Audio – Science in the Service of Art

by

Floyd E. Toole, Ph.D.
Vice President Engineering, Harman International
Senior Vice President Acoustical Engineering, JBL and Infinity

INTRODUCTION

Audio products must sound good. That is a given. However, the determination of what constitutes "good sound" is a matter that has been controversial. Some assert 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". This would place audio manufacturers in the category of artists, trying to appeal to a varying public "taste". Others, like the author, take a more pragmatic view, namely 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. Interestingly, though, this process has created new "artists", the recording engineers, who are free to editorialize on the impressions of direction, space, timbre and dynamics of the original performance, as perceived by listeners through their audio systems. Other creative opportunities exist at the point of reproduction, as audiophiles tailor the fundamental form of the sound field in listening rooms by selecting loudspeakers of differing timbral signatures and directivities, and by adjusting the acoustics of the listening space with furnishings or special acoustical devices.

To design audio products, engineers need technical measurements. Historically, measurements have been viewed with varying degrees of trust. However, in recent years, 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. Measurements inevitably involve 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 unmeasurably 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 considerable gains have been made.

This paper reviews scientific work performed by the author and his colleagues, that aimed to determine the extent to which listeners agreed on their preferences in sound quality and, beyond that, to identify relationships between listener preferences and measurable performance parameters of loudspeakers. Given that loudspeakers, listeners and rooms form a complex acoustical system, some of the effort was necessarily devoted to identifying the aspects of performance that maximize the performance of the entire system in real world circumstances.

ART, SCIENCE AND AUDIO

Science has allowed us to do some remarkable things. It has enabled us to travel safely back and forth to the moon, to drive around Mars, to harness atomic energy for peace and to unleash it for war. Its principles allow engineers to build tall buildings, immense bridges, and vehicles that roll on the earth and those that fly through air and space. These and most other aspects of scientific accomplishment are associated with "things", or theoretical concepts far removed from our everyday lives. However, that same foundation of scientific knowledge, principles, and methodology has given us the ability to enjoy the excitement, emotions and sheer beauty of music, whenever or wherever the mood strikes us.

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.

However, in spite of its many capabilities, science cannot describe music. Beyond the crude notes on a sheet of music, science has no dimensions to measure the evocative elements of a good tune. It cannot technically describe why Pavarotti's tenor voice is so revered, or why the sound of a Stradivarius violin is held as an example of how it should be done. Nor can science distinguish, by measurement, the mellifluous qualities of trumpet intonations by Wynton Marsalis, and those of a music student who simply hits the notes. Those are distinctions that must be made subjectively, by listening. Some scientific effort has gone into musical instruments and, as a result, we are getting better at imitating the desirable aspects of superb instruments in less expensive ones. We are also getting better at electronically synthesizing the sounds of acoustical instruments. However, the determination of what is aesthetically pleasing, remains firmly based in subjectivity and the arts.

Our audio industry is based on a seemingly simple sequence of events. We capture a musical performance with microphones, whose outputs are blended into an electronic message stored on tape or disc, which is subsequently reproduced through two or more 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 point to noises and distortions that were not in the original sounds, rendering even the most eloquent tunes in a brittle and harsh fashion. 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, as we have known it, 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? 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 the science of *electroacoustics* a blend of mechanical and electrical engineering with a dose of physics. Once the signal is in the electrical domain, the science and engineering expertise in *electronics* is brought to bear in preserving the integrity of the musical signal. Contamination by noise and distortions of any kind must be avoided, as it is prepared for storage and when it is recalled from storage during playback.

Emerging from the tape or disc recorder, radio or TV, 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. If there is a seriously weak link in this entire process, this is probably it. 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*, and from that we could predict that there would be problems of the kind that we experience. 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

careful acoustical design and electronic signal manipulations, we are finding ways to make speakers more "friendly" to rooms, both in recording studios and in homes.

THE STEREO PRESENT 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, then, 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 old argument that we only have two ears is valid only in the context of binaural recordings, where it is necessary to send specific sounds to each ear independently.

The excitement and realism of the reproduction is enhanced if we have more channels and more loudspeakers. This is a relatively new development in music recording and, understandably, it will take some time for artists and recording engineers to learn how to use the new format. Examples from the abortive quadraphonics era, and from some current efforts, show that not everyone has learned how to merge multichannel technology with good taste. However, properly used, the combination of digital discrete multichannel technology, and the appropriate loudspeakers can provide the basis for some amazing sound experiences – both realistic and contrived.

And, what might the appropriate loudspeakers be? 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 theories and practices 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 all can be great fun, some of them even very good, but they are all different.

At the reproduction end of things, loudspeakers 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, have been employed in attempts to coax, from a spatially-deprived medium, a rewarding sense of space and envelopment. In the two-channel world, therefore, the artists could not anticipate 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.

Nowadays we have elegant active-matrix technologies, such as Lexicon's Logic 7, and Citation's 6-Axis that allow us to play conventional stereo recordings through multichannel systems, to introduce us to being truly enveloped in sound. Whether the experience is realistic, or even tasteful, will depend on how the recording was made – remember, there are no standards in stereo.

All of this will get much better when recordings are created to be reproduced through multiple channels. Then, if we do not like what we hear, we can legitimately complain to the artist. Among the discoveries in this new era, may be that some of the loudspeaker designs that were flattering to stereo recordings, will be less appropriate for multichannel sound. For example, many listeners have come to favor loudspeakers having wide dispersion, or even multi-directional radiating characteristics, for stereo recordings. The principle at work here is that the reflected sounds in the listening room embellish the sounds from the two loudspeakers, pleasantly enhancing the impressions of acoustical warmth, depth and spaciousness. This is as it should be.

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: first, in movies:

❑ close-up speech, intimate whispers (listener in a strong direct sound field),
❑ being surrounded by reverberant space or cheering fans at a game or concert (multidirectional discrete sounds),

❑ fly-overs, drive-bys, ricochets, etc. (multidirectional discrete sounds), and in musical performances:

❑ being in a superb concert hall or club (direct sounds from an orchestra in front, with reflected and uncorrelated sounds from beside and behind), and

❑ being in the middle of a band (multidirectional direct sounds).

