Loudspeakers and Rooms for Sound Reproduction—A Scientific Review*

FLOYD E. TOOLE
(ftoole@harman.com)

Harman International Industries, Inc., Northridge, CA 91329, USA

The physical measures by which acousticians evaluate the performance of rooms have evolved in large performance spaces—concert halls. They rely on assumptions that become progressively less valid as spaces get smaller and more acoustically absorptive. In listening rooms the loudspeakers and the rooms interact differently below and above a transition region around 300 Hz, similar to the Schroeder frequency in large rooms. Above this transition we need to understand our reactions to reflected sounds; below it the modal behavior of the space is the dominant factor. A review of the scientific literature reveals that natural reflections in small rooms are at levels where they are perceptible, and their subjectively judged effects range from neutral to positive. At low frequencies the long-standing problem of room resonances can be alleviated substantially through the use of multiple subwoofers, thereby providing similarly good bass to several listeners in a room. A provocative observation has to do with human adaptation to the complexities of reflective rooms, and the extent to which it allows us to localize sounds correctly in direction and distance, and to hear much of the true timbral nature of sound sources. In the case of loudspeakers, an analysis of comprehensive anechoic data is found to be sufficient to provide a good prediction of sound quality, above the low-bass frequencies, as subjectively judged in a normal room. Although the interactions of loudspeakers and listeners in small rooms are becoming clearer, there are still gaps in our understanding. A number of these are identified and are good opportunities for future research.

0 INTRODUCTION

The design of loudspeakers, listening rooms, and a combination of these for sound reproduction has evolved with relatively little direction from the acoustic and psychoacoustic sciences. Only recently has significant effort been put into understanding the relationships between what is measured and what is heard in rooms of domestic size. The acoustical measures by which the excellence of rooms is judged originated in investigations of performance spaces—concert halls and auditoriums. The sound sources, voices and musical instruments, were considered, as a group, to be essentially omnidirectional, and acoustical measurements in such spaces use omnidirectional sound sources. The premise underlying these measures was that the later reflections in the sound field developed into something that was essentially diffuse and statistically random, and that this diffuse—reverberant—sound field extended to the bounds of the space. Reverberation time, a measure of the rate at which this sound field decays, has been integral to acoustic studies of large architectural spaces ever since Sabine's pioneering work over a century ago.

More recent studies have examined which aspects of the sounds arriving at listeners' ears contributed to each of the identifiable perceptions that, in combination, yielded a satisfying concert-hall experience. So even though the sound fields in large, relatively reverberant halls were strongly diffuse, it was recognized that human listeners could respond with some degree of independence to components of that sound field occurring at specific times, or arriving from specific directions. Consequently there now are several measures of sound as it decays in specific time intervals, and the proportions of sound arriving from different directions, all as functions of frequency [1].

The essential point about this situation is that the architectural space—concert hall, auditorium, or jazz club—is part of the original performance. To such an extent is this so that classical composers routinely adapted their compositional styles and instrumentation to fit the acoustical character of specific performance spaces [2].

In sound reproduction the situation is different. Measured reverberation times in domestic spaces fall in the range of 0.3–0.6 s, in contrast to the objectives for performance spaces of about 1.5–5 s, depending on the type of

*Manuscript received 2006 February 17; revised 2006 April 10.

music. This means that it is unlikely that the decaying sound field in a listening room would significantly impact the apparent decay of a concert-hall recording, or those made in larger recording studios.

In addition, loudspeakers have well-defined directional characteristics and most are significantly more directional than groups of musical instruments. The small rooms in which we listen for recreation or monitoring have larger scattering objects and more absorbing material in them (in proportion to their height and volume) than performance spaces. The absorbing material is also not distributed uniformly, but is concentrated in rugs, carpets, drapes, large pieces of furniture, and so on. This means that the sound field cannot be diffuse and, since all rooms are different, it is clear that not all consumers are hearing the same reproduced sounds. Yet there appears to be widespread pleasure and satisfaction with music and movies, even to the point of some general agreement on which ones sound good.

Recording control rooms have drifted in several directions driven by fashion, convenience, and opinion, and they now cover the range from home studios that acoustically resemble domestic listening spaces, through various "engineered" spaces (reflections enriched by diffusers or depleted by deflection or absorption), to those that are nearly anechoic. Because the control room is a factor in determining what is recorded, it is clear that recordings are not created in a standardized fashion. Yet successful records are made in all of these venues, and one would have great difficulty trying to guess the monitoring circumstances by listening to a recording.

Therefore in spite of gross acoustical variations in the recording and reproduction domains, the music industry has prospered for decades. Could it be that good music and musicianship can overcome this acoustical anarchy? Or could it be that humans have a remarkable ability to recognize and compensate for these variations?

Whatever the answer, it seems reasonable to strive to achieve sound reproduction that is essentially independent of the acoustical playback space. Timbral accuracy and impressions of direction, distance, and space ought to be delivered by the multichannel reproduction system with minimal variation attributable to the listening space. The loudspeakers and the listening room should therefore be neutral conveyors of the artistic experience—to both professionals and consumers. Neutral, in this context, does not mean acoustically "dead." We need some room sound, but what kind of room sound?

It is interesting to note, at this stage, prescient comments by some early workers in the field. Gilford, in 1959 [3], studied how well recordings of speech made in different BBC studios survived the transition to different broadcast monitoring environments. Observing that they survived very well, he concluded: "The fact that the listening room does not have a predominant effect on quality is very largely due to the binaural mechanism." In contrast, he showed that test sounds generated in a recording studio, picked up by a studio microphone, reproduced in a listening room, and then quantified using a microphone at the listening position in the playback room, showed that the listening room "had the principal effect." Conclusion: we measure differences that we seem not to hear. His colleague, James Moir, added in discussion: "Finally, in my view, if a room requires extensive treatment for stereophonic listening there is something wrong with the stereophonic equipment or the recording. The better the stereophonic reproduction system, the less trouble we have with room acoustics."

These observations imply that some of the problem lies in our interpretations of measurements made in small rooms. The horrendously irregular steady-state "room curves" that we see simply do not correspond to what we hear. Did our problems begin when we started to make measurements? Are we incapable of hearing these things? Or is it that we hear them, but they simply become part of the acoustical context within which other acoustical events occur, and we have some ability to separate the two? The answer turns out to be some of each.

