Audio - science in the service of Art

by Floyd E. Toole, Ph.D.
Vice President Acoustical Engineering, Harman International Industries, Inc.
8500 Balboa Boulevard, Northridge, CA 91329 www.harman.com

Audio is science

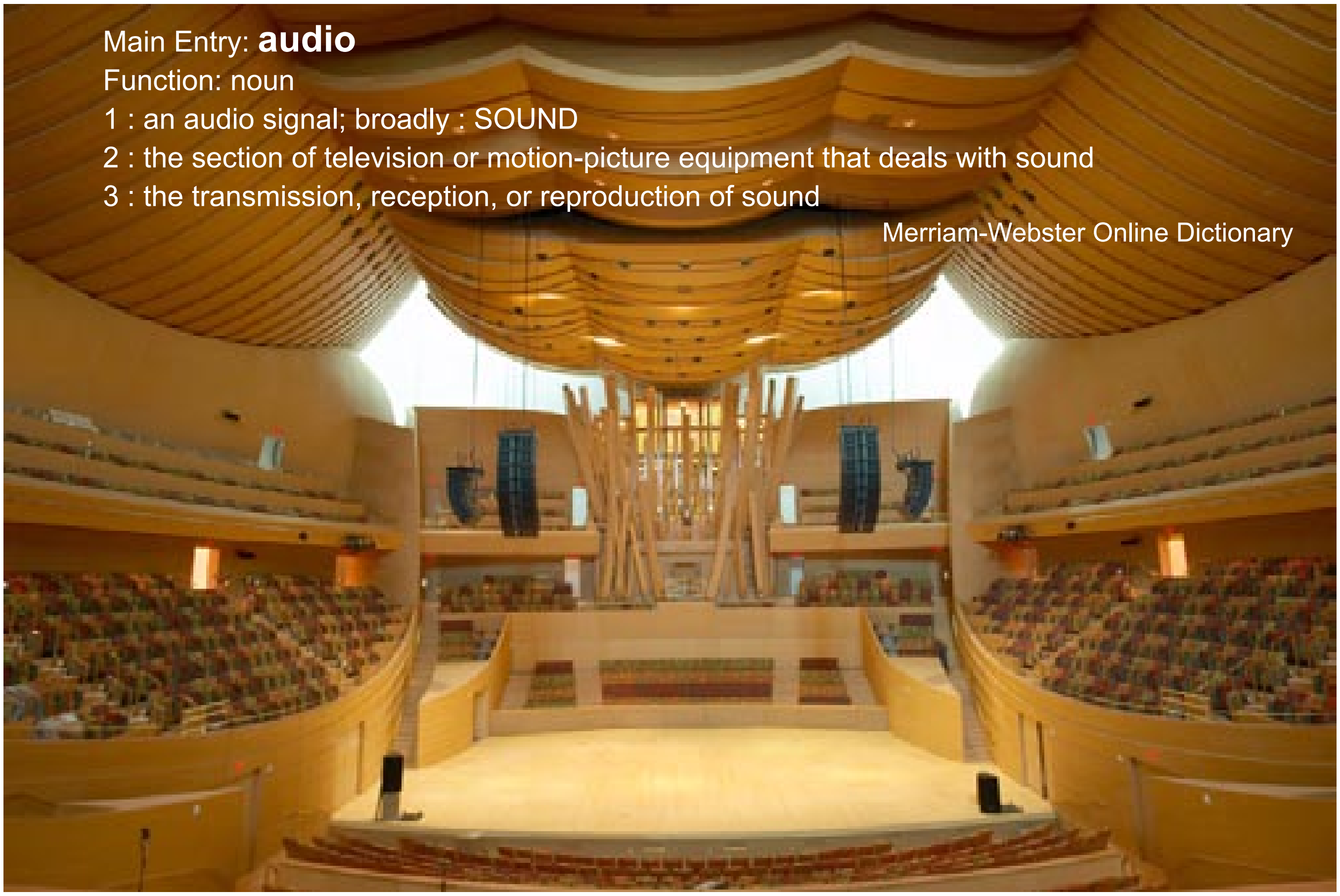
All of the devices used to capture, store and reproduce musical, theatrical and film performances are designed by engineers who apply the laws of physics as they pertain to the functioning of these devices. Perfection, in this context, is defined as not degrading any part or any aspect of the audio signal: the art itself. Determining how close we are to perfection is ultimately determined subjectively, in carefully-controlled listening tests. However, accurate and comprehensive measurements, interpreted using the ever-expanding ‘rules’ of psychoacoustics, are remarkably accurate predictors of what is and what is not audible and, ultimately, of subjective preferences. This allows us to design better sounding and more cost effective products.

Audio is art

The melodies and rhythms of music, the nuances of tone and inflection of instruments and voices, the emotive effects of sometimes subtle, sometimes explosive, sound effects in movies … are aspects of art for which there are no scientific or technical measures. Yet, subjectively we have little difficulty describing our reactions, positive and negative, when we hear these sounds. Perfectly reproduced, they should sound exactly as they did at the live performance or in the control room where the recording was mixed. Achieving this lofty goal is a considerable challenge, since the final audio component is the listening room, and all rooms are different. The interactions among loudspeakers, listeners and rooms are the focus of continuing scientific investigations.

Infinity Prelude MTS

Revel Salon



Sound in transition. The magnificent new Walt Disney Concert Hall, in Los Angeles, where live music is performed traditionally, without electronic aids, is shown being outfitted with JBL loudspeakers for performances in which the artists are reinforced or supplemented by recorded sounds. The line between live and reproduced sound has become blurred as technology adds more dimensions to the artistic palatte.

INTRODUCTION

We expect quality audio products to sound good. However, the determination of what constitutes "good sound" has been controversial. Some claim that it is a matter of personal taste, that our opinions of sound quality are as variable as our tastes in "wine, persons or song", and that the only opinion that truly matters is one's own. This would place audio manufacturers in the category of artists, trying to appeal to a varying public "taste". Others, like the author, take the view that artistry is the domain of the instrument makers and musicians and that it is the role of audio devices to capture, store and reproduce their art with as much accuracy as technology allows. The audio industry then becomes the messenger of the art – science in the service of art.

Interestingly, this process has created new "artists", the recording engineers, who freely editorialize on the impressions of direction, space, timbre and dynamics of the original performance. In fact, in studio creations, the original performance really occurs in the recording control room, through loudspeakers. Other creative opportunities exist at the point of reproduction, as audio enthusiasts tailor and tweak the sound in listening rooms by selecting loudspeakers of differing timbral signatures and directivity, and by adjusting the acoustics of the listening space with special acoustical devices. In an industry that lacks meaningful standards for recording or playback, why not?

To design audio products, engineers need technical measurements. Historically, measurements have been viewed with varying degrees of trust. However, the value of measurements has increased dramatically as we have found better ways to collect data, and as we have learned how to interpret the data in ways that relate more directly to what we hear. With measurements we can set objectives, telling us when we are successful. Some of these design objectives are very clear, and others still need better definition. All of them need to be moderated by what is audible. Imperfections in performance need not be immeasurably small, but they should be inaudible. Achieving this requires knowledge of psychoacoustics, the relationship between what we measure and what we hear. This is a work in progress but, right now, we know a great deal.

Since loudspeakers, listeners and rooms interact within a complex acoustical system, it is not possible to consider one without the others. What we hear is a combination of the loudspeaker and the room, and what we measure, and can see with our eyes, is not 'linearly' related to what we hear. Psychoacoustics and acoustics are tightly interwoven.

ART, SCIENCE AND AUDIO

Music, itself, is art, pure and simple. The composers, performers and the creators of the musical instruments are artists and craftsmen. Through their skills, we are the grateful recipients of sounds that can create and change moods, that can excite and animate us to dance and sing, and that form an important component of our memories. Music is part of all of us and of our lives.

$E=mc^2 \pm 3$ dB

Here Einstein's consequential and far-reaching equation, the most famous in all of physics, is modified by a tolerance that has its origins in what someone thought would be subjectively acceptable in systems for reproducing music. I saw this first many years ago, was amused by the juxtaposition, and never forgot it.

In spite of its many capabilities, science cannot describe music. The crude notes on a sheet of music provide a basic description to a musician, but science has no technical measures for the evocative elements of a good tune or good musicianship. It cannot, with numbers or graphs, describe why Pavarotti's tenor voice is so revered, or why a Stradivarius is held as an example of how violins should sound. Those are distinctions that must be made subjectively, by listening. The determination of what is aesthetically pleasing remains firmly based in subjectivity. The requirements for accurate *reproduction* of the sound creating the aesthetic experience, though, must eventually have a purely technical description – an important distinction. Right now, we rely on a combination of subjective and objective data.

Our audio industry is based on a sequence of events. We capture a musical performance with microphones, whose outputs are blended into an electronic message traditionally stored on tape or disc, which is subsequently amplified and reproduced through loudspeakers. This simple description disguises a process that is enormously complicated. We know from experience that, in some ways, the process is remarkably good. For decades we have enjoyed reproduced music of all kinds with fidelity sufficient to, at times, bring tears to the eyes, and send chills down the spine. Still, critics of audio systems can sometimes point to timbral characteristics that are not natural, that change the sound of voices and instruments. They hear noises and distortions that were not in the original sounds. They note that closing the eyes does not result in a perception that the listener is involved in the performance, enveloped in the acoustical ambiance of a concert hall or jazz club. They point out that stereo is an antisocial system – only a single listener can hear the reproduction as it was created.

For all of these criticisms there are solutions, some here and now, and some under development. All of the solutions are based on science.

How can science, a cold and calculating endeavor if ever there were one, help with delivering the emotions of great music? It is because, in the space between the performers and the audience, music exists as sound waves. Sound waves are physical entities, subject to physical laws, amenable to technical measurement and description and, in most important ways, predictable. The physical science of acoustics allows us to understand the behavior of sound waves as they travel from the musician to the listener, whether the performance is live or recorded.

To capture those sound waves, with all of the musical nuances intact, we need transducers to convert the variations in sound pressure (the sound waves) that impinge on them into exact electrical analogs. Microphone design uses *electroacoustics,* a blend of mechanical and electrical engineering sciences conditioned by an understanding of sound fields in rooms. Once the signal is in the electrical domain, expertise in *electronics* aims to preserve the integrity of the musical signal. Contamination by noise and distortions of all kinds must be avoided, as it is prepared for storage, and when it is recalled from storage during playback.

Emerging from the storage or broadcast medium the musical signal is too feeble to be of any practical use, so we amplify it (more *electronics*), giving it the power to drive loudspeakers or headphones (more *electroacoustics*) that convert the electrical signal back into sound. Again, this all must be done without adding to or subtracting from the signal, or else it will not sound the way it should.

Finally, the sound waves radiating from the loudspeakers propagate through the listening room to our ears. And this is where our tidy process goes astray. Rooms in our homes are different from any that would have been used for live performances, or for monitoring the making of a recording. Our eyes and our ears both recognize that. The acoustical properties of rooms, large and small, are in the scientific domain of *architectural acoustics.*

> … reproductions of … recorded musical performances should ideally be independent of the listening room

In live musical performances, the room – concert hall, jazz club, etc. – is part of the total event. Good recordings try to capture and convey this acoustical ambiance as a setting for the music. In contrast, reproductions of those recorded musical performances should ideally be independent of the listening room, so that what we hear is not changed by where we happen to be, whether that is a recording control room or a domestic listening space. The complex interactions between room boundaries and speaker directivity at middle and high frequencies, and speaker and listener position at low frequencies, are powerful influences in what we hear. With a combination of acoustical design, loudspeaker configurations and electronic signal manipulations, we are finding ways to make speakers more "friendly" to rooms, both in recording studios and in homes.

THE STEREO PAST AND THE MULTICHANNEL FUTURE

In stereo there are only two loudspeakers to reconstruct an illusion of the complex three-dimensional sound field that existed in the original hall, club or studio. The choice of two channels was based on the limitations, back in the 1950s, of what could be stored in the groove of an LP. Even at that time, it was known that two channels were insufficient to recreate truly convincing illusions of a three-dimensional acoustical performance.

The excitement and realism of the reproduction is enhanced if we have more channels and more loudspeakers. Multichannel audio is familiar in movies, but it is new to mainstream music recording and, understandably, it will take some time for artists and recording engineers to learn how to fully exploit the new formats. Skillfully used, the combination of digital discrete multichannel technology, and a competently set up loudspeaker/room combination can provide the basis for thrilling sound experiences whether the objective is realism or pure abstract artistry. As gratifying as concert hall performances and two-channel stereo have been, we now have an expanded spatial and directional palette for musical artists to play with and within. Inevitably, there will be debates about good taste and individual preferences, but this is art and anything goes. Just don't shoot the messenger (multichannel audio) if you don't like the message.