These will be convincing only if the loudspeakers can deliver the appropriate sounds to the listeners' ears. If this "spatial dynamic range", as I call it, is to be achieved in normal – i.e. not acoustically treated – listening rooms, we will very likely need loudspeakers of differing directivities in different locations and, if we are truly fussy, we will need more than five channels.

We have begun a voyage of discovery and, depending on how we approach it, it can be long and tortuous, or short and sweet. If it is approached with an appropriate blend of art and science, it could be the latter. It would be wonderful if the music industry could agree on a standard methodology for multichannel sound. However, based on experience, that is extremely unlikely.

At this point it is necessary to introduce the science 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. Such knowledge is absolutely necessary if we are to make significant progress in designing better products, especially when price is a concern.

THE SCIENCE OF AUDIO

Individual points of view are a part of human nature. They enrich our lives in many ways. It would be a boring world if we all were attracted to the same music, food, wine and people. However, it is a serious disadvantage if one is trying to design products that will be attractive to the majority of the consumer population.

A point of view, commonly expressed, is that sound is "subjective", that we all "hear differently", and therefore not all of us prefer the same loudspeakers. 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 kettle drums. 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? I don't believe it . . . and it doesn't happen that way.

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 the "east coast / west coast" sounds associated with some powerful 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 attraction to reproduce sounds as accurately as possible.

Now, what about those individual preferences in sound quality? This interesting issue was settled by conducting many, many listening tests, using many, many listeners and many, 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 surprisingly clear. When the data were compiled, it turned out that most people, most of the time, liked and disliked the same loudspeakers. There were also differences in the way listeners performed; most were remarkably consistent in repeated evaluations of the same product, while some others changed their opinions of the same product at different listening sessions.

HIGH FIDELITY AND STUDIO MONITOR SPEAKERS

■ ONE STANDARD DEVIATION FOR EXPERIENCED LISTENERS WITH NORMAL HEARING (3% - 5%)

Infinity

FIGURE 1: *Judged on a FIDELITY scale of 0 to 10, terrible to perfect, this figure shows results of many listeners evaluating many loudspeakers from different categories. To the same scale are shown, symbolically, the variability of repeated judgements by experienced listeners with hearing levels close to audiometric zero, and those who exhibited broadband hearing loss. Also shown is the variability of data combined from several listeners with normal hearing.*

Figure 1 shows, as might be expected, a rather large range in performance of the loudspeakers themselves. It also shows that, in repeated judgements, listeners with normal hearing exhibited standard deviations small enough for high statistical significance to be associated with small (about 0.5 scale unit) rating differences between products. This important observation is paralleled by another, perhaps even more important one: that groups of such listeners closely agreed with each other. 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 defined within hearing losses that would not be regarded as alarming by conventional audiometric criteria, see Figure 2. 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 1kHz. 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.
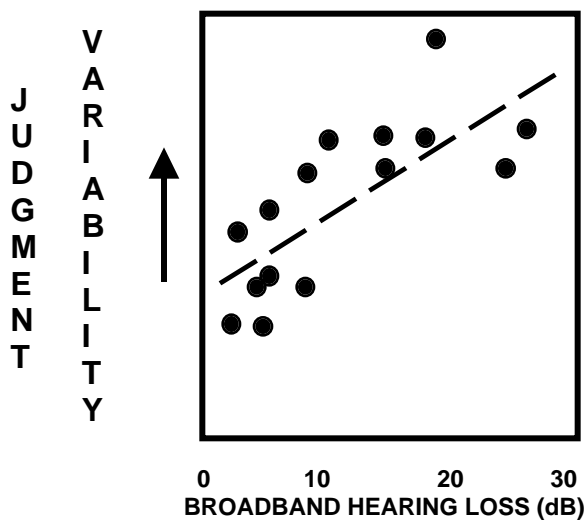

JUDGMENT VARIABILITY

BROADBAND HEARING LOSS (dB)

FIGURE 2

CHOOSE YOUR LISTENERS WITH CARE

FIDELITY

● LISTENERS WITH LOW JUDGMENT VARIABILITY / NORMAL HEARING

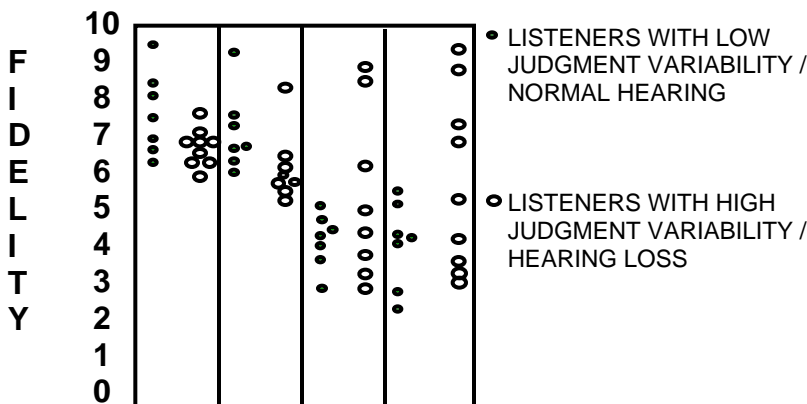○ LISTENERS WITH HIGH JUDGMENT VARIABILITY / HEARING LOSS

FIGURE 3: A comparison of loudspeaker FIDELITY evaluations performed by two groups, one with low variability in their judgments and the other with high judgment

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 3 is that the group of normal-hearing listeners substantially agree in their ratings. Interestingly, the second 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 individual opinions, either C or D could be the best or worst speaker in the world. It appears that their disabilities prevented some of the listeners from hearing certain of the deficiencies. Sadly, some listeners who fall into the problem category are talented and knowledgeable musicians or 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.

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 about 75%, or more, of the population – an acceptable target audience for most commercial purposes. This is not an "elitist" criterion.

PRACTICE MAKES PERFECT – USE TRAINED LISTENERS

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 relationships between technically-measurable parameters and their audible importance became clearer, it was possible to design training sessions that focused on improving 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.