Once a loudspeaker is in a room, it may be impossible to identify the contribution that the loudspeaker and the room each make to the combined measured result. Without those separate data, appropriately targeted remedial measures for problems are not possible. All of this can only be deciphered through an understanding of how the sound from a loudspeaker radiates into three-dimensional space, is communicated through a room, and then is perceived by a listener. The performance of a loudspeaker is much more complex than anything revealed in an on-axis anechoic measurement. The perceptual processes of two ears and a brain are vastly more complex than anything revealed in a room curve or a reverberation time.

We need to start again. The following is a review of acoustic and psychoacoustic research from disparate fields, put into a framework that is familiar to the audio community. The result is intriguing. It turns out that some of our common practices are less than optimum, and some popularly held notions might need to be revised. It will be clear that not everything is known, and that there are several opportunities for significant and useful research.

1 A PERSPECTIVE ON SOUND FIELDS IN ROOMS

Reading some of the current literature on the acoustics of small rooms, one could easily believe that reflected sound is a serious problem in need of immediate and expensive acoustical treatment. Yet attending live performances in the world's great concert venues is to experience almost nothing but reflections, and we would not wish to be deprived of any of them.

To understand the reason, one must consider the acoustical nature of both sound sources and human listeners. Examining the frequency-dependent directivity of musical instruments, it immediately becomes clear that no single axis is an adequate representation of the timbral identity as heard by the audience in a highly reflective concert venue [4], [5]. Deliberately reflective recording studios spatially integrate the differing off-axis sounds into a pleasing whole. Still, microphone placement is critical, and some amount of equalization in a microphone channel is a common thing—the portion of the sound field sampled by the microphone often exhibits spectral biases that are not in the overall integrated sound of the instrument.

At the receiving end we have ears with frequency responses that, on a single axis, are anything but flat. More important, they are different on different angular axes. The complex forms of these are known as head-related transfer functions (HRTFs), and a quick look at them reveals why we would not wish to be forced to listen only along a single axis, as in an anechoic chamber or outdoors [6]. HRTFs change as a function of the incident angles of incoming sounds, helping us to localize where sounds are coming from, but the perceived timbres of sounds in rooms are the result of spatial averaging, that is, reflected sounds arriving at our ears from many angles. These reflected "repetitions" of the direct sound have a second benefit, increasing our sensitivity to the subtle resonances that give sounds their distinctive timbres [7].

So when an investigation reveals that a reflection at a certain level relative to the direct sound is just audible, as in a threshold experiment, this is not necessarily an indication that problems have begun. More likely, it could indicate the beginning of something perceptually interesting and beautiful. As we get into the details, it will be seen that reflections from certain directions, at certain amplitudes and delays, are more or less advantageous than others, and that collections of reflections may be perceived differently from isolated reflections. We humans like reflections, but there are limits (too much of a good thing is a bad thing) and ways to optimize desirable illusions.

1.1 Acoustical and Psychoacoustical Sense of Scale

In the course of this review it is useful to understand the origin of the scientific knowledge that will be discussed. Fig. 1 is a pictorial portrayal of the scope, showing spaces large and small, tall and short. The science underlying room acoustics originated from very different motivations. Concert halls should ensure the delivery of sound with high sound quality and musical integrity to all members of a large audience. In work spaces investigations were driven by the need to understand the sound propagation of noise from HVAC systems and manufacturing machinery. Smaller spaces in homes, studios, and cars are used for listening, as part of the creative process of recording, or to enjoy the results. The physics of sound propagation are the same in all of them and the processes of binaural hearing are the same throughout. Still, it is unlikely that exactly the same measurements and interpretive rules will be equally useful in all of these spaces. From the perceptual point of view it is an interesting conceit that we try to reproduce the illusions of a vast concert hall in a car, and yet it can work remarkably well. This suggests that there may be some "elasticity" in the relationships between what we measure and what we hear.

1.2 Diffuse-Field Theory—Large Rooms

Classic concert-hall acoustical theory begins with the simplifying assumption that the sound field throughout a large relatively reverberant space is diffuse. In technical terms that means it is homogeneous (the same everywhere in the space) and isotropic (with sound energy arriving at every point equally from all directions). That theoretical ideal is never achieved because of sound absorption at the boundaries, by the audience, and in the air, but it is an acceptable starting point. Absorption in these spaces is minimized in order to conserve the precious acoustical energy from musical instruments and voices. An active reflected sound field ensures the distribution of that energy to all seats in the house. The challenge is to preserve the sound energy in the reflections without obscuring the temporal details in the structure of music. This is why reverberation time remains the paramount acoustical measure in performance spaces.

It is worth noting that in the calculations of reverberation time, it is assumed that the acoustical activity occurs on the room boundaries, and that the volume of the room is empty. In a concert hall the height is such that the audience can be treated as a "layer" of material with a certain average absorption coefficient placed on the floor.

Fig. 2 shows a classic portrayal of the inverse-square-law decay of the direct sound from an omnidirectional sound source (located well away from room boundaries) until it encounters the underlying steady-state reverberant sound field that is assumed to extend uniformly through-

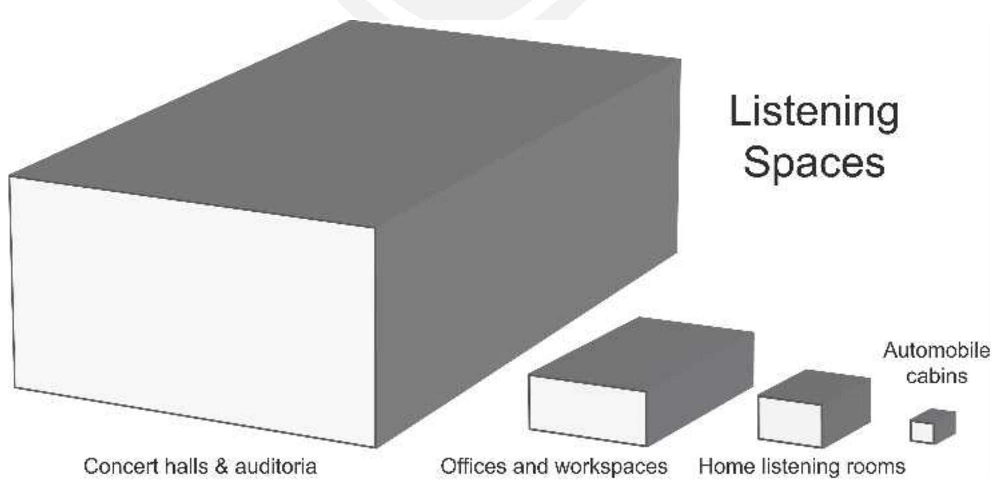

Fig. 1. Pictorial representation of the range of sizes and shapes of spaces for which we need acoustical measurements and rules explaining their relationships to perceptions of direction, distance, space, and sound quality.

out the space [8], [9]. The distance from the source at which the direct sound equals the level of the reverberation is the critical distance (also known as reverberation distance, reverberation radius). The dashed-curve sum of these is what would be measured by a sound level meter as it is moved away from the source—a "draw-away" curve. In real halls the level of the reverberant sound field is not constant, but gradually falls with increasing distance, as energy is dissipated in the room boundaries, audience, and air.