Two-channel stereo, as we have known it, is not a "system" of recording and reproduction.

Two-channel stereo, as we have known it, is not a "system" of recording and reproduction. The only "rule" is that there are two channels. At the recording end of the chain, there are many quite different methods of miking and mixing the live performance, ranging from the purist simplicity of two coincident microphones to multi-microphone, multi-track, pan-potted and electronically-reverberated mono. They are all different. It needs to be remembered that whatever ends up in a recording has been listened to – and approved of – by listening through the loudspeakers in the control room of the recording studio. The reality is that, whether it is a massive professional installation or a home studio, in terms of sound quality they run the gamut from superb to truly mediocre. No wonder recordings are variable.

At the reproduction end of things, loudspeakers used for stereo have taken many forms: forward facing, bipolar (bidirectional in-phase), dipolar (bidirectional out-of-phase), omnidirectional, and a variety of multi-directional variants. These, and various sum/difference and delay devices, were developed over the years in attempts to coax, from a spatially-deprived medium, a rewarding sense of space and envelopment. The principle at work here is that the reflected sounds in the listening room add to the sounds coming directly from the two loudspeakers, enhancing the impressions of acoustical warmth, depth, air and spaciousness. Of course the room now becomes an important part of the playback process and the one thing we know about rooms is that they are all different - one more significant variable in the stereo "system".

In the two-channel world, therefore, the artists *could not* anticipate precisely how their performances would sound in homes.

In the two-channel world, therefore, the artists *could not* anticipate precisely how their performances would sound in homes. It was left to the end user to create something pleasant. Stereo, therefore, is not an encode/decode system, but a basis for individual experimentation. The fact that both audiophiles and the audio industry have thrived in spite of this situation is a tribute to the power of human adaptation. Given time, we can come to believe that many different variations on the truth are apparently equally entertaining.

Nowadays we go even further, using active-matrix technologies to convert conventional stereo recordings into multichannel performances reproduced through 5, 6 or 7-channel systems. Prominent among these is Harman's excellent Logic 7 system, developed by Lexicon, and available in Harman/Kardon, JBL, and Mark Levinson consumer products and premium automotive audio systems supplied directly to several car manufacturers. Dolby's ProLogic IIx and DTS Neo:6 are other such systems. These are not arbitrary 'sound everywhere' surround sound systems. There is a serious effort to keep the musicians up front, and the frontal soundstage intact, but ambience, reverberation and other ambiguously localized sounds are separated out and sent to the surround channels. Whether the experience is realistic, or even pleasant, will depend on how the recording was made – remember there are no standards in stereo. However, most recordings translate beautifully into a multichannel experience that makes the original two-channel rendering sound flat and lifeless. Some recordings, though, don't respond well to this kind of post-processing, and sound better in the original stereo mode. You have to try it and see.

All of this is much improved when recordings are created for reproduction through multiple discrete channels. Then, if we do not like what we hear, we can legitimately blame the artists involved in the process. Multichannel recordings are created with the knowledge that there are loudspeakers positioned around the room in locations optimized to create different directional and spatial illusions. This is a powerful advantage, in that it permits the artist to create an enormous range of effects that can include some highly localized sounds coming from specific channels/loudspeakers and other, more ambiguously localized sounds, such as the comfortable reverberation of a concert hall.

Adding unfortunate confusion to the present situation are fundamental differences in the methods used by the film and music industries to record and reproduce multichannel audio.

Good sound is critical to the success of a movie. However, it is clear that sound is subordinate to the picture in the sense that only rarely do strongly localized sounds appear to come from directions far away from the screen. Typically, one hears momentary sounds like gunshots, ricochets, aircraft fly overs, door slams, etc. In general, the surround channels are used for "atmospheric" sounds, reverberation, and ambiguously localized music. For this reason, in cinemas the surround sounds are presented by several loudspeakers located down the sides and across the back of the auditorium. In home theaters that have been designed primarily for movies, it is common to use multidirectional loudspeakers (e.g. bipoles or dipoles) well above ear level on each of the side walls for the surround channels, creating a lot of reflected sounds, thus 'softening' the localization effects of the surround channel loudspeakers. Bass almost always would be provided by a bass-managed subwoofer channel.

For music, however, if we were to follow the lead of certain studios, we would be using five identical full-range loudspeakers, and some recording engineers would be encouraging us to place the surround speakers behind us, on the back wall (a lingering effect of quadraphonics?). These approaches lead to very different listening experiences compared to music reproduced through systems configured for movies. Surround loudspeakers located beside, and slightly behind, the listeners are able to generate the highly desirable sense of envelopment that makes live performances in halls and clubs so distinctly pleasurable. Placing them totally behind, on the back wall, compromises that, but allows for more dramatic front/back localization contrasts in pop music. What we want and

Adding unfortunate confusion to the present situation are fundamental differences in the methods used by the film and music industries to record and reproduce multichannel audio.

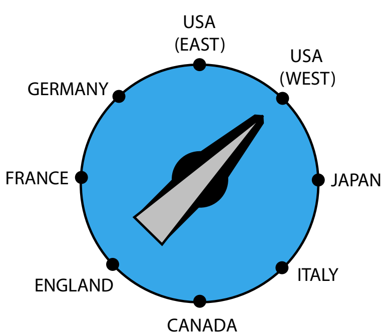
need is to be able to do both, simultaneously. This is why we now have 6.1 and 7.1 channel systems. Ironically, it is films that use the additional channels, while music is sticking to 5.1 channels. Clearly there are concerns about public acceptance of the increased number of channels, but recordings can be made for 7.1 channels that would gracefully degrade to 5.1, or even stereo, for those customers with basic systems.

Multichannel sound should not be an impediment to the arts. It should be a liberator. It would be wonderful if the film and music industries could agree on a standard methodology for multichannel sound. However, based on experience, that is unlikely. Pity.

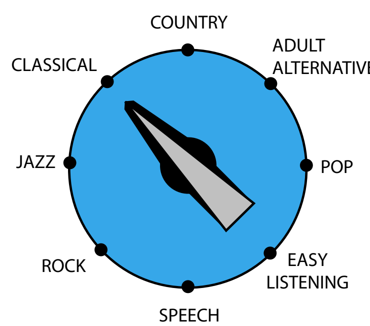
THE SCIENCE OF AUDIO

The scientific method requires measurements and, in audio, we do two kinds: subjective and objective. Then we enter the domain of *psychoacoustics*, the study of relationships between physical sounds and the perceptions that result from them. Psychoacoustics allows us to understand and interpret measurements in ways that relate to what we hear. However, it is all based on the premise that human listeners agree on what is, and is not, good sound.

Individual points of view are a part of human nature. They enrich our lives in countless ways. The world would be a boring place if we were all attracted to the same music, food, wine and people. A commonly expressed point of view is that sound also is "subjective", that we all "hear differently", and therefore not all of us prefer the same loudspeakers, amplifiers, etc. It is also alleged that different nationalities, and regions have different preferences in sound. I have always regarded these assertions with suspicion because, if they were true, it would mean that there would be different pianos for each of these regions, different trumpets, bassoons and kettledrums. Vocalists would change how they sang when they were in Germany, Britain, and the U.S. I wonder what Pavarotti's Japanese timbre sounds like? Of course, it doesn't happen that way.



Certain beliefs would have these controls on our loudspeakers. One is as silly as the other. A truly good loudspeaker is just that – no further adjustments are necessary.



The entire world enjoys the same musical instruments and voices in live performance, and the recording industry sends the same recordings throughout the world. True, from time to time, there have been regional influences that have made differences. I can recall a short period of "east coast / west coast" sounds associated with some powerful loudspeaker brands in those locations in the USA. In Britain, the British Broadcasting Corporation (BBC) provided loudspeaker designs that became the paradigm for a few years. Everywhere, there are magazines and reviewers that are influential. All of these factors change with time. In truth, they are really minor variations on a common theme. Underlying it all is a powerful desire to reproduce sounds as accurately as possible. Since there are universal standards of excellence for live sound performances, there should be comparable universal standards for reproductions of that sound.

So, what about those individual preferences in sound quality? This issue was settled by conducting many listening tests, using many listeners and many loudspeakers. To reduce the influences of price, size and style – factors that we absolutely know can reveal individuality – all of the tests were conducted blind. Other physical and psychological factors known to be sources of bias were also well controlled [refs. 1-3]. The results were very clear. When the data were compiled, it turned out that most people, most of the time, liked and disliked the same loudspeakers.

… it turned out that most people, most of the time, liked and disliked the same loudspeakers.

Still, there were differences in the way listeners performed. While most were remarkably consistent in repeated evaluations of the same product, some others changed their opinions of the same product at different listening sessions. Those people could like a product during a listening evaluation, writing verbose descriptions of praiseworthy characteristics, and then, during a later test in which the same product appeared, the results would be dramatically different, with detailed notes describing intolerable problems.

When this puzzling behavior was examined, it was found that the explanation lay in the hearing performance of the individuals. All of these erratic listeners had hearing loss. Listeners with relatively normal hearing tended to be quite consistent in their judgments from session to session.

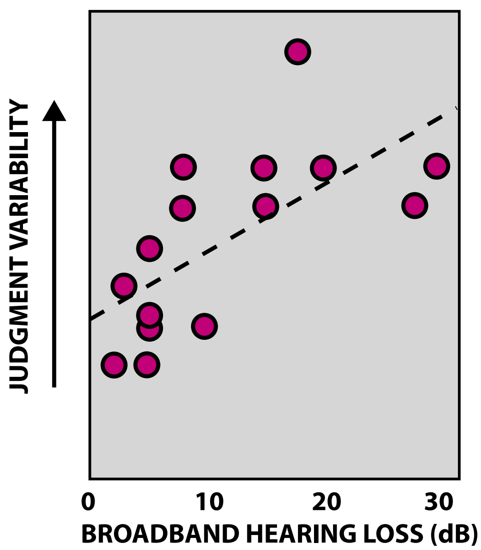


Figure 1 The relationship between judgment variability and hearing loss averaged over the frequency range below 1 kHz. From Ref. 4

This important observation is paralleled by another, perhaps even more important one: that groups of such listeners closely agreed with each other. In other words, viewed in the broad perspective, good sound quality is *not* a matter of individual taste. When dealing with loudspeakers that are similarly good, there will be small biases in the results attributable to the music itself (as it is able to reveal the problems) and listeners (as they reveal their individual tolerances to the problems they hear in the musical selections). However, the delineation of overall sound quality on a scale of good to bad is abundantly clear.

The finding that hearing loss is a factor is not surprising – listeners who cannot hear all of the sound must make less reliable judges. What is surprising is that the deterioration in performance is so rapid. It is well developed in people with hearing losses that would not be regarded as alarming by conventional audiometric criteria, see Figure 1. This difference, one assumes, is related to the difficulty of the task: judging sound accuracy, as opposed to understanding the spoken word. The hearing loss, in this case, was defined by the average threshold elevation at frequencies below 1 kHz. Those exhibiting this form of hearing loss, also tended to have loss at high frequencies. High-frequency loss, by itself, was not a clearly correlated factor. Reference 4 covers this in detail.

Listeners with hearing loss not only exhibit high judgment variability, they can also exhibit strong individualistic biases in their judgments. This comes as no surprise, since such individuals are really in search of a "prosthetic" loudspeaker that somehow compensates for their disability. Since the disabilities vary enormously, so do the biases.

The evidence of Figure 2 is that the group of normal-hearing listeners, who exhibited relatively low variability, substantially agree in their ratings. Interestingly, the high-variability group shares the opinion of the truly good speakers, A and B. However, speakers C and D exhibit characteristics that are viewed as problems by the normal group, but about which the second group has substantially no opinion. Based on the opinions expressed by certain individuals, either C or D could be the best or worst speaker in the group. It appears that their disabilities prevented some of the listeners from hearing certain of the deficiencies. Sadly, listeners in the problem category include some talented and knowledgeable musicians and audio professionals whose vocations may have contributed to their condition. However articulately their opinions are enunciated, their views are of value only to them personally and, possibly, only at that moment.