BLIND vs. SIGHTED TESTS – SEEING IS BELIEVING

Knowledge of the products that are being evaluated is generally understood to be a powerful source of psychological bias. In scientific tests of many kinds, and even in wine tasting, considerable effort is expended to ensure the anonymity of the devices or substances being subjectively evaluated. In audio, though, things are more relaxed, and otherwise serious people persist in the belief that they can ignore such factors as price, size, brand, etc. 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 differences that are in the range of extremely small to inaudible. That debate shows no signs of slowing down. 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.
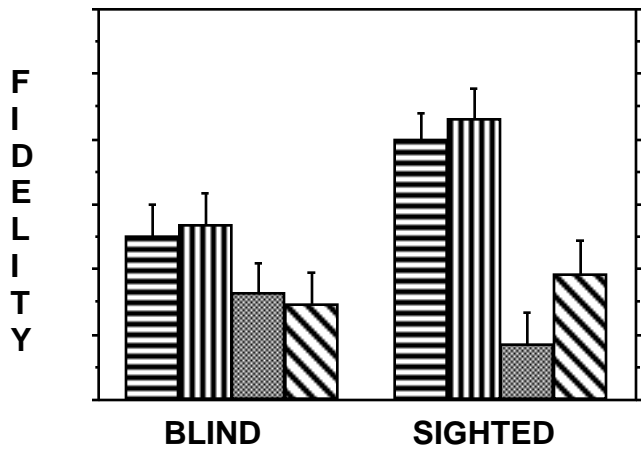
The results are very clear, and strongly supportive of the scientific view. Figure 4 shows that, in subjective ratings of four loudspeakers, the differences in ratings caused by knowledge of the products is as large or larger than those attributable to the differences in sound alone. The two left-hand striped 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 three-piece 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 beautiful 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 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 [6].

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".

FIGURE 4: A comparison of blind vs. sighted evaluations of the same products by the same group of listeners.

"ZOOMING IN" ON THE PROBLEMS

Inevitably, as we make progress in improving products, the listening tests no longer include examples of bad performance. On the 0 to 10, junk to jewels, "fidelity" scale, therefore, one ends up listening exclusively to devices that score in the top ten percent, or so. When reminders of bad sound are removed from the tests, an interesting thing happens: listeners spontaneously expand the scaling of their responses to fill more of the range. In the absence of reminders about how bad things can really be, we become more critical of the relatively good sounds we are evaluating.
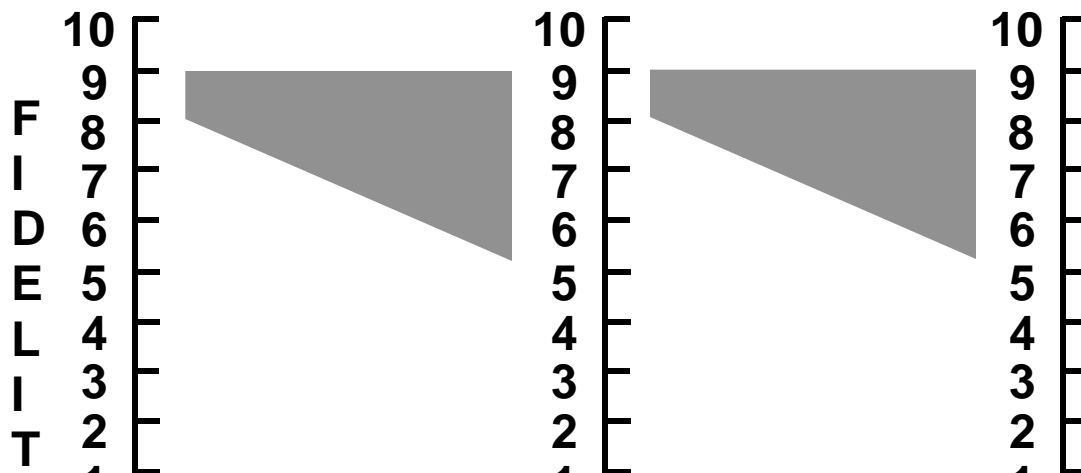
FIGURE 5: What happens when listeners evaluate products that are closely ranked at the top of a range of products. Without any "anchor products" to remind them of how things sound at lower ratings, the response scale expands to fill a "comfortable" range.

A consequence of the phenomenon illustrated in Figure 5, is that, when they are auditioned in isolation, listeners tend to exaggerate the importance of small differences. If one is focusing attention on the small differences, that is a good thing. If, however, one is attempting to arrive at an evaluation of a phenomenon in a manner that relates to its importance in a total system, it is a distorted judgment. Real-world examples of this occur when, for example, one does comparisons of devices that are fundamentally similar. Many electronic devices fall into this category, as do loudspeakers that differ only in minor details. 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.