In these large rooms all listeners beyond the front rows are in a predominantly reverberant sound field. The three dashed curves and the downward-pointing arrow illustrate that as absorption is added to a room the steady-state level of the reverberation drops. As a result, the critical distance increases (horizontal arrow). As the directivity of the sound source increases—in the direction of the listener or microphone—the critical distance also increases. In thinking about what may happen in the small rooms of interest to us, assuming no other differences, the critical distance will be larger because these rooms have proportionally more sound-absorptive material and the sound sources have significant directivity, and are aimed at the listener. As a result, we may find that we are not listening in the reverberant sound field.

1.3 Offices and Industrial Spaces

Occupying the middle ground between large, high ceilinged performance spaces and domestic rooms are those with large floor areas and lower ceilings: offices, factories, and the like. Most such spaces have significant amounts of absorption, much of it on the ceiling or floor, or both. They also have large sound-absorbing and scattering objects distributed throughout the floor area, desks, people, office cubicles, machines, production lines, and so on. If the objects in these spaces are significantly large relative to the height and volume of the rooms, they cannot be treated as a "layer" of sound-absorbing material on the floor. Sounds propagating across such spaces behave distinctively.

Different dimensional ratios, differing deployment of absorbing materials, and scattering objects, all result in different sound propagation characteristics. However, there are some strong common features. Close to the sound source, sound reflected (backscattered) from objects in the space can cause the sound level to exceed that of the direct sound, especially at high frequencies. Over much of the distance, the draw-away curve falls at a rate of approximately −3 dB per double distance (dB/dd), at least for combined middle and high frequencies. Hodgson [10] discusses several models for predicting the actual rate, which is frequency dependent. At longer distances this trend may continue, or, depending on the room geometry, the distribution of absorbing material and the presence of significantly large scattering objects, the rate of decay can accelerate [11]. Fig. 3 shows two simplified theoretical predictions for the tendencies of draw-away curves, the popular −3 dB per double distance, and a more elaborate prediction by Peutz [12], as compiled and reported by Schultz, in a very insightful document [9]. Real draw-away curves measured by Hodgson [11] in several industrial spaces exemplify both trends, with a fair amount of scatter caused by differing behavior at different frequencies.

Late reflections are attenuated rapidly with distance from the source. Over almost the entire draw-away distance, including the range of listening distances typical of small rooms, listeners are in what can best be described as a prolonged transitional sound field, neither direct nor reverberant. This means that critical distance is not an appropriate concept.

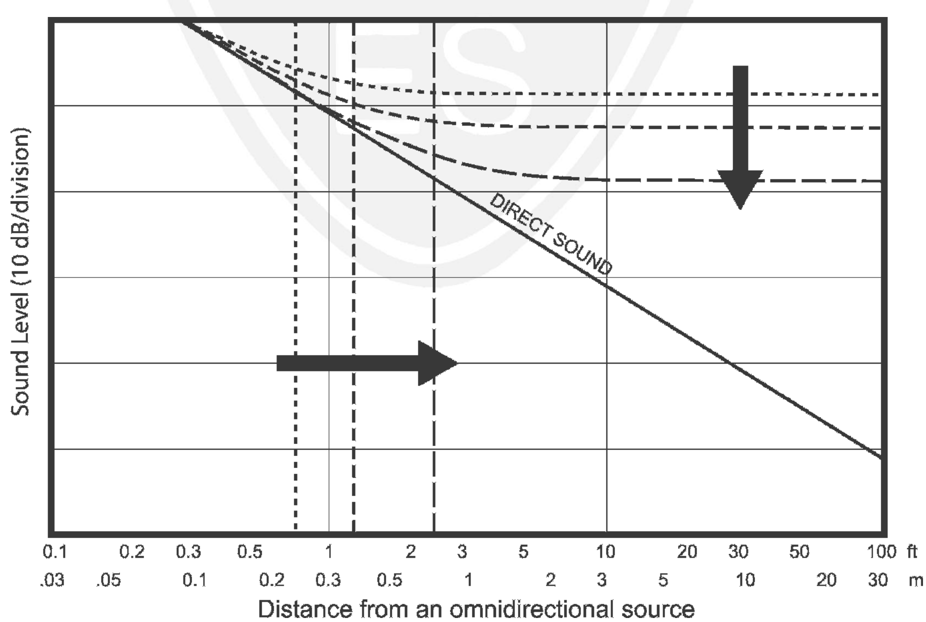

Fig. 2. Classic depiction of sound level as a function of distance from an omnidirectional sound source in a large, reverberant, irregularly shaped room such as an auditorium or concert hall (upper dashed curves). Horizontal portions of curves fall in level as absorption in room increases and reverberation time decreases. Critical distance, shown by vertical dashed lines, defines the point at which direct sound and steady-state reverberant field are equal in level. Critical distance increases with sound absorption in room [8], [9].

1.4 Domestic Listening Rooms and Control Rooms

When the floor area shrinks from office or factory to domestic dimensions there is reason to believe that the basics of this behavior will continue because key features of the commercial spaces are present. Large portions of one or more surfaces have significant absorption in the form of carpet, drapery, and, perhaps, acoustical ceilings. There are also sound-absorbing and scattering objects, such as sofas, chairs, tables, cabinets, and vertically stepped arrangements of bulky leather chairs in custom home theaters, all of which are large relative to the ceiling height in typical homes.

Schultz measured draw-away curves in several living rooms [9]. He used A-weighted measurements of broadband, omnidirectional, or at least widely dispersing, calibrated noise sources—an ILG fan, a blender, a saw, and a drill. The sound field was found to decline at a rate of approximately −3 dB per double distance. This was confirmed in draw-away measurements done by the author in two entertainment rooms using loudspeakers of various directional characteristics: omnidirectional, bipole, dipole, and forward firing. The combined data from nine sound sources in six rooms are shown in Fig. 4. The monotonic decline in sound level shown in all of the draw-away curves indicates a source-to-sink energy flow at increasing distance from the source.