The conclusion is clear. If there is any desire to extrapolate the results of a listening evaluation to the population at large, it is essential to use representative listeners. In this context, it appears to be adequate to employ listeners with broadband hearing levels within about 20 dB of audiometric zero. According to some large surveys, this is representative of at least 75% of the population – an acceptable target audience for most commercial purposes. This is not an "elitist" criterion.

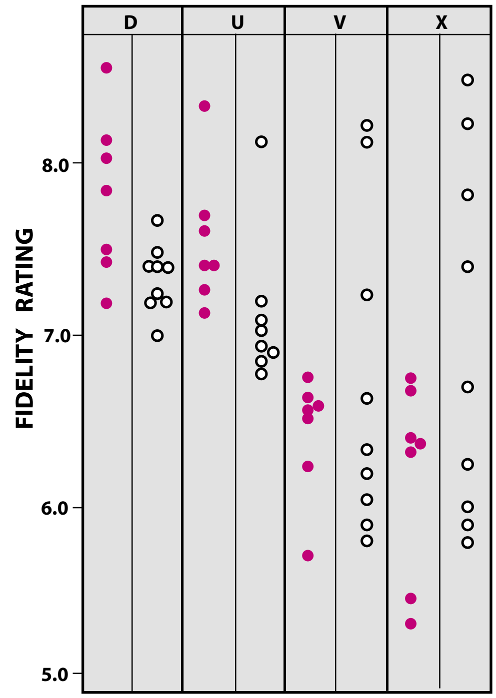


Figure 2: A comparison of FIDELITY evaluations of four loudspeakers performed by two groups of listeners, one with low variability in their judgments (closed circles) and the other with high judgment variability (open circles). From Ref. 4.

PRACTICE MAKES PERFECT – USE TRAINED LISTENERS

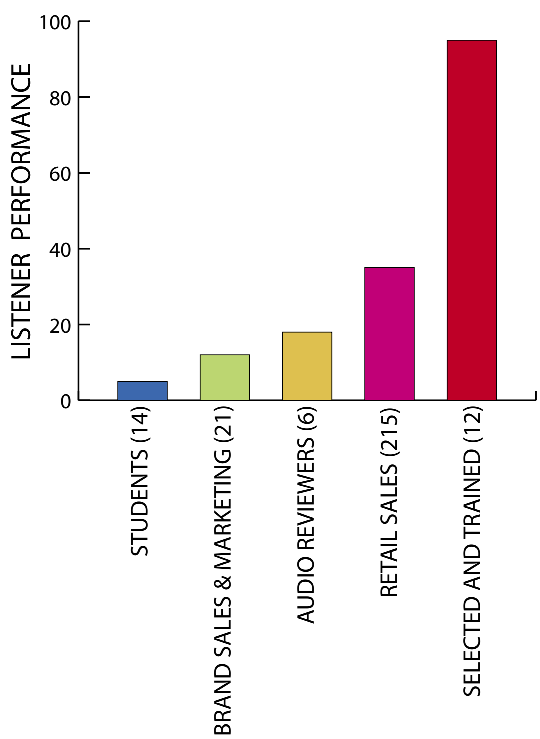


Figure 3: How different categories of people performed in comparisons of three or four high-end loudspeakers. Performance, in this context, measures the ability of listeners to clearly separate the ratings of different sounds, and to repeat those ratings consistently in sequential tests. Compiled from data in Ref. 6.

Listeners in the early tests gained much of their experience "on the job", while performing the tests. Some listeners were musicians, and others had professional audio experience, but most were simply audio enthusiasts. Probably the single most apparent deficiency of novice listeners was the lack of a vocabulary to describe what they heard. Without such descriptions, most listeners found it difficult to be analytical in forming their judgments, and to remember how various test products sounded. It was also clear that, without the prompting of a well-designed questionnaire, not all listeners paid attention to all perceptual dimensions, resulting in judgments that were highly selective.

As the understanding of technically-measurable parameters and their audible importance increased, it was possible to design training sessions that improved the ability of listeners to hear and to identify specific classes of problems in loudspeakers. With the aid of computers, this training has been refined to a self-administered procedure, which keeps track of the student's progress [5]. From this we have also been able to identify program material that is most revealing of the defects that are at issue, thus improving the efficiency and effectiveness of the tests.

Pulling all of this together is an important study [6] in which the opinions of 12 selected and trained listeners are compared to those of 256 listeners from various backgrounds. The relative ratings of the products were essentially the same for small groups of listeners extracted from each population. The consequential difference was in the statistical confidence one could place in the opinions. The selected and trained listeners were much more reliable in their ratings, meaning that trustworthy results could be obtained in much less time. The trained listeners also provided comments that were easily interpreted by design engineers to help them focus on aspects of performance that needed working on, while other listeners tended to use less technically descriptive terms.

BLIND vs. SIGHTED TESTS – SEEING IS BELIEVING

When you know what you are listening to, there is a chance that your opinions might not be completely unbiased. In scientific tests of many kinds, and even in wine tasting, considerable care is taken to ensure the anonymity of the devices or substances being subjectively evaluated. Many people in audio follow the same principle, but others persist in the belief that, in sighted tests, they can ignore such factors as price, size, brand, etc. and arrive at the unbiased truth. In some of the "great debate" issues, like amplifiers, wires, and the like, there are assertions that disguising the product identity *prevents* listeners from hearing small differences. "Proof" of this is the observation that perceived characteristics that seemed to be obvious when the product identities were known, are either less obvious or non-existent when the products are hidden from view. The truth is not always what we wish it to be.

In the category of loudspeakers and rooms, however, there is no doubt that differences exist and are clearly audible. To satisfy ourselves that the additional rigor was necessary, we tested the ability of some of our trusted listeners to maintain objectivity in the face of visible information about the products.

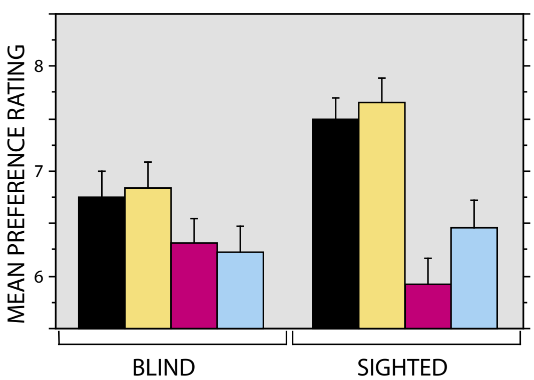


Figure 4. A comparison of blind and sighted evaluations of the same four loudspeakers by the same group of listeners. From Ref. 7.

The results are very clear. Figure 4 shows that, in subjective ratings of four loudspeakers, the differences in ratings caused by knowledge of the products is as large or larger than those attributable to the differences in sound alone. The two left-hand bars are scores for loudspeakers that were large, expensive and impressive looking, the third bar is the score for a well-designed, small, inexpensive, plastic sub/sat system. The right-hand bar represents a moderately expensive product from a competitor that had been highly rated by respected reviewers.

When listeners entered the room for the sighted tests, their positive verbal reactions to the big speakers and the jeers for the tiny sub/sat system foreshadowed dramatic ratings shifts – in opposite directions. The handsome competitor's system got a higher rating; so much for employee loyalty.

Other variables were also tested, and the results indicated that, in the sighted tests, listeners substantially ignored large differences in sound quality attributable to loudspeaker position in the listening room and to program material. In other words, knowledge of the product identity was at least as important a factor in the tests as the principal acoustical factors. Incidentally, many of these listeners were very experienced and, some of them thought, able to ignore the visually-stimulated biases [7].

At this point, it is correct to say that, with adequate experimental controls, we are no longer conducting "listening tests", we are performing "subjective measurements".

HOW BIG ARE THE DIFFERENCES . . . REALLY?

Subjective scales of judgment are notoriously elastic. A comparison of product A with product B might, depending on circumstances, yield opinions suggesting a close race or a run-away victory. It all has to do with subjective scaling, and the context within which the test is conducted. In Figure 5 the left-hand column shows results of a large test including loudspeakers that are truly good and truly bad. The result is a portrayal of relative performance over a wide quality range. Keeping some good and bad "anchor" products in subsequent listening tests will tend to stabilize the scale on which other products are judged.

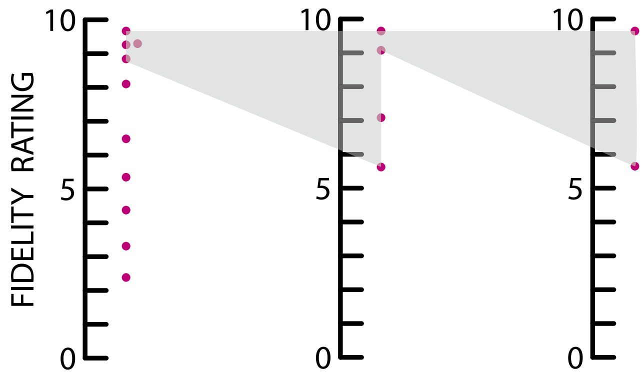


FIGURE 5: What happens when listeners evaluate products that are closely ranked at the top of a range of products. Each dot represents the rating of a different loudspeaker. Without any "anchor products" to remind them of how things really sound at lower ratings, the response scale expands to fill a "comfortable" subjective range of "good" to "less good". The inverse of this also happens.

But, if we focus only on a group of the best products, we get the results shown in the middle column. With reminders of bad sound removed from the tests, listeners spontaneously expand the scaling of their responses downward. A selection of low-rated products from the first column, would have resulted in middle-column ratings that drift upwards. Such is the nature of relative judgment.

The worst of all tests, in this sense, is the A vs. B test, where there is a high probability that differences will be exaggerated. Real-world examples of this occur when, for example, one does comparisons of devices that are fundamentally similar, like most electronic devices and a few superb loudspeakers. The listener impressions may be that there is an undisputed winner, a 9 versus a 6, but the reality is that, if there were any other variables in the test, the ratings may differ only by fractions of a point, and may even be in a different order. Both products could be significantly flawed but, since the flaw is identical in both, it is not noticed. However, as a means of determining if a *difference* exists, there is no better test.

Because of this, another response-scaling method is used when we are interested only in the relative performance of devices. It is a preference scale. The function of this kind of response scaling is to establish how listeners respond to differences between sounds. Since the scale is not "anchored" by reference to sounds known to be very good and very bad, there is nothing to indicate where they stand in any absolute sense. Listeners respond on a LIKE/DISLIKE scale, and are encouraged to separate scores by fixed amounts, which increase as the strength of the preference increases.

MEASUREMENTS: TECHNICAL ⇔ SUBJECTIVE

We know from always bitter and sometimes expensive experience, that a loudspeaker can sound very different in different rooms, and in different positions within the same room. There are two main causes for this variable performance: first, the enormous influence of room resonances at low frequencies and, second, the differing spectral output from the loudspeakers at different angles, and the interactions with reflecting surfaces in the room. This means that, if we hope to be able to anticipate how a loudspeaker will sound, it must be measured in a way that is related to how we listen to it in a room [8]. The following diagrams illustrate the basis of a measurement system that has been very helpful in understanding how

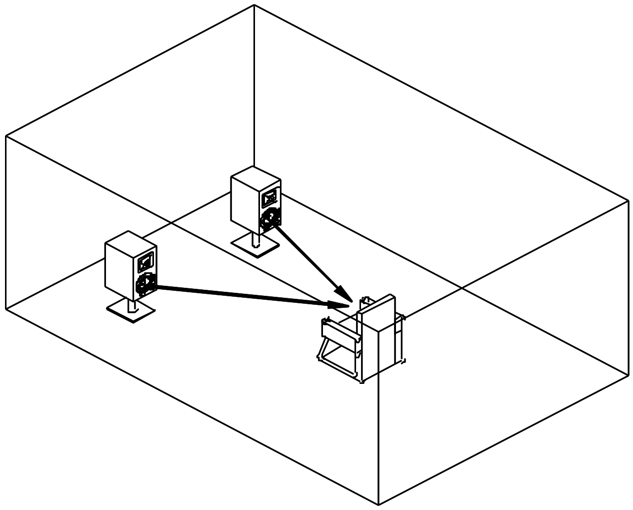


Figure 6. The direct sound

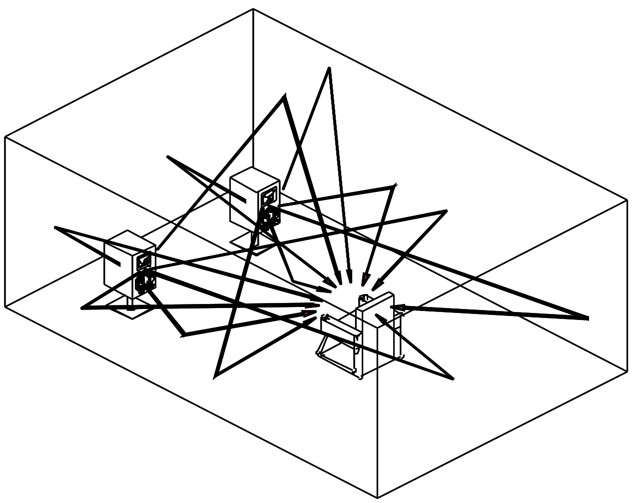


Figure 7. The early-reflected sound.