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.

```
P                1
R       R        9     REALLY LIKE            ■   SLIGHT PREFERENCE
E       A        8
F       T        7     LIKE
E       I        6
R       N        5     NEITHER LIKE NOR DISLIKE  █ MODERATE PREFERENCE
E       G        4
N                3     DISLIKE
C                2
E                1     REALLY DISLIKE         █   STRONG PREFERENCE
                 0
```
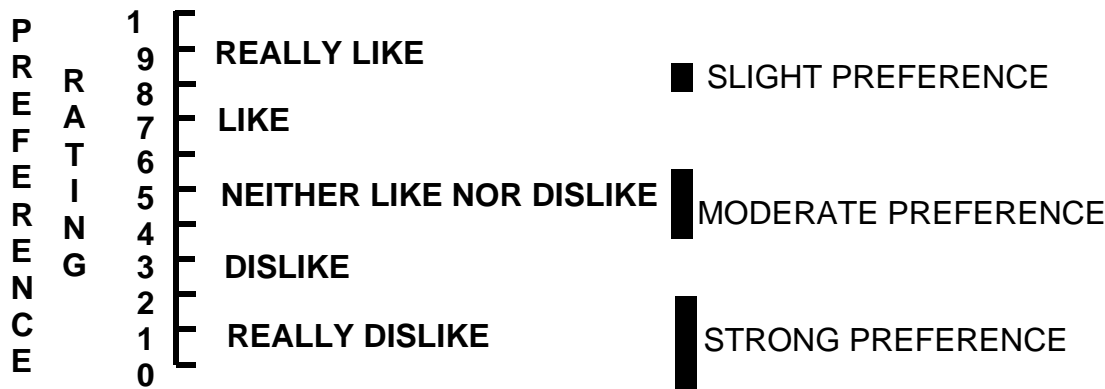
FIGURE 6: The preference rating scale that is used when comparing sounds on a relative, rather than an absolute, basis. On the right are the suggestions for rating <u>differences</u> according to the strength of the preference.
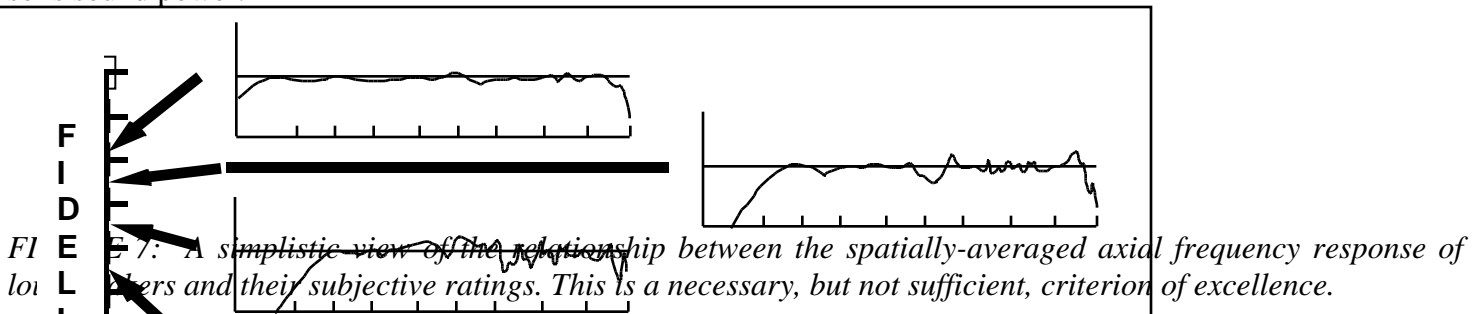
The function of this kind of response scaling is to establish how listeners respond to differences between sounds. Since the scale is not "anchored" by reference to sounds known to be very good and very bad, there is nothing to indicate where they stand in any absolute sense.

TECHNICAL MEASUREMENTS ↔ SUBJECTIVE MEASUREMENTS

With statistically reliable, repeatable, numerical subjective ratings of loudspeakers in hand, it is an unavoidable temptation to look for orderly relationships with technical data on the same products. Figure 7 shows that the axial response of a loudspeaker needs to be very smooth, flat and wide-band in order to achieve high subjective ratings. One can rightly conclude that this is associated with a perception of timbral neutrality - a lack of coloration. But, there is more to the story. Since loudspeakers function in rooms, it is necessary that the sounds radiated in other directions also be well behaved, so that the ears receive similarly neutral sounds after single and multiple reflections from floor, walls, etc. The extension to the requirement, therefore, is that the off-

axis frequency responses, including sound power, also be well-behaved [7,8]. The design target, therefore, is a smooth, flat, axial response, with a constant directivity as a function of frequency. Just as axial behavior, by itself, is an insufficient criterion, so is sound power.

FIGURE 7: A simplistic view of the relationship between the spatially-averaged axial frequency response of loudspeakers and their subjective ratings. This is a necessary, but not sufficient, criterion of excellence.

Since absolute perfection in transducers and enclosures is still a remote possibility, it is important to be able to identify, in measurements, the presence and the audibility of defects such as resonances. In this way designers can choose to build the best product possible or, by making appropriate compromises, the best product at a given price level. The first step in this process is to develop a measurement method that allows the eyes to identify the presence and magnitude of defects that are audible. The second step is to identify the measured level at which the defect is audible – the detection threshold below which the defect ceases to be a problem.
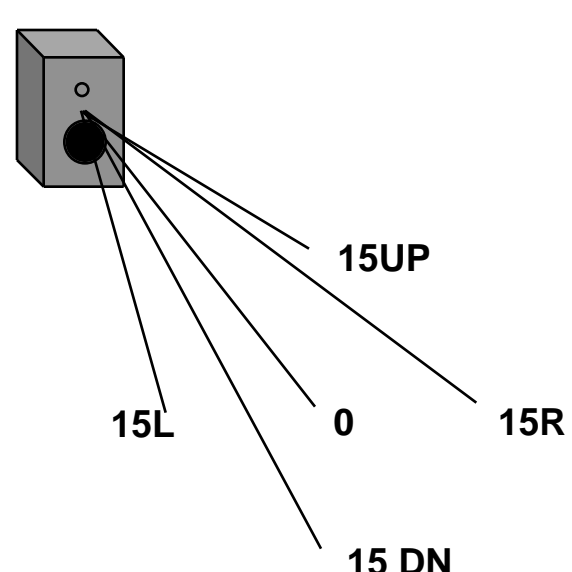

FIGURE 8: A simple form of spatial averaging, in which measurements on several axes can be viewed individually, or in combination.

A frequency response curve contains evidence of both resonances and acoustical interference. It is important to be able to identify which detailed features in the curve are attributable to each phenomenon. Why? Because, in a room, resonances will be easy to hear, and interference will be perceptually attenuated.

Spatial averaging is a simple way to identify resonances since, in the data for each of the microphone positions, the resonances will be relatively unchanged, while evidence of acoustical interference will change as a function of microphone position. When the collection of frequency response curves is averaged, the vis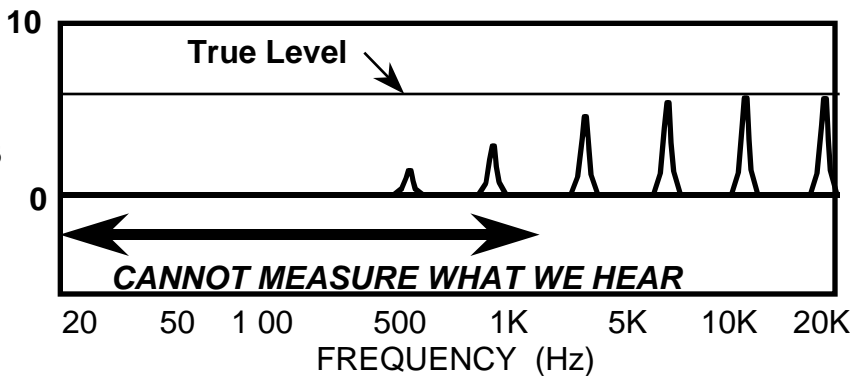ual evidence of resonances remains, while that for the interference is diminished. It is a very simple, but very effective analytical method.