Considering the distances at which we listen in our entertainment spaces and control rooms, it is clear that we are in the transitional region, where the direct and early reflected sounds dominate, and late reflected sounds are subdued, and progressively attenuated with distance. The sound field is not diffuse, and there is no critical distance, as classically defined.

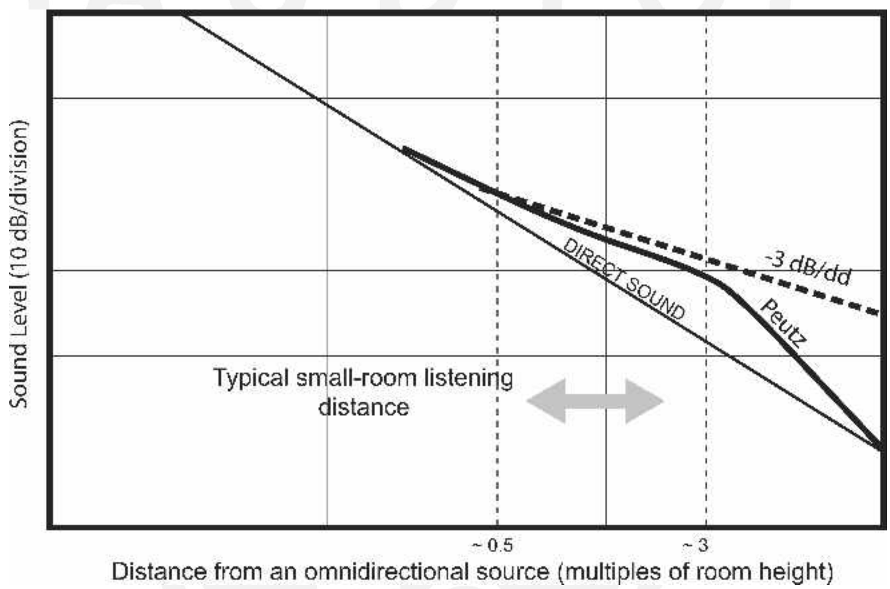

Fig. 3. Anticipated A-weighted sound level as a function of distance from an omnidirectional sound source in an office or industrial space. −3 dB/dd line—approximation of predictions according to several theories [10]; ▬—predictions according to a theory by Peutz, described in [9]. Horizontal scale is appropriate for Peutz prediction.

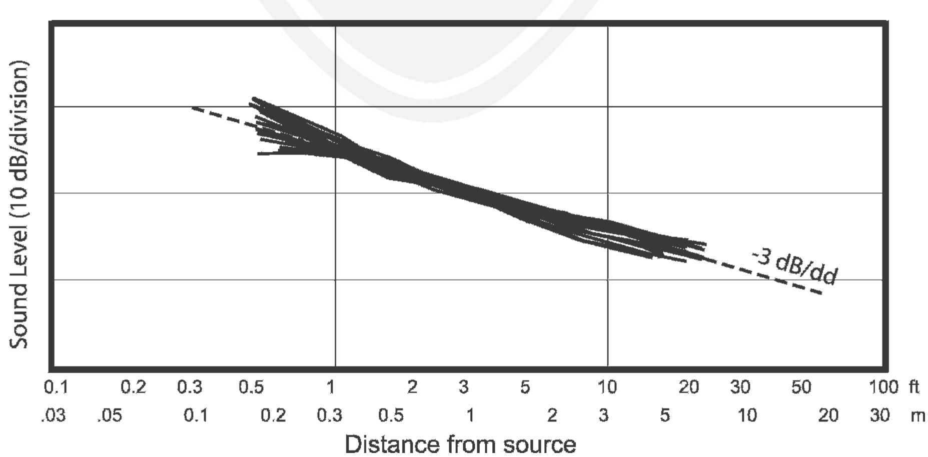

Fig. 4. A-weighted sound level as a function of distance for four widely dispersing noise sources [9] in four different living rooms, and five loudspeakers of differing directivity (omnidirectional, dipole, bipole, and two different forward firing) measured in two different listening rooms. Curves were normalized to a similar middistance sound level so as to reveal average shape and slope. Variations seen at distances less than about 1 ft are due to near-field effects of large loudspeaker sources. Those seen at maximum distances could be due to reflections from back wall of room.

1.4.1 Measuring the Lack of Diffusion

A recent paper provides hard evidence of what is going on in the sound fields in some small rooms [13]. Using a novel spherical steerable-array microphone, the authors explored, in three dimensions, the decaying sound field in several small rooms. None of them exhibited isotropic distributions at the measurement locations. Strong directional features were associated with early reflections. Small meeting rooms and a videoconferencing room with reverberation times of 0.36–0.4 s, like listening rooms, had anisotropy indices and directional diffusion measures that fell roughly halfway between anechoic and reverberant conditions. Moreover, the values changed with time, with later sound showing increased anisotropy and even changing orientation in the room according to the surfaces that were more reflective (Fig. 5).

None of this is necessarily bad. A diffuse sound field may be a worthy objective for performance spaces and recording studios, where the uniform blending of multiple sound sources and the reflected sounds from those multidirectional sources is desired. However, it is conceivable that such a sound field may not be a requirement for sound reproduction through multiple, somewhat directional loudspeakers surrounding and directed toward a listener. This becomes especially so when it is considered that, in popular applications like movie and television sound tracks and traditional music recordings, all of the loudspeakers are not allocated equivalent tasks—front loudspeakers predominantly create real and phantom “sound stage” images, while side and rear loudspeakers provide occasional directional cues, but are mainly utilized to create enveloping ambient and spatial illusions. This notion might need rethinking if “listener-in-the-middle-of-the-band” recordings become the standard.