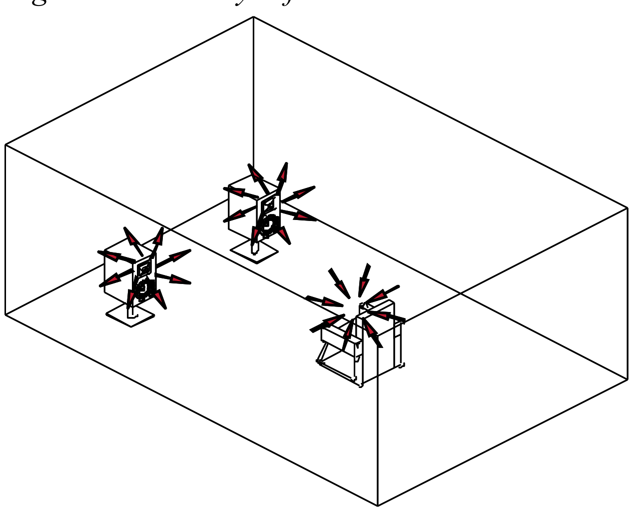


Figure 8. Sounds radiated from the loudspeakers in all other directions will eventually reach the listener, but after several reflections. This is called reverberation.

to design better loudspeakers – loudspeakers that increase the probability of good sound in a variety of different rooms. These loudspeakers are ‘room friendly’.

Figure 6 show the first sound to arrive at a listener’s ears – the direct sound. It is unaffected by the room but, as we shall see, it is a tiny portion of the sound that we hear in a room. Depending on how the loudspeaker is aimed, this may or may not be described by the on-axis response of the loudspeaker.

Figure 7 shows a few of the many paths that reflected sound can travel on the way from a loudspeaker to a listener. These sounds have been reflected only once in transit and, in most domestic-size rooms, will arrive within approximately the first twenty to thirty milliseconds (ms) after the direct sound. They are called early reflections. Those strong reflections arriving later, are called “late reflections”. If these are delayed by 30 ms or more, as can happen in large rooms, they may be heard as echoes.

In small rooms, such as we have in our homes, echoes are almost never a problem, but early reflections are major determinants of what we hear from loudspeakers. Obviously, the strength of individual early reflections is determined by the directivity of the loudspeakers and the reflectivity of the walls. So, when measuring the loudspeaker we must do so at the appropriate angles away from the forward axis. Doing this means that we need information about typical rooms and the arrangements of loudspeakers and listeners within them.

Wide dispersion or multidirectional loudspeakers generate lots of early reflections, meaning that, for such speakers the acoustics of the room are major determinants in how they sound.

Figure 8 shows the rest of the sound radiated from the speakers. Predicting what precisely might happen to this sound in any room is a hopeless task, so we lump it all together and consider the total sound power radiated by the loudspeaker to be a reasonable estimate of it. Sound power is a measure of all of the sound radiated from a loudspeaker in all directions.

The message from this is that, all sound radiated by a loudspeaker, in whatever direction, eventually reaches the listener, and all of it will influence what we perceive in terms of sound quality and spatial and directional effects.

So, why not just measure the loudspeaker in a room? Because, compared to two ears and a brain, a microphone is a "dumb" device. It accepts sounds from any angle, at any time, and treats them equally. In contrast, a human distinguishes between direct sounds and later arrivals (the precedence effect and forward masking), and between sounds from one direction and those from another (binaural discrimination). Steady-state "room curves" are not at all analytical in these respects, offering little insight into what may be responsible for any interesting features – the loudspeaker itself or its interaction with the room boundaries.

To make measurements that allow us to anticipate what may happen in real rooms we need statistical data about what happens in real rooms, all of which are different in some respects. Such a survey was done, and the data analyzed to show the angular statistics for the various significant reflections, including delays and propagation attenuations. A measurement system was then designed that yielded a set of frequency response curves indicating the direct, early reflected and reverberant sounds expected from the loudspeaker under test, when used in an 'average' listening room [9].

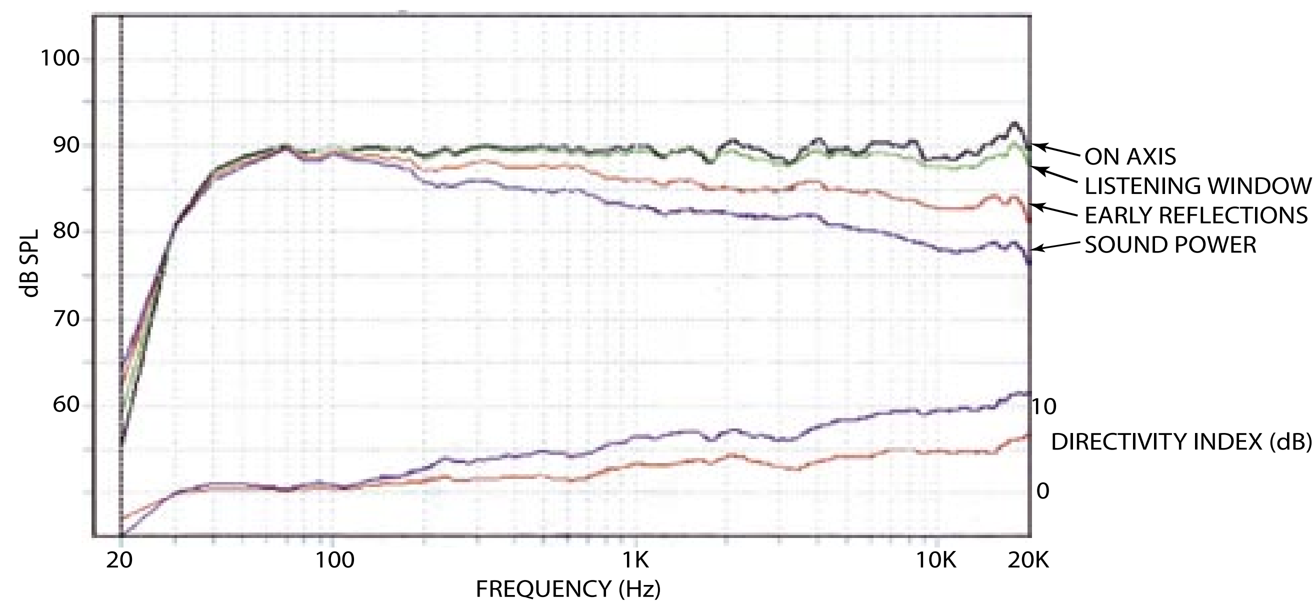


Figure 9. Measurements that give us a very good prediction of how this loudspeaker will sound in a room. In a large anechoic chamber, at a distance of 2 m, a total of 70 frequency-response measurements were made at 10° intervals on horizontal and vertical orbits. The data were processed to yield these curves. The on-axis curve describes the direct sound, the first sound to arrive at the ears of somebody seated in the 'sweet spot'. The listening window describes the average direct sound for listeners seated or standing within a ±10° vertical, ±30° horizontal region directly in front of the loudspeaker – the entire audience in a home theater, for example. The next curve describes the sound of the average strong early reflections from the room boundaries, and the sound power is a measure of the total sound output of the loudspeaker without regard to direction. The bottom two curves describe the directivity (DI = Directivity Index) of the loudspeaker: how uniformly it radiates its sound into the room at all frequencies. The top DI curve is an adaptation of the 'classic' one, based on a comparison of listening window measurement and sound power. The lower DI curve is based on the 'early reflections' curve. In a truly good sounding loudspeaker, the on-axis and listening window curves will be smooth and flat, and the other curves will be smoothly and gradually changing. This is an example of an exceptionally good loudspeaker, a three-way studio monitor listing at $1400 each.

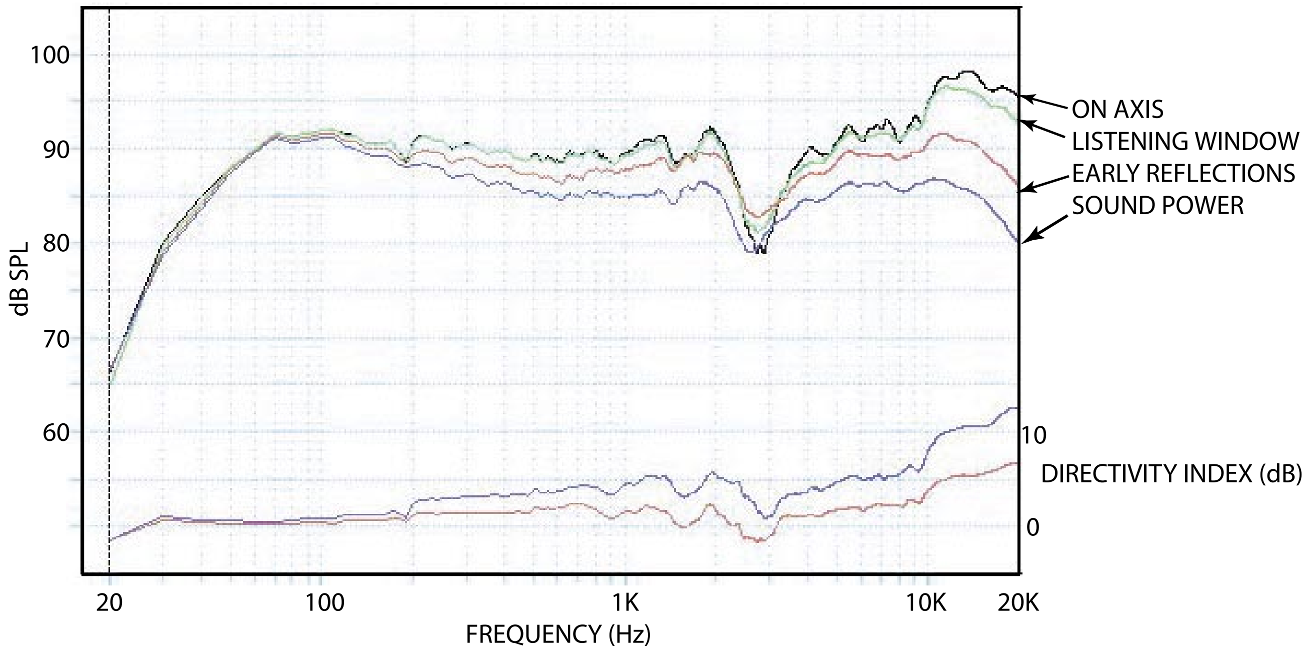
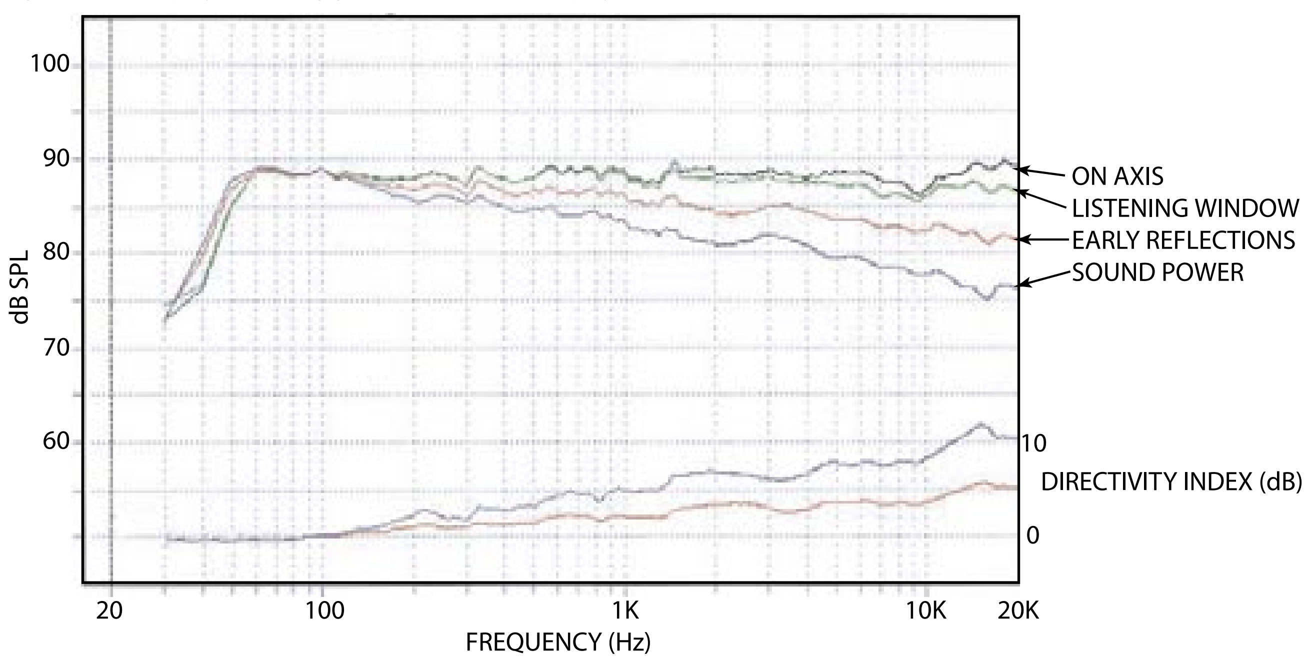