FIGURE 9: A frequency response measurement before (top curve) and after (bottom curve) spatial averaging.

Data processing of multiple measurements, as in Figure 9, allows us to see, in high resolution, the amplitude and "Q" of resonances in complex mechanical and acoustical systems. These data were gathered in an anechoic chamber, where it is possible to measure with high resolution over the entire frequency range.

A practical problem is that many measurements these days are made with FFT or TDS 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 limited frequency resolution, most noticeable at the low frequency end of the spectrum. As a result, it is not possible to see high-Q phenomena at low and middle frequencies.

FIGURE 10: Identical high-Q resonances (Q=50) at equal intervals between 20 Hz and 20 kHz, adjusted to the threshold of audibility, as measured by an FFT measurement system having a time window of 17 ms, corresponding to a frequency resolution of 60 Hz.

Resolution limitations of the kind shown in Figure 10, and worse, are common among loudspeaker designers and reviewers, because of the scarcity and expense of anechoic chambers, and the practical difficulties in measuring outdoors, in nature's own anechoic space.   It means, simply, that many commonly-used and published measurements 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.

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. Spatial averaging adds information, whereas spectral averaging removes spectral details, making curves look smoother and prettier. The sound, though, remains unchanged. Is it any wonder that some people mistrust measurements?
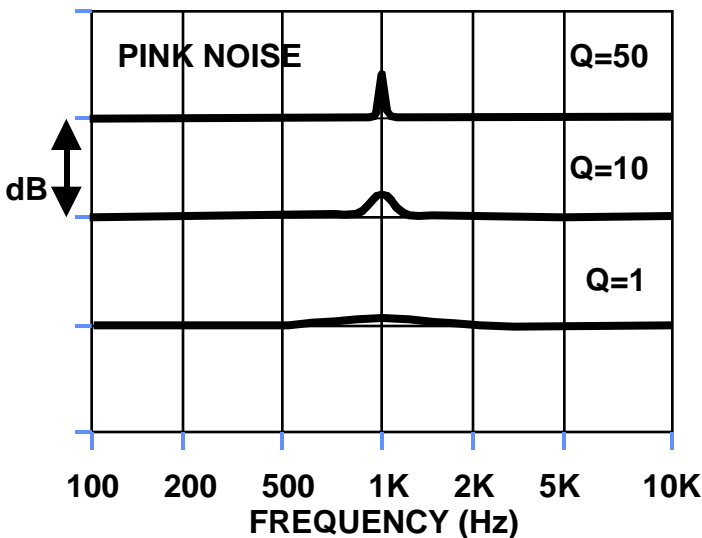
Serious loudspeaker manufacturers need to be able to see, in measurements, anything that might result in an audible defect. In terms of the measurement of frequency response, it means expensive anechoic chambers, facilities for gathering data over an entire sphere, and elaborate post-processing of the data. Doing this quickly and accurately is neither inexpensive nor simple.

THE AUDIBILITY OF RESONANCES

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 of voices and instruments. Medium- and low-Q resonances define the timbres of sounds, 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.

In order to be effective, design engineers need to know when a resonance is present, and if it is audible. The techniques described above, help to identify the presence of resonances. Reference 9 describes, in detail, how the audibility of resonances is related to measurements.

FIGURE 11: The amplitude response measured in an otherwise perfect system, after a single resonance has been added at the threshold of detection, when listening to the most revealing signal, pink noise.

In that study, resonances of different Q, at different frequencies, were added to an otherwise excellent system, and the amplitudes at which they were just detected by listeners were determined by experimentation. Different kinds of program material yielded different thresholds, as did different listening environments.

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.

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 11, the low-Q resonances were detectable at much lower amplitudes than resonances of higher Q. If this is so, it means that a treasured belief is in jeopardy. The popular belief is that prolonged ringing, by itself, is a reliable indicator of an audible problem. To test this, we performed the measurements in the time domain, with the following results.
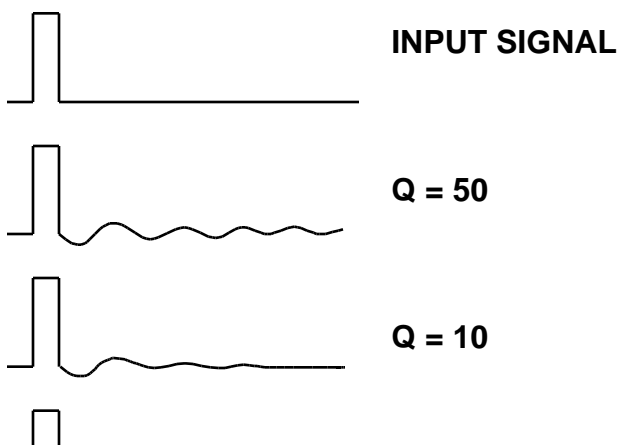
FIGURE 12: Pulse responses of an otherwise perfect system (top curve) to which resonances have been added at the thresholds of detection.

Looking at Figure 12, it is important to note that, in terms of audible changes in timbre, these responses are all equal. To the eyes, though, the long tail of the Q=50 resonance is quite alarming, and the Q=1 response appears almost perfect. How can this be so?

One important factor would appear to be related to the portion of the spectrum influenced by a resonance.

High-Q resonances are very narrow-band phenomena. In order for one of these to be energized by music, a sound would need to be closely centered on the frequency, and remain there long enough to impart significant energy to the resonant system. The higher the Q, the longer is the "build-up" period. In music, pitches are constantly changing, and voices and instrumental sounds frequently have vibrato, a fluctuating pitch. Such sounds would more often drive a low or medium Q resonance to maximum output, than they would a high-Q resonance. Statistically, therefore, with music and speech, lower-Q resonances would be heard more frequently than those of higher Q.