1.5 Interim Summary of Small-Room Acoustics

1.5.1 What Is a Small Room?

Diffuse-field theory may not apply perfectly to concert halls, but it applies even less well to other kinds of rooms. In the acoustical transition from a large performance space to a “small” room, it seems that the significant factors are a reduced ceiling height (relative to length and width), significant areas of absorption on one or more of the boundary surfaces, and proportionally large absorbing and scattering objects distributed throughout the floor area. Different combinations of these characteristics result in basically similar acoustical behavior in large industrial spaces [10], [11] and, with minor adjustment, in domestic listening spaces. Sound radiating from a source is either absorbed immediately on its first encounter with a surface or object, or the objects redirect the sound into something else that absorbs it. Thus the late reflected sound field is greatly diminished with distance from the source. These are not “Sabine” spaces, and it is not appropriate to em-

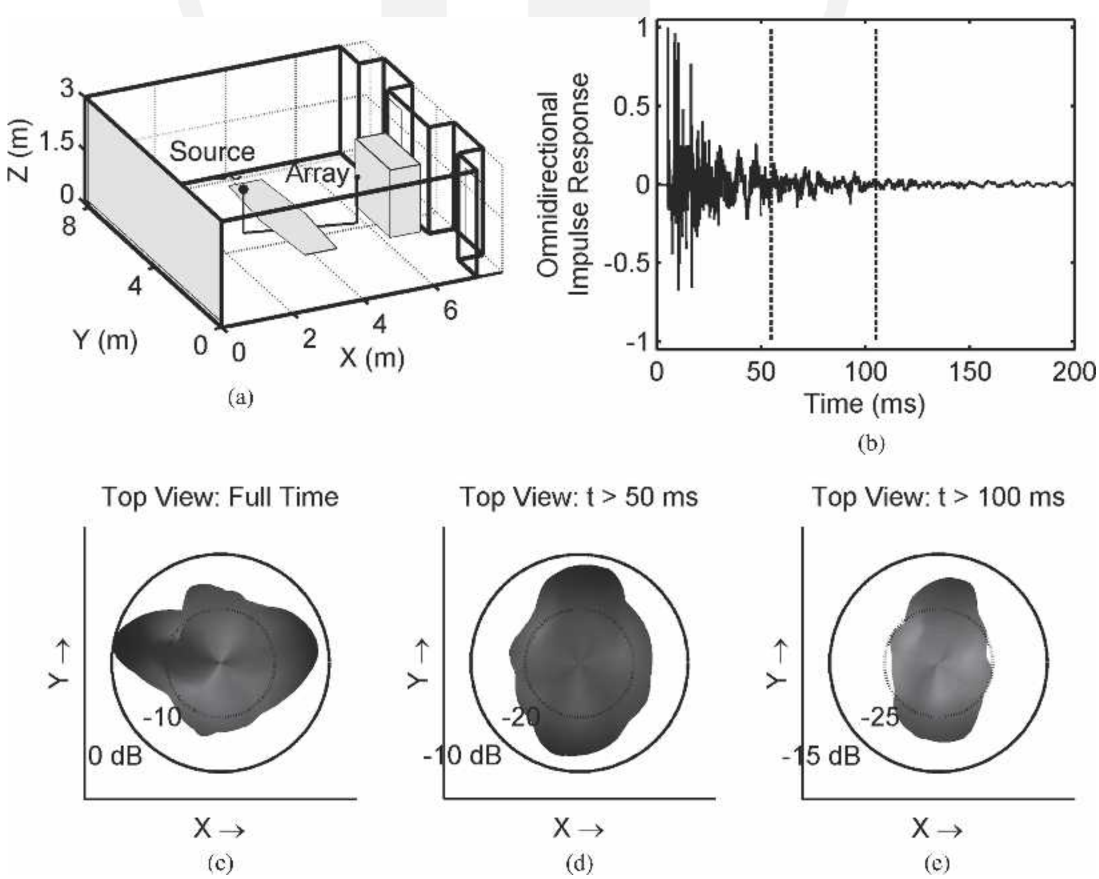

Fig. 5. Directional diffusivity in small room for different time intervals in a decaying sound field. A perfectly diffuse sound field would yield a circular pattern. In this room the field exhibits strong directivity, and that directivity changes and becomes more exaggerated with time [13].

ploy calculations and measurements that rely on assumptions of diffusivity [9].

1.5.2 Value of Conventional Acoustical Measures in Small Listening Rooms

A measurement of reverberation time (RT) in a domestic-sized room yields a number. When the number is large, the room sounds live, and when the number is small, the room sounds dead. The implication is that there should be an optimum number. In spite of this, there are serious minds saying that RT is unimportant or irrelevant [14]–[17]. The numbers measured are small compared to those in performance spaces, and so the question arises if the late reflected sound field in a listening room is capable of altering what is heard in the reproduction of music. Yet RT is routinely included as one of the measures of small listening and control rooms for international standards, even to the point of specifying allowable variations with frequency.

Reverberation time is a property of the room alone, and a correct measurement of it should employ an omnidirectional sound source capable of "illuminating" all of the room boundaries. The reason: it is assumed that the boundaries consist of areas of reflection and absorption, and that the central volume of the room is empty. The several formulas by which we estimate RT confirm this. Some practitioners incorrectly use conventional sound-reproduction loudspeakers as sources. The directivity of these is such that the resulting reflection patterns and decays are not properties of the room, but of the room and loudspeaker combination—a very different situation.

The result of a correct RT measurement is a number, or a set of numbers for different frequency bands, describing the decay rate over a range of sound levels, maybe 20 or 30 dB, and then extended by multiplication to give a number for a 60-dB decay. It is common to look at the mid-frequency reverberation time and the variations with frequency. The former is a measure of the suitability of a performance space for different styles of music. The variations with frequency are important because it is undesirable to change the spectral balance of voices and musical instruments by excessive absorption in narrow frequency bands. This is critical in large performance spaces, because almost all of the listeners are in a sound field dominated by reverberation.

In a small listening room we are in a transitional sound field, consisting of the direct sound, several strong early reflections, and a much diminished late reflected sound field. What we hear is dominated by the directional characteristics of the loudspeakers and the acoustic behavior of the room boundaries at the locations of the strong early reflections. RT reveals nothing of this. As a measure, it is not incorrect; it is just not useful as an indicator of how reproduced music or films will sound. Nevertheless, excessive reflected sound is undesirable, and an RT measurement can tell us that we are "in the ballpark," but so can our ears, or an acoustically aware visual inspection.

This transitional sound field appears to extend over the entire range of listening distances we commonly employ in small rooms. It is therefore necessary to conclude that the large-room concept of critical distance is also irrelevant in small rooms. This said, there is much anecdotal evidence of a perceptible transition occurring at some distance from loudspeakers in a room. None of this appears to have been systematically investigated in terms of examining the nature and consequences of this perceptual change. Since critical distance is not the appropriate measure, a new one is needed. A reasonable hypothesis is that it is related to the ratio of direct to early reflected sound; a topic for some useful research.

All of the other acoustical measures employed in evaluating performance spaces: early/late decay rates, energy ratios, lateral fractions, and others having to do with impressions of articulation, direction, image size, apparent source width, and spaciousness, could be applied to sounds reproduced over a multichannel reproduction system. However, in doing so, one is also evaluating the recording, and the manner in which it captured, or was processed to simulate, those attributes. That is another worthy and challenging area of investigation, and it could conceivably lead to improvements in recording technique and multiple-loudspeaker configurations. But, again, so far as the performance of the listening space itself is concerned, these are more acoustical measures that find themselves in the wrong place. The numbers produced by the measuring instruments, while not totally irrelevant, are simply not direct answers to the important questions in small-room acoustics. What, then, are the important questions? They have to do with reflections, but not in a bulk, statistical sense.

