeakers and Rooms – The Stereo Past and the Multichannel Future", invited paper at the 109th Convention, Audio Eng. Soc., Preprint no. 5201 (2000 Sept.).
2. F.E. Toole, "Loudspeakers and Rooms for Stereophonic Sound Reproduction", Proceedings of the 8th International Conference, Audio Eng, Soc. (1990 May).
3. 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).
4. 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).
5. F.E. Toole and S.E. Olive, "Subjective Evaluation", in J. Borwick, ed. "Loudspeaker and Headphone Handbook - Third Edition", (Focal Press, London, 2001).
6. 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).
7. Mäkivirta and Anet, "The Quality of Professional Surround Audio Reproduction – A Survey Study", Proc. 19th Audio Eng. Soc. Conference (2001 June).
8. T. Welti, "How Many Subwoofers are Enough?", 112th Convention, Audio Eng. Soc., Preprint No. (2002 May)
Relevant technical papers are also downloadable from www.harman.com, under "white papers".