The apparent contradiction between perception and measurement, then, begins with the observation that, as conventionally measured, the frequency response is a "steady-state" measurement, showing the resonance outputs at their maximum amplitude. With music, high-Q resonances are rarely driven to their maximum outputs, and so are less audible than the measurement indicates. The problem is not that the measurements are wrong, or irrelevant, it is that they are non-linearly related to the perceptual mechanism in humans, and therefore must be interpreted. Time-domain measurements are similarly problematic, since they suggest audibility in the prolongation of ringing. Such phenomena are wonderfully visible in pulse responses, tone-burst responses and the highly ornamental "waterfall" diagrams that digital measurement systems permit. Truth is that, without careful interpretation, these are just as misleading as the frequency responses. The conclusion is that, in practical situations, if a resonance does not make itself apparent in an accurate, high-resolution spatially-averaged, frequency-response measurement, then it is probably not audible. If a resonance is visible in a frequency response measurement, its audibility must be assessed by comparison with data from Figure 11, or better, from the detailed analysis in reference 9.

An interesting fact now emerges: that the conventional method of specifying the excellence of frequency response - $\pm\,x$ dB – is almost useless unless the tolerance is very, very small. For equal audibility, high-Q phenomena could be $\pm\,5$ dB, while moderate-Q resonances could be $\pm\,3$ dB and low-Q and other broadband deviations could be $\pm\,0.5$ dB. Clearly, frequency response curves must be *interpreted*, there is no simple "catch-all" kind of tolerance specification that is truly meaningful. Such is life.

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 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 move uniformly, as a piston, at all frequencies, so measurement at a single point can be misleading. In practice, portions of the surface 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 of playing safe, has been to build enclosures from massive, dense and stiff materials. If cost, size and weight are not considerations this method works well. However, enclosures with excellent acoustical performance can be built from mundane materials, at much lower cost, with due regard to modal prediction and analysis.   In lower cost wooden enclosures, and especially with plastic enclosures, this becomes normal engineering practice.

LOUDSPEAKERS AND ROOMS

As important as the enclosures behind loudspeaker diaphragms are, it is the ones in front that, as often as not, give us serious problems. The listening room is the final audio component, and it is the one over which the loudspeaker manufacturer has little or no control. The first important step is to design loudspeakers so that they have a reasonable chance of sounding good in a room.

In a room, 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 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 directivity as a function of frequency.

The directivity itself can take different forms for different applications, but is should not change with frequency. This is a difficult requirement to fill, and it requires special measurement techniques to allow engineers to judge their success as products are developed. In brief, it requires that we know:

❑ the nature of sounds radiated in the direction of the listener (the on-axis/listening-window performance),
❑ the nature of sounds that will be reflected from adjacent room surfaces (vertical and lateral "early-reflection" performance), and
❑ the nature of sounds that will generate the diffuse reverberant sound field (the sound power).

From these the directivity index can be calculated. It is the combination of these that matter; any one by itself is insufficient data.

Loudspeakers that meet these requirements also tend to be the ones that win listening tests, and there can be no better reinforcement for a methodology than that. One of the factors in listening tests is imaging, and that raises an issue for which there is not a complete answer at the moment: what is the ideal directivity?

LOUDSPEAKERS FOR STEREO

As discussed earlier, 40-some years with two-channel stereo have yielded nothing in the way of a clear direction. Although the majority of loudspeakers sold are traditional forward-facing "cone and dome" designs, it is also true that the majority of listeners are not very critical about the imaging of their systems. Among those who are, the "high end audiophiles", those designs figure strongly in their preferences. However, so do designs of very different kinds, like, dipole, bipole and directional horns. The perceptual consequences of speakers this diverse are not subtle. Bipole designs have approximately omnidirectional sound radiation properties, and therefore produce energetic reflected sound fields in listening rooms. Conventional forward-facing systems, and horn-loaded systems will place the listener in a sound field in which the direct sound is more prominent. The more directional the system, the more dominant will be the direct sound.

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 the specific images can be rather vague – in other words, not unlike real concerts. A positive effect of this vagueness is that the stereo listening region is enlarged.

However, there is also a category of listeners who respond unfavorably to this kind of reproduction, and prefer to have a very specific, almost pinpoint, sense of image position. Interestingly, this category includes many recording engineers who, in their studios, require that they be able to hear, very precisely, the results of their manipulations. 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 the more spacious version of stereo. So, go figure.
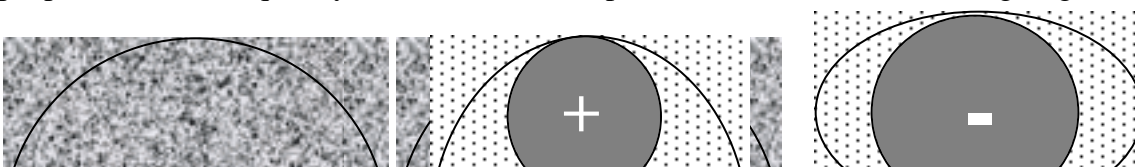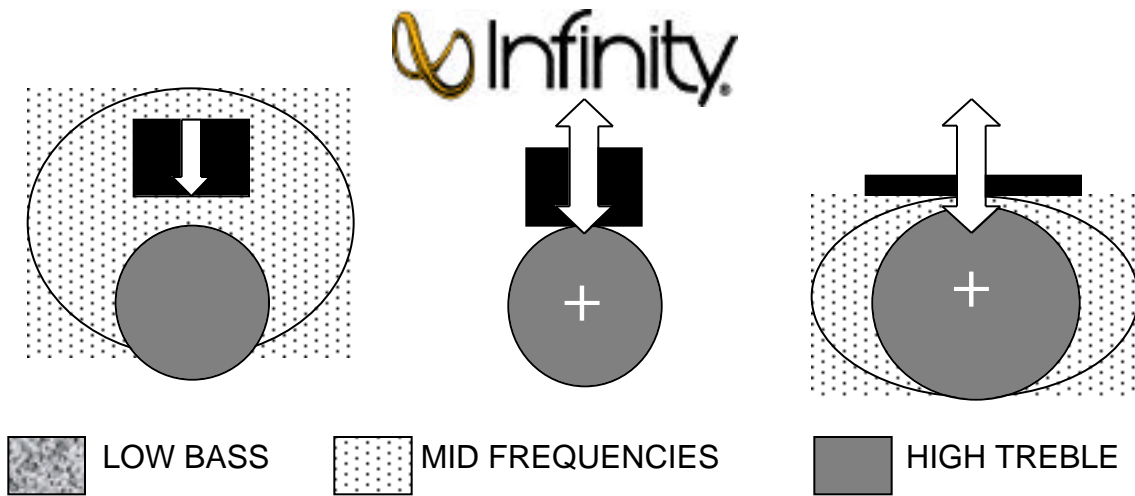
FIGURE 13: Forward-firing (left), bidirectional-in-phase – Bipole (center) and bidirectional out-of-phase – Dipole (right), are just some of the very different directional patterns that are used in loudspeakers for stereo systems. The differences in imaging precision, spaciousness, and soundstage depth are not subtle. Once selected, though, the characteristics are applied to all kinds of music, whether it is appropriate or not.