2 AUDIBLE EFFECTS OF REFLECTIONS—A SURVEY

In anticipation of experimental results showing listener reactions to reflections in a variety of circumstances, it is useful to have some perspective on what might be found. The following is a quick review of these known effects:

- Localization (direction)—the precedence effect
- Localization (distance)
- Image size and position
- Sense of space
- Timbre: comb filtering, repetition pitch
- Timbre: audibility of resonances
- Speech intelligibility.

2.1 Effects on Localization (Direction)—The Precedence Effect

Over the years the terms Haas effect and the law of the first wavefront have also been applied to this effect, but current scientific work appears to have settled on the term precedence effect. It has to do with the well-known phenomenon wherein the first arrived sound, normally the direct sound from a source, dominates our sense of where sound is coming from. Within a time interval often called the fusion zone we are not aware of reflected sounds as separate spatial events. All of the sound appears to come from the direction of the first arrival. Delayed sounds arriving outside the fusion zone may be perceived as separate auditory images, coexisting with the direct sound, but the direct sound is still perceptually dominant [6]. At long delays the secondary images are perceived as echoes, separated in time as well as direction.

It needs to be emphasized that, within the fusion interval, there is no masking—all of the reflected sounds are audible, making their contributions to timbre and loudness, but the early reflections simply are not heard as spatially separate events. They are perceived as coming from the direction of the first sound.

Recent research [6], [18]–[21] suggests that the precedence effect is cognitive, meaning that it occurs at a high level in the brain, not at a peripheral auditory level. Its purpose appears to be to allow us to localize sound sources in reflective environments where the sound field is so complicated by multiple reflections that sounds at the ears cannot be continuously relied upon for accurate directional information. This leads to the concept of "plausibility" wherein we accumulate data we can trust—such as occasional high-frequency transients or visual cues—and persist in localizing sounds to those locations at times when the auditory cues at our ears are contradictory [22].

At the onset of a sound accompanied by reflections in an unfamiliar setting, it appears that we hear everything. Then after a brief buildup interval, the precedence effect causes our attention to focus on the first arrival, and we simply are not aware of the reflections as spatially separate events. This spatial suppression of later sounds can persist for up to about 9 s, allowing the adaptation to be effective in situations where sound is not continuous.

A change in the pattern of reflections, in number, direction, timing, or spectrum, can cause the initiation of a new buildup without eliminating the old one. We seem to be able to remember several of these "scenes." All of this buildup and decay of the precedence effect needs to be considered in the design of experiments where spatial or localization effects are being investigated; namely, are the reported perceptions before or after precedence-effect buildup?

Important for localization, and very interesting from the perspective of sound reproduction, is the observation that the precedence effect appears to be most effective when the spectra of the direct and reflected sounds are similar [4], [18], [20]. This appears to be an argument for constant-directivity loudspeakers and frequency-independent (that is, broad-band) reflectors, absorbers, and diffusers.

2.2 Effects on Localization (Distance)

Our ability to judge the distance of sounds is greatly improved when there are reflections, especially, it seems, early reflections. In rooms we improve with time, meaning that we learn certain aspects of the sound field and, once learned, it transfers to different locations in the same room and, to some extent, to rooms having similar acoustical properties [23]–[29]. It is another perceptual dimension with a cognitive component. All of this is clearly relevant to localizing the real sources—the loudspeakers. However, success in doing this may run counter the objectives of music and film sound, which is often to "transport" listeners to other spaces. How do we react to a combination of real and recorded patterns of reflections?

The single-channel, hard-panned signal of a news reader is perceived as originating in the center-channel loudspeaker localized within the listening room; the loudspeaker itself is the real source of sound. If spatial cues, in the form of real or simulated reflections, are incorporated into a good multichannel recording, they can make the loudspeakers less obvious and can cause the apparent sound source to seem farther away and the room to seem larger. A psychoacoustic perspective on what is happening in these instances would be able to indicate the characteristics of both local and recorded early reflections necessary to establish dominance in our perception of distance. Hints that the perception of distance is more driven by monaural cues than binaural cues [27] are encouraging, given the limited number of channels available in our audio systems. However, if there is even an element of "plausibility" in distance perception, it may be difficult not to be influenced by walls and loudspeakers that we can see. More research is needed on this important topic.

2.3 Effects on Image Size and Position

The precedence effect will guide our attention to the first arrived sound as an indicator of the location of the source. However, early reflections can cause this directional impression to shift slightly, or the "image" of the source to be enlarged. One can logically speculate that the image shifting or spreading is associated with a reduction in interaural cross correlation (IACC), in the well-known tradeoff with spaciousness—increasing spaciousness tends to be correlated with image broadening [30]. Haas noted that adding a delayed lateral sound caused a "pleasant broadening of the primary sound source" [31]. Olive and Toole determined the delays and levels at which a single lateral reflection caused a perceptible change in the size or location of the primary image [32].

In concert-hall acoustics the nearest equivalent effect is described as apparent source width (ASW), where it is regarded as beneficial, allowing the sound of the orchestra to appear larger than the visual spread of performers. Many audiophiles appear to put value in a soundstage that extends beyond the physical span of the loudspeakers, a phenomenon that can be created through deliberate or accidental binaural effects in recordings, or by lateral reflections from adjacent walls in the reproduction space.

Whether a change in image position or size is good or bad, then, is a subjective judgment. Because what was intended by the recording artists is normally not known, such judgments must be based on what appears to be plausible or personally preferable. If a video or movie image is involved, visual cues normally provide that data and dominate the overall impression almost regardless of the sounds at the two ears.

2.4 Effects on the Sense of Space

The term spatial impression incorporates two separable perceived dimensions: ASW, described in Section 2.3, and listener envelopment (LEV), which appears to be the more important in live performances [33]. This impression, of being in a specific acoustical space, is perhaps the strongest argument for multichannel audio as an advance on stereo—the fact that the recording engineer has some degree of independent control over the senses of direction, ASW, and LEV. It will be seen that even a single reflection is sufficient to generate a rudimentary sense of spaciousness.