Figure 10 (above). A three-way loudspeaker (7" woofer, 7" midrange, 1" tweeter), at $750/pair. It has several audible problems. Using the mid-frequency band as a reference, there is a slight bass exaggeration around 70-100 Hz. At 200 Hz a discontinuity suggests a mechanical resonance in a woofer or enclosure panel. The dip around 3 kHz is caused by a problematic midrange-tweeter crossover transition. It is serious, since it appears in all of the curves, even the sound power. A resonance just below the dip, around 2 kHz, aggravates the coloration. The steeply rising high-frequency output adds 'sparkle', 'sizzle' and 'air' that are not in the original program. Looking at the directivity indices, one can see the gently rising directivity of the 7-inch midrange, which is too large for a smooth transition to the 1-inch tweeter, resulting in a discontinuity at the 3 kHz crossover frequency.

Figure 11 (below). A loudspeaker with a similar driver configuration, at $700/pair. Superior engineering, including better crossover design and a waveguide added to the tweeter to better match directivities at crossover, yield a loudspeaker that is a consistent high scorer in listening evaluations. Its all-round good behavior makes it very "room friendly", with a high probablility of sounding good in a wide variety of rooms.



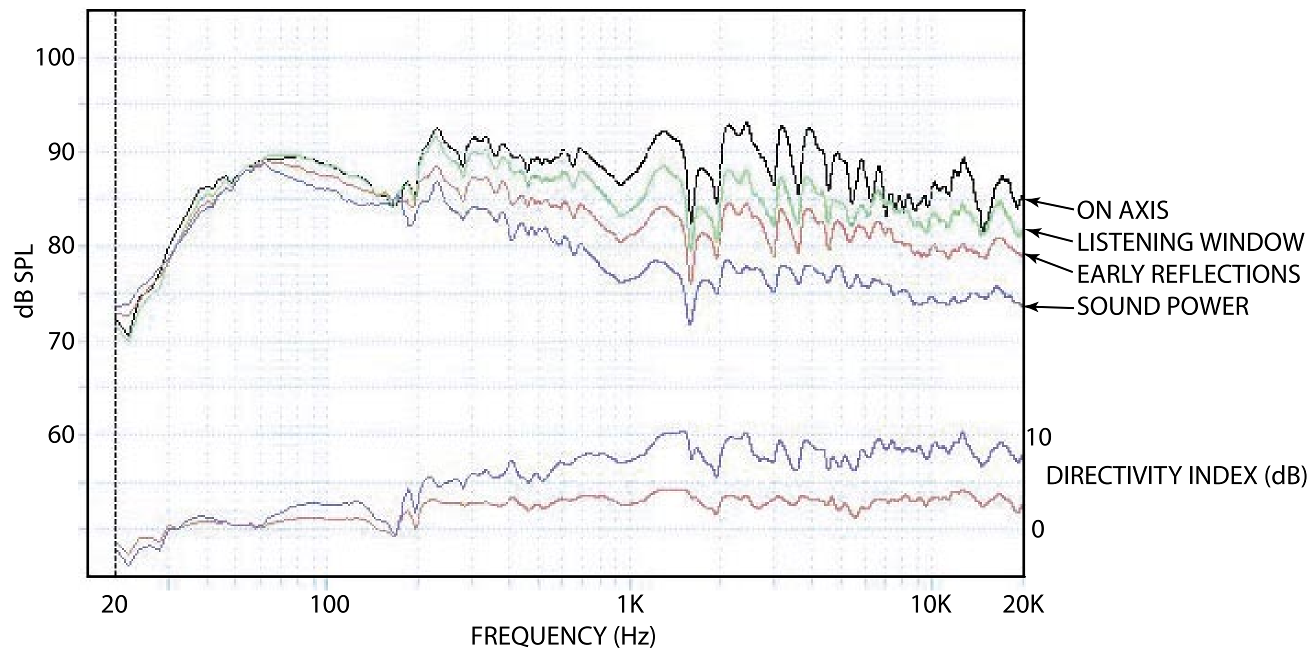


Figure 12. A hybrid loudspeaker consisting of a large-panel mid/high frequency bidirectional (dipole) radiator mated with a conventional woofer. Price: $11,000/pr. The transition frequency is clearly visible at 200 Hz, where the directional characteristics change. From the omnidirectional woofer (directivity index near zero) the radiation pattern jumps to one that is quite directional, hovering around 7-8 dB from 600 Hz up. The fact that there is a consistent 2-3 dB, or more, difference between the on-axis response and the listening window, means that listeners at different horizontal angles from the loudspeaker, on a sofa for example, will not hear the same spectral balance. The curves are also not smooth, indicating the possibility of resonances. This possibility is confirmed as a reality when the same pattern is seen in all of the curves, up to and including the sound power. The clear differences in overall level and bass balance among the curves poses some perceptual challenges in that the direct, early-reflected and reverberant sounds each exhibit different spectral balances. The high directivity, means that the direct sound will be relatively stronger than reflected sounds in rooms. This characteristic, and the collection of resonances, are the principal distinguishing features in the sound of this loudspeaker.

We could go on, examining many more examples of good and less-good designs, but the pattern is emerging: measurements of the right kind are good predictors of how loudspeakers sound. But, are there other predictors? After 25 years of doing evaluations of this kind it can be said that there are no safe assumptions when it comes to choosing loudspeakers. Price is certainly an unreliable indicator of sonic excellence, as is transducer type. For each example of a good cone/dome, electrostatic, panel, or whatever, kind of loudspeaker, it is possible to find ones that are not well designed.

The competence of the engineering, and the attitude of brand marketing towards achieving excellence and neutrality in sound quality, are the real differentiating factors in how loudspeakers sound. Sadly, some manufacturers lack the skills and engineering resources to do a really good job even if they wanted to. Others deliberately design loudspeakers with sounds intended to "stand out" in a showroom. The distinctive "booms" and "tizzes" eventually become tiresome, and another disillusioned customer has been created.

Magazine reviews, sadly, are not always what they should be. The loudspeaker shown in Figure 12, for example, has received rave reviews and awards. Ref. 6 shows results of listening tests with 268 listeners, who persuasively indicated that they heard substantial problems, relative to some well-designed loudspeakers. They were listening in a double-blind situation. The reviewers, almost certainly, were not.

Price is certainly an unreliable indicator of sonic excellence.

Listening tests require facilities, just as do technical measurements. Since one is evaluating the combination of loudspeaker and room, we must do something to ensure that the room is a neutral factor [10]. One way to do this is to bring each loudspeaker being compared to precisely the same location in the same room. Our solution has been to build an elaborate 'shuffler' mechanism to physically move the loudspeakers [11]. This is combined with careful loudness balancing, and the appropriate randomization of musical programs and loudspeaker identifiers.

Figure 13. A computer-controlled, pneumatically-operated loudspeaker mover eliminates loudspeaker position as a variable in listening tests. The photo on the left shows one stereo pair of loudspeakers in the active location, while another pair is parked quietly against the back wall. The listener controls the test, switching among the comparison loudspeakers while forming opinions (there can be up to four single loudspeakers, four stereo pairs, or three L,C,R combinations). Of course the tests are double-blind.

INTERPRETING DETAILS IN THE CURVES

The ideal frequency response curve might be a straight line, or at least a gently curving one, but at what point does a deviation, a peak or a dip, become an audible problem? Designing good-sounding loudspeakers that are cost effective requires that knowledge.

The overall trends in frequency response are not hard to see. The lower and upper frequency limit can be easily found and, in between it is possible to visualize general spectral shapes and trends. The tricky part comes when one attempts to identify whether a specific peak or dip has been caused by a resonance (serious) or by acoustical interference (not so serious). One of the reasons why acoustical interference is harder to hear is that, in rooms, there are many reflected sounds, each one with a distinctive acoustic interference pattern (see Figures 7 and 8). In the perceptual summation of all of them, the audible effects are largely averaged out.

The sensitivity of acoustical interference to microphone position means that, in measurements, spatial averaging (combining measurements made at several microphone locations) will attenuate these effects, while leaving those due to resonances substantially unchanged. In this sense, spatial averaging adds information, while spectral averaging, e.g. 1/3-octave smoothing, removes it.

In the curves shown in Figures 9–12, only the on-axis curve is not spatial averaged. Peaks that appear in it, and that persist through the other curves, especially up to the sound power curve, which is a full 360° spherical average, are most likely due to resonances. It is a simple, but very effective, analytical method.

Once the resonances have been identified, their audibility can be assessed by referring

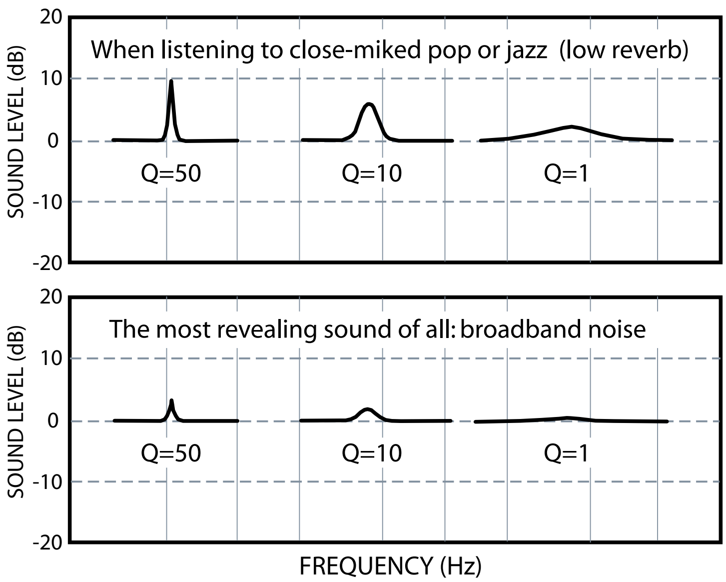


Figure 14. High-, medium- and low-Q resonances shown at the detection thresholds, below which the resonances cease to be audible. It depends on what you are listening to. The locations of the illustrative peaks on the frequency scale are arbitrary. The audibility is similar at all frequencies except, perhaps, at very low frequencies. Adapted from Ref. 12.

to Figure 14. Why is it that resonances are so important? Because they are the fundamental building blocks of almost all of the sounds we are interested in hearing. High-Q resonances define the pitches. Medium- and low-Q resonances define the timbres, allowing us to distinguish between different voices and instruments. It is subtle differences in the resonant structure of sounds that are responsible for the nuances and shading of tone in musical sounds. Our ears are very highly attuned to the detection and evaluation of resonances, and it is therefore no surprise that listeners zero in on them as unwanted "editorializing" when they appear in loudspeakers.

As can be seen, there really is such a thing as being "good enough for rock and roll". Complex instrumentation and reverberation result in lower detection thresholds, and continuous noise yields thresholds that are lowest of all – presenting a challenge for acoustical measurement systems to reveal them accurately, and for engineers to design transducers and enclosures that meet the criterion.