LOUDSPEAKERS FOR MULTICHANNEL AUDIO

With the introduction of multichannel audio, things are no less complicated. Multichannel sound should mean that things become more controlled, that listeners stand a better chance of hearing the senses of direction and space that the artists created. In film applications, this is actually attempted. The film industry, has had basic standards for playback systems and environments for many years. THX improved on this, and attempted to translate it into the home. For the Dolby Surround films of that era, it was moderately successful. However, things now are much more confused, with Dolby Digital and DTS sound tracks and music recordings, and the Audio DVD around the corner. One senses an impending free-for-all in which anything goes.

If it were approached logically, however, it seems that there is a scheme that makes sense. The purpose of multiple channels with loudspeakers located around the listeners, is to allow for a large variety of predictable localization and spatial effects. Since one of these is a sense of intimacy, wherein the sound from the front loudspeakers does not energize the listening room reverberant field, there is a requirement for front loudspeakers that are predominantly forward firing, with directional control in both vertical and horizontal planes. The well-known THX requirement has been for some directional control in the vertical plane only. Many of the implementations have been somewhat less than ideal; simple vertical arrays of drivers cause severe lobing at the off-axis angles at which floor and ceiling bounces occur. If we are to address this issue properly, we need to incorporate horizontal directional control as well – our ears are in the horizontal plane, and it is horizontal reflections that are primarily responsible for the impressions of spaciousness [10]. We also need to focus on how well the speakers behave at the off-axis angles of importance – the adjacent boundary reflections and sound power.

Predictable directional control is possible with horns, waveguide-loaded tweeters or complex two-dimensional arrays, all of which can deliver excellent sound, using today's technology. The alternative, unattractive in most practical situations, is to move quantities of sound absorbing material into the room and cover large areas with it. The lack of attraction has two components, visual and acoustical. Visually, areas of sound absorber run contrary to popular themes of interior décor. Acoustically, absorbers dissipate sound energy that one has paid good money to create, thus making the speakers work even harder.

In a multichannel system, the impressions of space and envelopment are provided by loudspeakers positioned to the sides of the listeners. Here we enter hostile territory, with monopoles, dipoles, tripoles and quadrapoles attempting to be the perfect solutions. Truth is that, while all of these names have meaning in the physics of sound, the products bearing them are only crude approximations to these forms of acoustic behavior. What really is at issue here is the matter of whether the listener should be in a predominantly direct or

predominantly reverberant 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, containing many reflections. If the decorrelation is in the recording, or is added electronically, spaciousness can be perceived in headphone listening. It can also be convincing through conventional forward-facing surround speakers. However, much (most?) existing recorded material is deficient in this respect, and additional decorrelation tends to be beneficial. Using multidirectional (or multiple) surround speakers is a good method of adding decorrelation through multiple reflected sounds. The actual performance, however, is dominated by the geometry and reflective properties of the room.

At present, movies are still made assuming that audiences will experience a multiplicity of speakers down the sides of the cinema, even though the digital discrete surround channels are all wideband and full range. At home, this argues for multidirectional surround speakers, strong electronic decorrelation, or multiple direct-radiating monopoles – i.e. business as usual. However, the notion of five identical loudspeakers is gaining favor. Multichannel music, still very much in its infancy, also supports the idea of five identical loudspeakers. DTS, in its multichannel music demos, goes a step further, and has been promoting the idea of moving the surround speakers to the rear of the room, mirroring the left and right fronts, and delivering sounds of individual musicians to these locations.

At this point the audience divides into two camps – those who like having musicians behind them, and those who do not . . . and the feelings can be very strong. Some see this as a cause to criticize multichannel audio. This is where we must invoke the old adage: if you don't like the message, don't shoot the messenger. Multichannel audio systems are just delivery systems [11].

SPEAKERS FOR MUSIC AND SPEAKERS FOR MOVIES

Are there fundamental differences between loudspeakers designed for movies and music? This oft-repeated concern is really the *wrong question*. The real question is: are there fundamental differences between loudspeakers designed for stereo and those designed for multichannel systems? The answer is that there may be.

Can multichannel delivery systems work equally well for music and films? They must or we will have consumers up in arms, and for very good reasons. A good loudspeaker is a good loudspeaker, and there is no practical reason for different standards of performance in loudspeakers intended for films and music. There *are* reasons for differing directivities and, depending on how the loudspeakers are used, listeners will experience different imaging and spatial effects depending on the manner in which loudspeakers radiate their sound into rooms. That has always been true for stereo, and it remains so for multichannel systems. Thus, the choice of loudspeaker directivity should not alter ones expectations of inherent sound quality.

GOOD BASS IN ROOMS – A BASIC PROBLEM

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 [12,13,14,15].

This 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 projects and products over the years. Short of hiring an acoustical consultant, and being willing to rearrange the furniture and possibly the walls, what can be done?

The short answer is "equalize". At this point, some people's hackles are rising, I am sure. Equalization has acquired a bad reputation over the years. To some proponents, it is a "cure all"; set up a microphone, follow the dancing lights of a "real-time analyzer", reach for the trusty multifilter equalizer and create a pretty curve. Such an exercise is almost certainly doomed to disappoint. Steady-state room curves are "dumb" measurements, in that the microphone simply adds up all of the incoming sounds, from whatever direction, after whatever time.

One-third octave analysis, which is typical of these devices, is a crude measurement.   Two ears and a brain are much more sensitive and analytical, directionally, temporally and spectrally.

To be successful, equalization should address the problems it has the potential to remedy. There are some things equalization can and cannot do:

1. It cannot make a poor speaker sound good. With laboratory measurements to work from it might make a good speaker sound better. However, there is nothing that an average consumer can measure in a listening room that would reveal problems in a speaker that can be repaired by equalization. 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.