2.5 Effects on Timbre—Comb Filtering, Repetition Pitch

The acoustical sum of a sound and a delayed version of the same sound produces two results. First, if measured, it yields a frequency response that looks a bit like a comb, with regularly spaced alternating (constructive interference) peaks and (destructive interference) dips. Second, if listened to, we can get any of several responses, including coloration at worst and a pleasant sense of spaciousness at best. In that sense, comb filtering is something akin to a measurement artifact.

The worst situation is when the summation occurs in the electrical signal path or within the loudspeaker itself. Then the direct sound and all reflected versions of it contain the same interference pattern. Another difficult situation is one with only a single dominant reflection arriving from close to the same direction as the direct sound. In a control-room context, this could be a console reflection in an otherwise dead room.

Fortunately such events are rare. Most reflections arrive from directions different from the direct sound, and perceptions vary considerably. Two ears and a brain have advantages over a microphone and an analyzer. The fact that the perceived spectrum is the result of a central (brain) summation of the slightly different spectra at the two ears attenuates the potential coloration from lateral reflections significantly [34]. If there are many reflections, from many directions, the coloration may disappear altogether [35], a conclusion to which we can all attest through our experiences listening in the elaborate comb filters called concert halls. Blauert summarizes: "Clearly, then, the auditory system possesses the ability, in binaural hearing, to disregard certain linear distortions of the ear input signals in forming the timbre of the auditory event" [6].

Superimposed on all of this is a cognitive learning effect, a form of "spectral compensation" wherein listeners appear to be able to adapt to these situations, and to hear "through and around" reflections to perceive the true nature of the sound source [36]–[38]. Put differently, it seems humans have some ability to separate a spectrum that is changing (the program) from one that is stationary (the transmission channel/propagation path). It is evident that we do not yet have all the answers, but it seems clear that the human auditory system is well adapted to dealing with reflective listening spaces.

2.6 Effects on Timbre—Audibility of Resonances

The Toole and Olive investigations of the audibility of resonances yielded the interesting fact that repetitions of a sound lowered the detection thresholds for medium- and low-Q resonances within the sound [7]. This is a quantitative confirmation of common experience: live unamplified music sounds better in a room than it does outdoors. In addition to the obvious spatial embellishments, it sounds richer and more rewarding as we are able to hear more of the timbral subtleties.

2.7 Effects on Speech Intelligibility

A common belief in the audio industry is that reflections are detrimental to speech intelligibility. No doubt there is a risk in PA systems that can elicit long delayed reflections in large environments, but what about small rooms? It turns out that, depending on the arrival times and sound levels of delayed sounds, speech can be made more, not less, intelligible and, further, can be made more pleasant to listen to. It is an interesting story.

3 AUDIBLE EFFECTS OF A SINGLE REFLECTION—EXPERIMENTAL RESULTS WITH SPEECH

If the perception of speech is unsatisfactory, it is fair to conclude that our ability to be informed and entertained has been seriously compromised. Consequently most of the classic psychoacoustic investigations have used speech as a signal. In examining the effects, it is convenient to separate the discussion into the following categories:

1) Localization—the apparent position in azimuth, elevation, distance, in this case also considering the perceived size of the sound source

2) "Disturbance" of speech by a reflection—the worst case

3) Intelligibility of speech as it is influenced by a reflection—the essential factor

4) Effect of a reflection on the preferred sound quality of speech—adding an aesthetic touch.

3.1 Localization

Fig. 6 shows a series of transitions between audible effects when a single lateral reflection is added to a direct sound (at 0°) in an anechoic environment. All of these curves were measured using speech as a signal.

The lowest curve describes the sound level at which listeners reported hearing any change attributable to the presence of the reflection. This is the absolute threshold since nothing is perceived for reflections at lower levels. Most listeners described what they heard as a sense of spaciousness [32]. (It should be noted that direct sound and a reflection arriving from the same vertical, medial, plane yielded impressions of timbre change, not spaciousness.) Throughout listeners reported all of the sound as originating at the location of the loudspeaker reproducing the first sound; the precedence effect was working.

The next higher curve is the level at which listeners reported hearing a change in size or position of the principal sound image, which the precedence effect causes to be localized at the position of the loudspeaker reproducing the earlier sound. This was called the image-shift threshold [32]. In general these changes were subtle, noticeable in A versus B comparisons, but in the context of a multiple-image soundstage one needs to know whether these playback effects are likely to be recognized as being distinct from those comparable effects that are created (accidentally or deliberately) during the multichannel recording process. An aesthetic subjective judgment of this attribute is especially difficult to put a value on, when what was intended by the recording engineer and artists is not known. In programs such as movies or television, where most of the information is presented through the center channel, it could be argued that some image spreading—a "softening" of the image—might be beneficial.

With the two curves that portray the third perceptual category a major transition is reached, because it is at this sound level that listeners report hearing a second sound source or image, simultaneously coexisting with the original one (at long delays there is a sense of a temporally as well as spatially separate echo) [39], [40]. This means that the precedence effect directional "fusion" has broken down. Although the original source remains the perceptually louder, spatially dominant source, there is a problem because two spatial events are perceived when only one should be heard.

The top curve is from the well-known work by Haas [31], in which he asked his listeners to adjust the relative levels of the spatially separate images associated with the direct and reflected sounds until they appeared to be equally loud. This tells us that, in a public address situation, it is possible to raise the level of delayed sound from

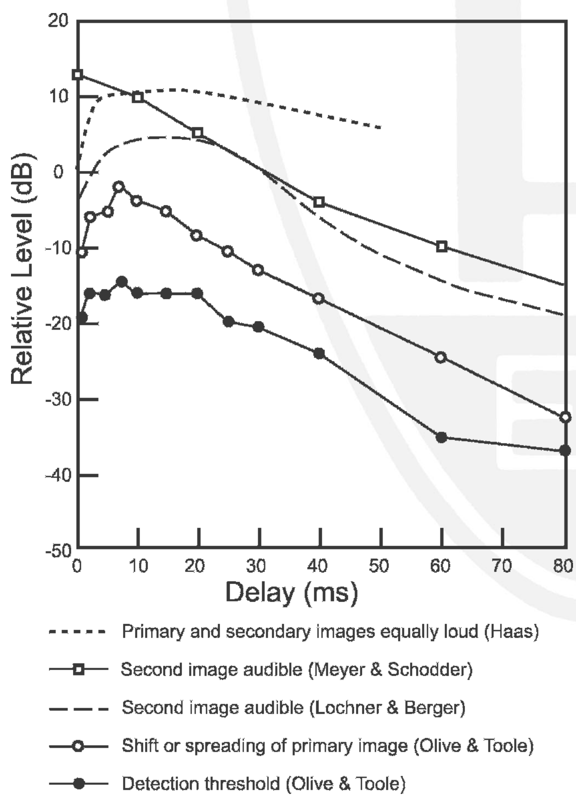

Fig. 6. Progression of thresholds for different audible effects of a single lateral reflection, as heard in an anechoic space. Bottom curve—detection threshold [32]. At this level the common sensation was one of spaciousness; all sound appeared to come from the location of the first arrived sound. Second curve from bottom: sound level at which listeners heard a change in the physical spread or position of the principal sound image [32]. Next two curves—levels at which listeners were able to identify a second image, at the location of the reflection, coexisting with the principal image [39], [40]. Top curve—level at which primary and secondary images were judged to be equal in loudness [31].