I think most people might be surprised to see that a high-Q resonance can exhibit a peak of 10 dB before it becomes clearly audible with *any* kind of music or sound. Similarly, why is it that the low-Q resonances, that hardly ring at all, are audible at such low levels? At least part of the answer lies in the ability of musical signals to excite the resonances. A high-Q resonance is very frequency specific, and it takes time to build up, just as it decays slowly. This means that a musical signal must hit the resonant frequency quite accurately, and stay there long enough to transfer significant energy to it. The changing nature of music, and vibrato, both conspire against this happening. In contrast, lower-Q resonances have wide footprints in the frequency domain, and so are more often and more readily excited.

Reverberation is another factor that lowers thresholds. The common example of this is in the experience of music being performed outdoors, where it is dry and lackluster, compared to the richness of performances in a hall, where the repetitions offered by reflections and reverberation allow timbral subtleties to be revealed [12].

A practical problem is that many measurements these days are made with FFT, TDS, and other systems that time-window the data so that anechoic measurements can be made in normal rooms. A result of the time windowing is that the frequency response data have poor frequency resolution (≥100 Hz is common), and cannot reveal high-Q phenomena at low and middle frequencies. Resolution limitations of this kind are common among loudspeaker designers and reviewers who measure in ordinary rooms. Anechoic chambers are expensive, and measuring outdoors, in nature's own anechoic space, has practical difficulties. It means, simply, that many commonly-used and published measure-

… there really is such a thing as being "good enough for rock and roll".

ments simply cannot reveal visual evidence of certain kinds of audible problems falling within a critical portion of the frequency range – that of the human voice and below.

Another common measurement is one in which the audible frequency range is divided into equal fixed-percentage bandwidths, such as 1/3 octaves, or in which a high-resolution measurement is heavily smoothed, or spectrally averaged, on a continuous basis. These spectral-averaging devices have extremely limited utility in the design and evaluation process.

Resonances reveal themselves in both the frequency-domain (amplitude and phase vs. frequency) and the time domain (impulse response / transient response). As defined by the Fourier Transform, if there is misbehavior in one domain, there will be misbehavior in the other, so we have two ways to look for problems.

Figure 15. A 4π (full space) anechoic chamber in which high-resolution measurements over the entire audible bandwidth are possible. This chamber is naturally anechoic down to 60 Hz, and calibrated down to 20 Hz.The platform on which the loudspeaker is located rotates to permit measurements over a full 360°orbit. The loudspeaker is then tilted to complete orbits over more of the total sphere surrounding it.

It is a matter of fact that high-Q resonances exhibit prolonged ringing in the time domain, and that low-Q resonances exhibit little ringing. The irony of this finding is that, as represented in conventional steady-state frequency response measurements, Figure 14, the low-Q resonances were detectable at much lower amplitudes than resonances of higher Q. This means that prolonged ringing, by itself, is not a reliable indicator of an audible problem Figure 16 confirms it, at least for middle and high frequencies.

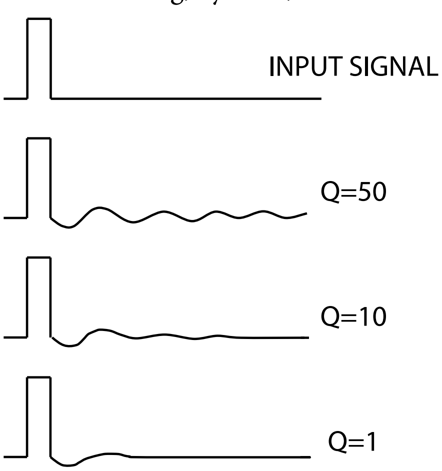


Figure 16. Pulse responses of an otherwise perfect system (top curve) to which resonances have been added at the thresholds of detection for pop music - see Figure 14.

At very low frequencies, the long wavelengths and periods allow ringing to be heard as an extension to bass sounds – boom. At these frequencies, both frequency domain and time domain data appear to matter. Fortunately, in practice, if a resonance does not make itself apparent in an accurate, high-resolution frequency-response measurement, then it is probably not audible. If a resonance is visible as a bump in a frequency response measurement, its audibility must be assessed by reference to the detailed analyses in references 12 and 13.

An interesting fact now emerges: that the conventional method of specifying frequency response, ± *x* dB, is useless unless the tolerance is very, very small. High-Q phenomena could be ± 5 dB, while moderate-Q resonances could be ± 3 dB and low-Q and other broadband deviations could be ± 0.5 dB, and *all* of them would be *equally* audible! Clearly, frequency response curves must be *interpreted.* Frequency responses must also be measured in a manner that reveals resonances as distinct from interference peaks and dips – i.e. spatial averaging. The sad conclusion, therefore, is that standard verbal descriptions of frequency response are not useful, and most published graphs are not of the right kind to allow for unambiguous interpretation. It is easy to understand how the popular belief that 'you can't measure what you can hear' came to be. Bad measurements and useless specifications are responsible.

… the conventional method of specifying frequency response, ± *x* dB, is useless …

Having identified the presence of a problematic resonance, it is the task of the design engineer to diagnose the cause, and to prescribe a remedy. All the while, it is essential to keep a close eye on costs. In this task, other specialized tools are needed, since the origins can be both acoustical and mechanical, and they can be associated with either the drivers or the enclosure. It is a curious phenomenon that the perception of "boxiness" in sound, may have nothing to do with the box itself. It is not uncommon for the offending resonance to have another origin.

In the analysis of resonances, a scanning laser interferometer/vibrometer is a powerful ally, in that it can show the complete vibratory behavior of a surface, such as a loudspeaker diaphragm or an entire surface of a loudspeaker enclosure. Vibration is important only if sound is radiated. Surfaces do not always move uniformly, as a piston, so measurement at a single point can be misleading. In practice, portions of the surface can move in opposite directions, simultaneously. Consequently, some vibratory modes radiate sound very effectively, others less so, and some not at all. It is important not to go chasing after modes that simply cannot be audible. The combination of finite element analysis, in the design of diaphragms and enclosures, and modal analysis, after the prototype is built, allow engineers to be intelligent about the way they use shapes, materials, braces, etc. in reducing the audibility of resonances. The traditional method has been to build enclosures from massive, dense and stiff materials, playing it safe. If cost, size and weight are not considerations this method works well. However, with attention to modal prediction and analysis, enclosures with excellent acoustical performance can be built at much lower cost.

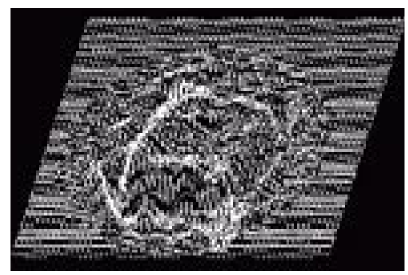


Figure 17. A scanning laser vibrometer image of a loudspeaker diaphragm "breaking up". Different parts of the cone are moving by different amounts, in different directions, all at the same time. This is not good.

PREDICTING SUBJECTIVE OPINION

A loudspeaker engineer's dream is a system by which technical data can be used to reliably anticipate listener preferences. We take it for granted that technical measurements are repeatable. If done carefully, they may also be accurate, and then they are repeatable when done by different people in different locations. If we do enough measurements, we may gain insights into why listeners react in certain ways to certain loudspeakers.

In contrast, our expectations are that subjective evaluations will be variable. However, if done with adequate controls, they turn out to be remarkably repeatable, even when done by different listeners – albeit not just *any* listener. The room is a variable too, so it is well to maintain it as a constant factor when doing comparisons among numbers of loudspeakers.

In the discussions so far, one can see that specific technical data correlate well with specific kinds of perceptions. However, what is missing is a method by which *all* of the relevant technical data can be combined in a computational model yielding a number that correlates with overall subjective preference as determined in carefully-controlled listening tests.

The concept is not new, but examples are rare. Perhaps the only ones of any practical note are the loudspeaker reviews in *Consumer Reports,* published by Consumers Union (CU), starting about 1970. Based on a comparison of subjective and objective data, a method of processing sound power measurements was developed that resulted in a percentage accuracy rating [14,15]. Products are rank ordered in quality according to this rating.

Over the years, it became a bit of sport, within the audio community, to be critical of this process, alluding to examples of loudspeakers that were believed to be better or worse than the CU ratings suggested. New knowledge about the acoustics and psychoacoustics of loudspeakers and rooms suggested that the CU method was flawed. Still, a viable alternative computational model has been lacking … until now.

With trustworthy technical data from the Harman measurement facility, and with trustworthy subjective data from his program of listening evaluations, Sean Olive has moved this process forward. Starting with 13 loudspeakers from a CU test, he performed listening tests of the kind described earlier, and compared the subjective preference ratings with the published accuracy ratings. The correlation was low, and negative: – 0.22 [16].

Next, models were developed employing different subsets of the anechoic data (of the kind seen in Figs. 9-12). One of these models predicted our own listening test results with a correlation of 1.0 [17]. However, the big variable in these exercises is the subjective data. All 13 CU loudspeakers were evaluated in one long series of listening tests. The ratings therefore had same kind of self-consistency seen on the left in Fig. 5. However, our routine tests are done in groups of four products, when doing competitive evaluations of new loudspeakers under development. These subjective ratings are victims of the upward and downward drifts and the scaling 'elasticity' shown in the middle data in Fig. 5. The ratings are consistent within each test of four loudspeakers but, when data from many groups of four are pooled, scaling distortions contribute artificial variability. Consequently, when the correlations were expanded to a total of 70 loudspeakers tested over a period of several months, the results were less impressive. However, "less impressive" does not mean that the results were unimpressive. A correlation coefficient of 0.86 between the predicted and real subjective ratings is remarkable. This is no accident.

Among the results are conclusions that measurements with1/3-octave resolution are not adequate, that sound power or in-room measurements alone are not sufficient to predict listener preferences, and that the flatness and smoothness of high-resolution on-axis curves need to be given substantial weighting. Powerful research, and the work continues.

LOUDSPEAKERS AND ROOMS

The enclosures behind loudspeaker diaphragms are very important, but it is the ones in front –rooms – that give us really challenging problems. The listening room is the final audio component, and it is the one over which the loudspeaker manufacturer has little or no control.

Loudspeakers can be designed so that they have a reasonable chance of sounding good in rooms. As shown earlier in Figures 6-8, listeners hear the direct sound first, followed quickly by early reflections from floor, ceiling and sidewalls. Then arrive the multitudes of sounds from many reflecting surfaces after several reflections – the reverberation. Ideally, all of these sounds should reinforce a similar timbral signature in the mind of a listener. This can only happen if the loudspeakers are designed to radiate similar sounds in all of the relevant directions. In technical terms, this reduces to a requirement for constant, or at least smoothly changing, directivity as a function of frequency.

This became evident in the curves of Figures 9-12, showing examples of loudspeakers that scored highly in listening tests, and some that did not. The winners are easy to pick; flat and smooth are beautiful.

But there is more. Within the criteria for good sound, it is possible to design loudspeakers with different directional characteristics, ranging from relatively directional forward-firing horns, through the common cone/dome configurations, to bipoles, dipoles, and various direct/reflecting configurations. Since these designs evolved in the age of two-channel stereo, it is proper to speculate whether there is a relationship with the delivery medium and, further, whether the relationship is the same in the new age of multichannel audio. The vast majority of loudspeakers sold are traditional forward-facing "cone and dome" designs. These, and horn-loaded variations, are standard equipment in recording studios for use in monitoring the creative process for stereo as well as multichannel recordings. In homes, however, other designs have crept in.

LOUDSPEAKERS FOR STEREO

With only two channels, stereo is a spatially-deprived medium. Yes, some recordings, from the perspective of the symmetrical sweet spot, deliver quite wondrous sounds. However, most recordings are not fully gratifying, with whole sections of an orchestra appearing to emerge from a single loudspeaker, phantom images inappropriately floating around, and disappointing sensations of depth and space. The "you are there" sensation is missing. Motivated by a desire for ever "deeper, wider, more spacious and airy" soundstages, some audiophiles gravitated to loudspeaker designs that directed substantial energy towards the walls, adding reflected sounds to the playback "mix". Wide-dispersion forward-firing cone/dome systems, and some bidirectional and multidirectional designs evolved to fill this need. Others experimented with additional loudspeakers and delay devices. It is probably correct to say that the majority of listeners find stereo to be pleasantly embellished if the room reflections are energetic. The sound tends to be open and spacious, with a good sense of depth, but specific images might be rather vague – in other words, rather like real concerts.

However, some listeners prefer a very specific, almost pinpoint, sense of image position. These people are attracted to highly directional horn and large panel loudspeakers. Similar effects can be achieved with other loudspeakers by attenuating early room reflections with large areas of acoustical absorbing material. Among these listeners are many recording engineers. Consequently, recording studios are often acoustically rather dead, and the loudspeakers directional (often horn loaded), or placed very close (so-called near-field listening). However, these same people, at home, frequently revert to a more spacious version of stereo. Let's hope that this is where they do their reality checks.