2. At frequencies below about 300 to 500 Hz, the system performance is dictated by the shape and size of the room, and the position of the speaker and listener within it. Three factors are interactively operational here:

   ❑ solid-angle gains (the proximity of the speaker and listener to adjacent room boundaries: walls and floor.) Addressing this requires a broadband "tone-control" kind of equalization.

   ❑ room resonances, which can cause strong peaks and dips, some with quite high "Q". Before these can be addressed, the resonances must be identified within a complicated confusion of peaks and dips caused by acoustical interference.   It is always unwise to try to fill dips – acoustical cancellations can be "bottomless pits".   Prominent peaks can be individually addressed with parametric filters set to the appropriate center frequency, Q, and gain/loss. Identifying those parameters accurately requires high-resolution measurements; much more detail than is revealed by the traditional 1/3-octave "real-time analyzers".

   ❑ Acoustical interference caused by the interaction of many reflected sounds within the room. These are non-minimum-phase phenomena, and they cannot be addressed with equalization. Fortunately, our ears are less sensitive to their effects than our measurement systems, so we really need to find a measurement process that diminishes their visibility.

Successful equalization begins with good measurements of what is happening in the room. Broadband spectral trends are easily identified even in elementary real-time analyzer (RTA) measurements. Separating resonances from acoustical interference requires spatial averaging, just as it does in speaker design. For this one needs to make measurements at several different locations throughout the listening area, and they need to be averaged. This requires data acquisition and post processing, it cannot be done with several microphones plugged into a simple mixer. The measurements need to have high resolution in the frequency domain, if there is to be any hope of identifying the key parameters of problematic resonances. For this job, the RTA fails.

Successful equalization requires a good equalizer. Ideally, this would be a multiple parametric-filter type.

The JBL Synthesis home theater systems incorporate all of the required characteristics for successful equalization.   The dedicated digital controller has 95 individually-configurable parametric filters distributed among the 5.1 channels. In the laboratory, based on high-resolution spatially-averaged anechoic measurements, some filters are preset to address small residual problems in the speakers themselves. The equalizer has helped to create better speakers. Once the system is installed in the customer's home, a trained installer arrives with a custom measurement system to adapt the system to match the room at low frequencies. The system employs five microphones connected to a multiplexer, coupled to a laptop computer which, in turn, is coupled to the digital processor. In a carefully controlled sequence, the appropriate test signals are sent through each of the channels, the measurements are compared to predetermined "target functions", and a set of filters is automatically designed that will allow the system performance to approach the target with minimum error. Built into the system are safeguards that attempt to prevent it from trying to equalize the unequalizable. Manual override is always an option, so human intervention can modify a bad decision by the computer, or accommodate customer preferences in spectral balance.

At present, this is an expensive and cumbersome system. However, we are learning from our experiences in the real world. We are learning what the optimum target functions are, how best to instruct the adaptive

equalization process, and what we may be able to leave out of a cost-reduced version of the system. Ideally, this should become a standard feature in popularly-priced active woofers and subwoofers.

A LITTLE FOURIER ANALYSIS

Before leaving this topic, let us address one of the common misunderstandings of equalization. It is frequently asserted that equalizers, because they are filters, add ringing, in the time domain. That is a fact. The assertion continues that, in addressing a frequency response problem, one may damage the transient behavior of the system. That may or may not be the case.

Resonances in speakers and rooms also ring. *If* the filter is used to correct a resonance, both the problem and solution are minimum-phase phenomena. *If* the measurement of the problem has been made accurately, and the parametric filter solution has been designed to match the problem, *then* the ringing of the filter is equal and opposite to the ringing of the problem resonance. The ringing is gone, just as the measured peak in the frequency response has been eliminated.

The real trick in this is trying to ensure that the filters are not used to "fix" something that cannot, and perhaps need not, be fixed by this means. In the days of RTA's and 1/3-octave multifilter equalizers, success of this kind was simply not possible. With today's elegant technology, it is possible, but difficult.

SUBJECTIVISM vs. OBJECTIVISM – IN CONCLUSION

In this lengthy summary we have covered a lot of topics. Much of it was matter-of-factly technical, driven by data and the need to measure, and much of it was subjective, driven by the desire to understand what we can hear. All of it was oriented towards creating loudspeakers that sound better.

The literature of audio continues to be sprinkled with letters and articles 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 30 years in audio science and engineering.

The simple fact is that, without science, there would be no audio as we know it. Without extensive and meticulous subjective evaluation, there would be no *audio* science as we know it. Without audio science, audio engineering reverts to trial and error. So, where does this leave us? Clearly, to be successful in this business, one must be actively involved with both of the objective and subjective sides.

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 proof, as always, is in the listening, and one MUST listen.

The Harman International loudspeaker companies, JBL, Infinity, 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 are equipped to continue those investigations, to push those frontiers.

The arrival of multichannel audio for films required some adjustments in the performance objectives of speakers, certainly at the high end. Multichannel music is another, as yet ill-defined, challenge. More speakers in rooms, means less consumer tolerance for large boxes. Merging loudspeakers with rooms is not easy, and it is the one remaining large challenge for our industry. We are working on all of these fronts. Stay tuned.

REFERENCES

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

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

3. "Subjective Evaluation", F.E. Toole, in J. Borwick, ed. "Loudspeaker and Headphone Handbook - Second Edition", chap. 11 (Focal Press, London, 1994).

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

5. "A Method for Training of Listeners and Selecting Program Material for Listening Tests", S. E. Olive, 97th Convention, Audio Eng. Soc., Preprint No. 3893 (1994 November).

6. "Hearing is Believing vs. Believing is Hearing: Blind vs. Sighted Listening Tests and Other Interesting Things", F.E. Toole and S.E. Olive, 97th Convention, Audio Eng. Soc., Preprint No. 3894 (1994 Nov.).

7. "Loudspeaker Measurements and Their Relationship to Listener Preferences", F.E. Toole, J. Audio Eng, Soc., vol. 34, pt.1 pp.227-235 (1986 April), pt. 2, pp. 323-348 (1986 May).

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

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

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

11. "The Future of Stereo", F.E. Toole, Part 1, Audio, vol 81, no.5, pp. 126-142 (1997 May), Part 2, Audio, vol.8, no.6, pp. 34-39 (1997 June).

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

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

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

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

5/8/98 rev. 8/19/99