This said, there is an abundance of everyday experience, and some scientific data [4], indicating that the dominant factor in the 'imaging' we perceive is the microphone and mixing technique used in the making of the recording. The loudspeakers have an effect, but it is a subordinate one, and most of that effect has to do with the relative strengths of early reflected sounds [18].

The problem with customizing the soundstage with a loudspeaker configuration is that it is totally inflexible. Once selected, it gets applied to every recording played through the system, whether it is ideal or not. Purists need to reflect on the fact that *only* forward-firing loudspeakers are used during the recording process, where the recording engineers and artists make their decisions.

LOUDSPEAKERS FOR MULTICHANNEL AUDIO

In a rational world, the uncertainties of stereo should be eliminated by multichannel systems. After all, with more channels, and independent control of each, what listeners hear should be less dependent on the room. Artists can predict the imaging and spaciousness to a much greater degree than stereo could permit. It remains only for consumers to set up similar loudspeakers in a similar arrangement, sit back and enjoy. Alas, it is not so simple.

As discussed earlier, there is the disparity in production methodologies and philosophies between movies and music, which leads to some differences in locations and types of loudspeakers, if listeners are to hear what they should. Fortunately, the locations and configurations of the front left, center and right loudspeakers are not controversial. Most home theater installations employ forward-firing loudspeakers in these locations, but the surround loudspeakers are a different matter.

In a multichannel system, the impressions of space and envelopment are mainly provided by the surround loudspeakers positioned to the sides of the listeners. Here is where tough decisions need to be made. What really is at issue here is the matter of whether the listener should be in a predominantly direct or predominantly reflected sound field from the surround loudspeakers.

There cannot be a single correct answer for the loudspeaker configuration until there is agreement at the production end of the process.

From the perceptual point of view, impressions of ambiguous localization and spaciousness exist when sounds arriving at the two ears are uncorrelated. This can be achieved by using multidirectional loudspeakers that generate many reflections. Some refer to this as a "diffuse" sound field. It isn't truly diffuse in the small, acoustically well damped, rooms we use for home theaters. Still, it is a multidirectional reflected sound field, and it can encourage the perception of ambiguous localization. Uncertainty arises because, for this sound field to be created, the room surfaces need to be reflective in the right locations relative to the loudspeakers and listeners. Multidirectional loudspeakers rely on reflective walls to deliver their sounds to listeners.

JBL K2 S9800

On the other hand, if there is decorrelation in the recording, or it is created electronically in a surround processor during playback, the same effect can be achieved using conventional forward-firing loudspeakers. The uncertainty is gone.

Movies are made assuming that audiences will experience a multiplicity of surround speakers down the sides and across the back of a cinema. At home, this argues for multidirectional surround speakers, additional electronic decorrelation, or multiple forward-firing loudspeakers.

Music, in contrast, is mixed in the expectation that listeners will have five identical loudspeakers, three across the front and one on each side, situated slightly behind. Movies use the surround channels for directionally ambiguous ambient sounds, music, and occasional momentary sound effects. Music recordings might put a whole musician in a surround channel, or a group of backing musicians in both surround channels.

There cannot be a single correct answer for the loudspeaker configuration until there is agreement at the production end of the process. So, in setting up a multichannel audio system, some serious thought must be given to how it is to be used.

GOOD BASS IN ROOMS – SCIENCE TO THE RESCUE

At low frequencies the room interactions take on a special flavor because of the way acoustical standing waves (resonances) dominate what we hear. The quantity and quality of bass sound is as much or more determined by the room and how it is set up, as it is the speakers themselves [10,19,20,21]. In the investigation of many rooms over the years, I would estimate that something like 80% have serious bass coloration - too much, too little, boomy, uneven, etc. Add to this the inevitable seat-to-seat variations in bass, and it is clear that a lot of our customers need help.

This situation is an enormous frustration for speaker manufacturers. Once the product is in a box, on a truck, we have lost control of how the speaker will sound at low frequencies. Gaining some control has been the objective of several research projects and product innovations over the years. Short of hiring an acoustical consultant, and being willing to rearrange the furniture and possibly rebuild the walls, what can be done?

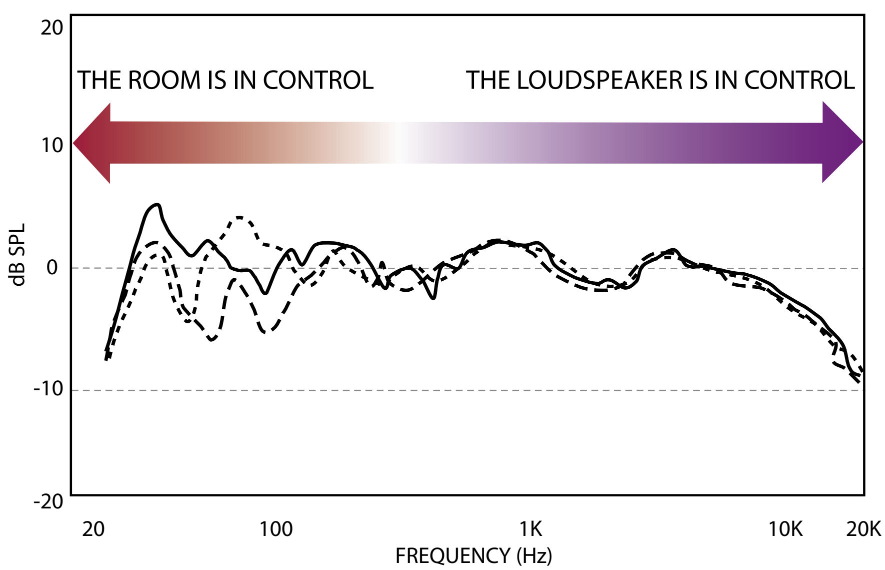


Figure 18. A loudspeaker placed in three locations within a 2 ft. radius and measured at the listening positions in a typical room. The graphs show the profound effects of room resonances at frequencies up to about 300Hz. Above that frequency the curves change relatively little indicating the dominance of the loudspeaker frequency response and directivity in determining the shape of the curve. The dividing frequency is a function of room size, moving lower as the room gets larger. This explains why bass sounds so good in concert halls.

There are two steps in the solution. The first step is to decide how many listeners should have expectations of good bass. Seat-to-seat variations in bass quality can be huge. If there is only a single listener, then proceed immediately to EQUALIZATION. If there are multiple listeners deserving of good bass, then visit SOUND FIELD MANAGEMENT for explanations of how to employ multiple subwoofers to achieve more uniform bass at several seats in a room.

SOUND FIELD MANAGEMENT

With a thorough understanding of acoustical standing waves in rooms, it is possible to manipulate them by using multiple subwoofers. In any given situation, it may be possible to alleviate a problematic resonance by the clever positioning of one or more subwoofers [15]. However, such a solution is not likely to work in another room – all rooms are different. The approach we adopted was to look for a more general solution to the problem, one that could be trusted to work in many rooms. It turns out that there are, in fact, two solutions: one for simple rectangular rooms, and one for rooms with more complicated shapes.

❑ Rectangular rooms. These rooms must have mostly flat walls, floor and ceiling, and no large openings to other spaces. Summarizing the results of an elaborate study, Figure 19 shows some of the recommended subwoofer arrangements. There is no improvement in using more than four subwoofers. The identical loudspeakers are driven with identical signals. Some of the arrangements permit the use of smaller subwoofers. See Reference 22 for more options, and more explanation.

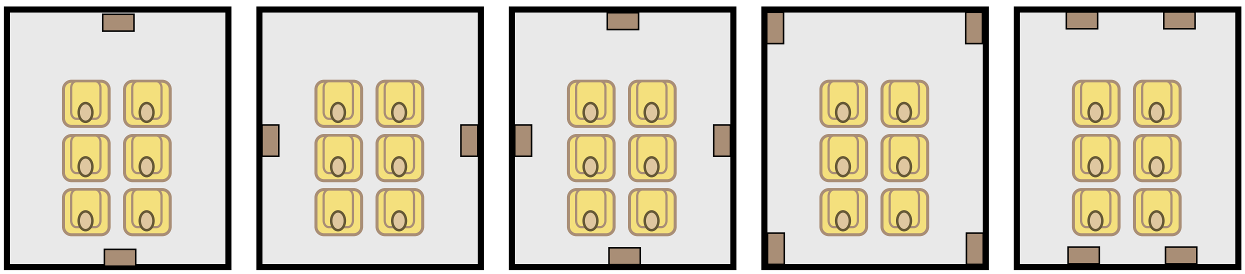


Figure 19. In rectangular rooms, some of the arrangements for two or four subwoofers that yield improved bass uniformity over the listening area. From Ref. 18.

❑ Non-rectangular rooms. Things get much more complicated in these rooms; there are no safe generalizations and every room must be treated individually. Inevitably, acoustical measurements are necessary, and some form of optimization algorithm is needed to work through the variables. In addition to the room, loudspeaker position, and listener position, the variables can now be expanded to include any or all of: delay, amplitude and parametric equalization in each of the subwoofer signal feeds. The questions are: how many subwoofers do we need, where should they go and what signals do we feed them? In this solution, the user is provided with some form of measurement and computational capability. This can be an elaborate stand-alone system for custom installers, or a simplified system for the mass market.

The setup of such a system begins by identifying possible locations for subwoofers, and specifying listener locations. Complex acoustical transfer function measurements are then made from each potential loudspeaker location to each listener location. Working with the stored data, an optimization algorithm then calculates the best solutions for the number of subwoofers the operator specifies. The output is a set of predicted frequency responses at each listening location. So, one can evaluate how well the system can work with two, three, four, or more subwoofers, looking to minimize variations in frequency response among the listening locations. The customer can then choose to go with the most cost-effective or practical arrangement, or with the very best one, if cost is not an issue. It is possible even to have different optimized solutions for different numbers of listeners. An elaborate custom system might have settings for "me", "me and my best friend" and "a full house". See Reference 23 for more information.

EQUALIZATION

With one subwoofer, and a lot of luck, it is possible to deliver good bass to a single listener. But what if you are not lucky? With one subwoofer, a device capable of measuring high-resolution (1/10-octave or better) frequency responses, and a parametric equalizer, it is possible to deliver good bass to a single listener – no luck required. Some other listeners in the room may also be content, but they now are the ones with the luck.

Sound field management, as described above, yields the desirable situation of several listeners in a room hearing similar bass sound. Now we need to apply equalization to make that sound good. Interestingly, in the process of reducing the seat-to-seat variability, we have also reduced the need for aggressive equalization – a win all around.

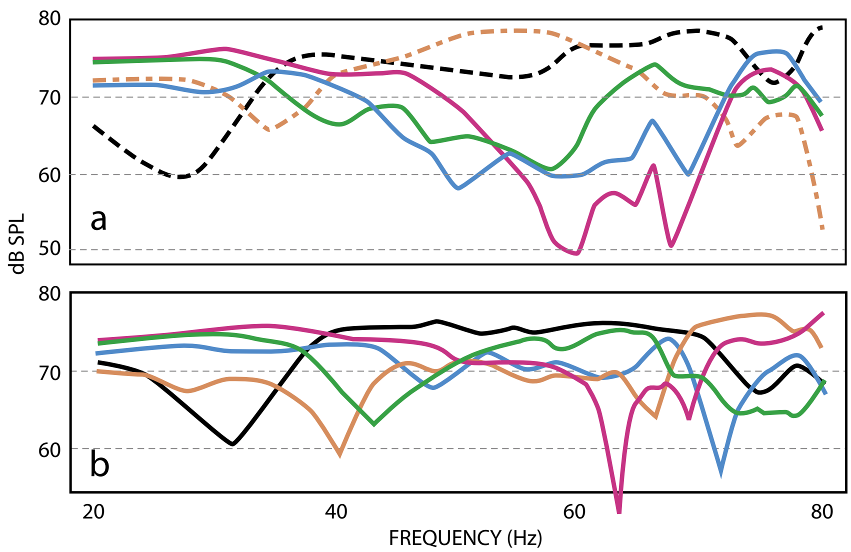


Figure 20(a). Using one subwoofer, frequency response measurements made at five seating locations in my listening room. The max/min variation among these curves is 28 dB! The two dashed curves are flatter than the others but, unluckily, my seat is one of the others. If one of the lower curves is equalized to be flatter, both of the dashed curves will acquire huge bass excesses.

Figure 20(b.) With four sound-field-managed subs, equalized, all of these curves track each other closely. The dips are very narrow, and appear to be inaudible. Bass is now similarly and superbly deep and 'tight' in all of the seats.

There are some things equalization can and cannot do:

1. It can help with some, but not all, loudspeaker problems. With comprehensive laboratory measurements (like Figs. 9–12) to work from, equalization can be used to make a good speaker sound better. Once a loudspeaker is in a normal room, we lose the ability to measure it in ways that allow us to be totally analytical. Without knowing what is wrong, we don't know whether equalization is the right solution for the problem. For example, equalization can change frequency response, but it cannot change directivity, yet together they determine the basic shape of a room curve. Poor directivity control, as a function of frequency, can only be cured by using a better loudspeaker. In general, if the speaker has been competently designed, it should probably be left alone at frequencies above about 300 to 500 Hz, whatever the room-curves look like. Since all manufacturers claim to be competent, consumers have a problem. Automated equalization systems that are not dedicated to specific loudspeakers have a problem. They have no special "inside information" about the loudspeakers, so those that make adjustments over the whole audio frequency range are, as they say in 'Vegas, rolling the dice. *Caveat Emptor.*

2. It can help with some, but not all, room interaction problems. At frequencies below about 300 to 500 Hz, the shape and size of the room, and the position of the speaker and listener within it dictate the system performance. Two factors are active here:

- ❑ room resonances, some with quite high "Q", can cause note-to-note fluctuations in loudness, and boominess in percussive drum and bass sounds. Since room resonances behave as minimum-phase systems, once the peak in the amplitude frequency response (See Figure 14) has been attenuated using a matching parametric filter, the time-domain misbehavior (the ringing seen in Figure 16) will also be tamed. This works.
- ❑ Acoustical interference caused by the interaction of many reflected sounds with-

in the room. These "comb filters" are non-minimum-phase phenomena, and they cannot be corrected with equalization.

Spatial averaging – measurements made at several locations and averaged – can help separate these phenomena by attenuating the visibility of acoustical interference, while leaving resonances relatively unchanged.

The misuse of equalization over the years has given it a mixed reputation.

The misuse of equalization over the years has given it a mixed reputation. Steady-state room curves are crude measurements, in that the microphone simply adds up all of the incoming sounds, from all directions, and without regard to relative timing. Two ears and a brain are much more sensitive and analytical – directionally, temporally and spectrally.

Allusions to perceptual *critical bands* frequently suggest that this is the resolution of the ear; that anything existing within a critical band is perceptually combined. This argument has been used to justify measurements with1/3-octave resolution, even though this is not an accurate representation of actual critical bandwidths.

In psychoacoustic research, it has been found that loudness summation, masking and multiple-tone identification all exhibit behavior that relates to critical bands. For example, two tones separated by less than a critical band may not be perceived as two recognizably separate tones. However, multiple tones occurring within a critical band don't disappear, nor do they lose all identity. In fact, they interact with each other non-linearly, generating beats and roughness that are components in the sounds of instruments and music. If anything happens to change the amplitude or frequency of these multiple events within a critical band, then the perception is changed. If one tone is significantly changed in amplitude, the beating and roughness that we hear will surely change. This means that, in order to preserve the perceived fidelity of all sounds, reproducing systems must be examined with resolutions much finer than the critical bands and their popular incarnations, 1/3-octave filters. Making things worse in the measurement world is the use of fixed-frequency filters. The ears do not have critical bands centered on internationally-standardized frequencies. The critical bands are *continuously variable* responding to the sounds that are present. A high resolution curve smoothed by a 1/3-octave sliding filter is a better approximation to *certain* aspects of hearing. However, in assessing the performance of sound reproduction equipment and rooms, where *all* aspects of hearing are brought to bear, greater resolution is required.

In our work with designing and evaluating loudspeakers there have been numerous occasions when listeners heard problems that were visible only with measurement resolutions of 1/10-octave or better. All of them were very high-Q resonant phenomena. Identifying the precise frequency and Q of resonances helps determine their origins and apply appropriate remedies. In the case of room resonances, high resolution data are necessary in order to design the correct filters to attenuate and damp the resonances. Fortunately, several computer-based measurement systems exist that allow all of this to be done at reasonable cost.

A simple implementation of these principles exists in the Infinity Room Adaptive Bass Optimization System (R.A.B.O.S.™), in which powered woofers and subwoofers are equipped with a single band of parametric equalization. Supplied with the loudspeakers is a CD of test signals, a sound level meter, and a clever device for identifying the Q, or bandwidth, of a problem resonance. Straightforward instructions, or automated website calculations, lead the customer or installer through a sequence of operations aimed at identifying and attenuating the single most objectionable room resonance. In practice, most rooms exhibit only a single powerful resonance, so a single filter can be a great improvement. The system is designed to attenuate only, so the uninitiated cannot try to fill acoustical cancellation holes in the frequency response.

JBL Synthesis systems have had elaborate, semi-automated digital measurement and equalization systems for several years. Revel, and electronic equipment manufacturers Lexicon, Harman/Kardon and Mark Levinson are all on track with various versions of measurement/equalization systems. Customized equalization has been used for years in Harman supplied automotive OEM audio systems. Without it, good sound in a car interior is all but impossible.

The JBL LSR 6300 series of studio monitor loudspeakers. The large three-way loudspeaker and the subwoofer are equipped with an equalizer that is set up using the measuring system shown. This brings room-adapted bass to the recording studio, including the increasingly popular home studios. Better and more consistent recordings should be the result.

The era of room-adaptive loudspeakers has begun. Sadly, many of those currently in the marketplace are poor implementations, thus perpetuating the "EQ is bad" notion. *Bad* equalization is bad. Equalization, done properly, is the icing on the cake.

IN CONCLUSION

The literature of audio continues to be sprinkled with letters, articles and internet discussions debating the merits of science in audio. The subjectivist stance is that "to hear is to believe", and that is all that matters. Some of the arguments conjure images of white-coated engineers with putty in their ears, designing audio equipment and not caring how it sounds, only how it measures. I have never met such a person in my nearly 40 years in audio.

The simple fact is that, without science, there would be no audio, as we know it. Without extensive and meticulous subjective evaluation to confirm the meaning of measurements, there would be no *audio* science, as we know it. Without audio science, audio engineering would be a trial and error exercise. Clearly, one must pay close attention to both the objective and subjective forms of product evaluation, because there is still more to learn.

The recent work showing strong correlations between predicted (from technical measurements) and real subjective ratings is, perhaps, the strongest evidence that we are on the right track. A faith in the scientific method is not a blind faith. It is a faith built on a growing trust that measurements can guide us to produce better sounding products at every price level, for every application. The final proof, though, is in the listening because, as much as we have come to trust measurements, we must always be alert to new variables.

Readers may have noted an absence of references to non-linear distortions. It is not that they are unimportant, but rather that current design practices have reduced them to the point where they are not normally factors in determining listener preferences. However, it does occasionally happen that they are, and that is another reason to listen. In the meantime, the search for more meaningful ways to quantify non-linear distortion continues. Present measures of harmonic and intermodulation distortion are crude in the extreme, exhibiting very poor correlation with how devices sound with music.

Equalization, done properly, is the icing on the cake.

The Harman International loudspeaker companies – JBL, Infinity, Harman/Kardon and Revel – have invested heavily in measurement facilities that allow them to take the fullest advantage of existing audio science. They have invested in talented engineers who understand and respect the scientific method, good sound and great music. They have invested in elaborate listening rooms where they can enjoy and criticize the fruits of their labors. There are people on staff with many years of experience in successfully probing the frontiers of knowledge in product design and audio science, and they continue to push those frontiers.

Perhaps the most gratifying outcome of this research is a genuine improvement in loudspeakers; one that increases the probability that they will sound good in any room. Coupled with this is the recent work on sound field management, which allows us, for the first time in the history of audio, to deliver similarly good bass to several listeners in a room or car. Digital technology has finally reached the cost level where all of this can now be implemented in chips costing a few dollars. Having the technical ability is one thing. Knowing what to do with that technical ability is what really counts. We have both. Stay tuned.

REFERENCES

1. Toole, F.E., “Listening Tests, Turning Opinion Into Fact”, J. Audio Eng. Soc., vol. 30, pp. 431-445 (1982 June).

2. Toole, F.E., “Listening Tests - Identifying and Controlling the Variables”, F.E. Toole, Proceedings of the 8th International Conference, Audio Eng, Soc. (1990 May).

3. Toole, F.E. and S. E. Olive, “Subjective Evaluation”, in J. Borwick, ed. “Loudspeaker and Headphone Handbook - Second Edition”, chap. 13 (Focal Press, London, 2001).

4. Toole, F.E., “Subjective Measurements of Loudspeaker Sound Quality and Listener Performance”, J. Audio Eng. Soc., vol. 33, pp. 2-32 (1985 January/February)

5. Olive, S.E., “A Method for Training of Listeners and Selecting Program Material for Listening Tests”, 97th Convention, Audio Eng. Soc., Preprint 3893 (1994 November).

6. Olive, S.E., “Differences in Performance and Preference of Trained versus Untrained Listeners in Loudspeaker Tests: A Case Study”, J. Audio Eng. Soc., vol. 51, pp806-825 (2003 September).

7. Toole, F.E. and S. E. Olive, “Hearing is Believing vs. Believing is Hearing: Blind vs. Sighted Listening Tests and Other Interesting Things”, 97th Convention, Audio Eng. Soc., Preprint 3894 (1994 Nov.).

8. Toole, F.E., “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).

9. Devantier, A., “Characterizing the Amplitude Response of Loudspeaker Systems”, 113th Convention, Audio Eng. Soc., Preprint 5638 (2002 October).

10. Olive, S.E., P. Schuck, S. Sally, M. Bonneville, J. Audio Eng. Soc.,“The Effects of Loudspeaker Placement on Listener Preference Ratings”, Vol. 42, pp. 651-669 (1994 September)

11. Olive, S.E., B. Castro and F.E. Toole, “ A New Laboratory For Evaluating Multichannel Systems and Audio Components”, 105th Convention, Audio Eng. Soc., Preprint 4842 (1998 Sept).

12. Toole, F.E. and S. E. Olive, “The Modification of Timbre by Resonances: Perception and Measurement”, J. Audio Eng, Soc., vol. 36, pp. 122-142 (1988 March).

13. Olive, S.E., P. Schuck, J. Ryan, S. Sally, M. Bonneville,“The Detection Thresholds of Resonances at Low Frequencies”, J. Audio Eng. Soc. Vol. 45, No. 3 (1997 March.)

14. Seligson, L., "How Consumers Union Tests Speakers", Stereo Review, pp. 62-66, Feb. 1969.

15. Anon, "Loudspeaker Accuracy: CU's Tests", Consumer Reports, p. 456, July 1973.

16. Olive, S.E., "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part I - Listening Test Results." 116th Convention, Audio Eng. Soc., Preprint 6113, 2004 May.

17. Olive, S.E., "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part II - Development of the Model." 117th Convention, Audio Eng. Soc., Preprint , 2004 Oct.

18. Olive , S.E., and F.E. Toole, "The Detection of Reflections in Typical Rooms", J. Audio Eng, Soc., vol. 37, pp. 539-553 (1989 July/August).

19. Toole, F.E., "Loudspeakers and Rooms for Stereophonic Sound Reproduction", Proceedings of the 8th International Conference, Audio Eng, Soc. (1990 May).

20. Schuck, P.L., S.E. Olive, J. Ryan, F. E. Toole, S. Sally, M. Bonneville, E. Verreault, K. Momtahan, "Perception of Reproduced Sound in Rooms: Some Results from the Athena Project", pp.49-73, Proceedings of the 12th International Conference, Audio Eng. Soc. (1993 June).

21. Olive, S.E., P. Schuck, J. Ryan, S. Sally, M. Bonneville,"The Variability of Loudspeaker Sound Quality Among Four Domestic-Sized Rooms", presented at the 99th AES Convention, preprint 4092 K-1 (1995 October).

22. Welti, T., "How Many Subwoofers are Enough", 112th Convention, Audio Eng. Soc. Preprint 5602. (2002).

23. Welti, T. and A. Devantier, "In-Room Low Frequency Optimization", 115th Convention, Audio Eng. Soc., Preprint 5942. (2003).

Sean Olive, Allan Devantier and Todd Welti work with the author in the Harman International, Corporate Engineering R & D Group, in Northridge, California.