a laterally positioned loudspeaker by as much as 10 dB above the direct sound before it is perceived as being as loud as the direct sound. It is important information in the context of professional audio, but it is irrelevant in the context of small-room acoustics. Unfortunately the audio engineering literature has several reinterpretations of this result, including the notion that there is masking or fusion in more respects than just localization, within an interval of about 20–30 ms, which has been called the Haas zone, Haas effect fusion zone, and other. This is not so. It is evident from an inspection of Fig. 6 that all of the curves are continuous. (There is no range of delays or time zone wherein certain things happen that do not happen elsewhere.) It is also evident that audible effects begin to be perceived at levels 30 dB or more below the level of the curve that Haas generated, meaning that masking in the simple sense of rendering other sounds inaudible is not evident. Haas himself noted that as the delayed sound was increased in level, there were perceptible changes in loudness, sound quality, liveliness, body, and, as mentioned earlier, a "pleasant broadening of the primary sound source." Superimposed on all of this is the fact that the Haas data are frequently interpreted as if they applied to music. The threshold curves for other kinds of sounds, especially transient sounds, are very different, as will be seen. Whatever conclusions may be drawn from Haas' work, they will apply to speech, and some, but not all, musical sounds.

In 1985 Benade published the following: ". . . there is an accumulation of information from the various members of the sequence [of reflections following a direct sound]. It is quite incorrect to assume that the precedence effect is some sort of masking phenomenon which, by blocking out the later arrivals of the signal, prevents the auditory system from being confused. Quite to the contrary, those arrivals that come in within a reasonable time after the first one actively contribute to our knowledge of the source. Furthermore members of the set that are delayed somewhat too long actually disrupt and confuse our perceptions even when they may not be consciously recognized. If the arrivals are later yet, they are heard as separate events (echoes) and are treated as a nuisance. In neither case are the late arrivals masked out" [4]. It is enlightening to read Benade's full description of what he calls the generalized precedence effect with an appreciation of when it was written.

If we take from Fig. 6 those things that are relevant to the unamplified reflected sounds that occur in sound reproduction from a single loudspeaker, we end up with Fig. 7, in which the Haas data have been removed and the "second-image" data have been combined into a single average curve. The shaded area under that curve can be considered to be the real-world precedence effect fusion zone for speech, within which any reflected sound will not be perceived as a spatially separate localizable event. The much quoted 30 ms (±) as the fusion interval for speech clearly applies only if the delayed sound has the same level as the direct sound. This is how the classic psychoacoustic experiments were conducted, but it is improbable in normal rooms. For reflections at realistically lower levels the fusion interval is much longer.

Thus we arrive at a partial answer to the localization issue. Individual reflections in normal small rooms are not likely to generate multiple images from speech produced by a person or reproduced by a loudspeaker. (The directivity of a human speaker is within the range of directivities for conventional cone/dome loudspeakers [41].) A single lateral reflection may cause the sound image to be slightly larger or slightly displaced from the position it would have in an otherwise anechoic space, but it remains to be seen if this is noticeable in a multichannel recording. Some evidence suggests that even these small effects might be diminished by experience during listening within a given room [42]. So far, in small rooms the precedence effect appears to be in control of the localization of speech.

3.2 Disturbance of Speech by Reflections

Fig. 8 shows data from two studies of the disturbing effects of delayed sound on speech [43], [44]. While this is an issue in large venues, it is evident that natural reflections in small rooms are too low in amplitude and delay to be problems in this respect.

3.3 Effects of Single Reflections on Intelligibility

It has long been recognized that early reflections improve speech intelligibility, so long as they arrive within the "integration interval" for speech, about 30 ms [45]. More recent investigations found that intelligibility improves progressively as the delay of a single reflection is reduced, although the subjective effect is less than would be predicted by a perfect energy summation of direct and reflected sounds [46]. Fig. 9 shows the effect of a single reflection arriving from different directions on the intelligibility of speech [47]. In this study the reflection was at the same sound level as the direct sound, which makes this a worst-case test. The fact that it was done in a quiet anechoic chamber means that the signal-to-noise ratio was not an issue. Within the time interval in which strong early reflections are likely to occur in listening and control rooms (about 15 ms) the effects on intelligibility are negligible for the most likely lateral reflections (30–60°), and really only noticeable for those arriving from 0° (ceiling and floor reflections). A survey of typical listening rooms [48] showed that the average floor bounce occurs at 1.7 ms and is attenuated by 1.5 dB.

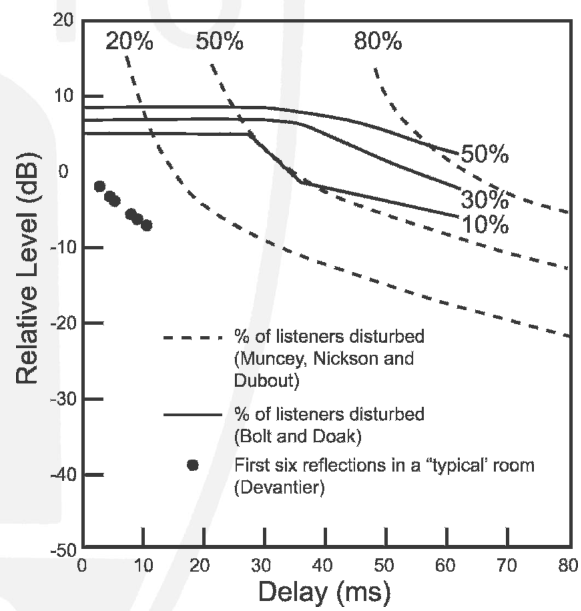

Fig. 8. Levels and delays at which a single reflection causes listeners to be disturbed while listening to speech, according to [44], [43]. Shown for comparison are the first six reflections in a typical listening room [48].

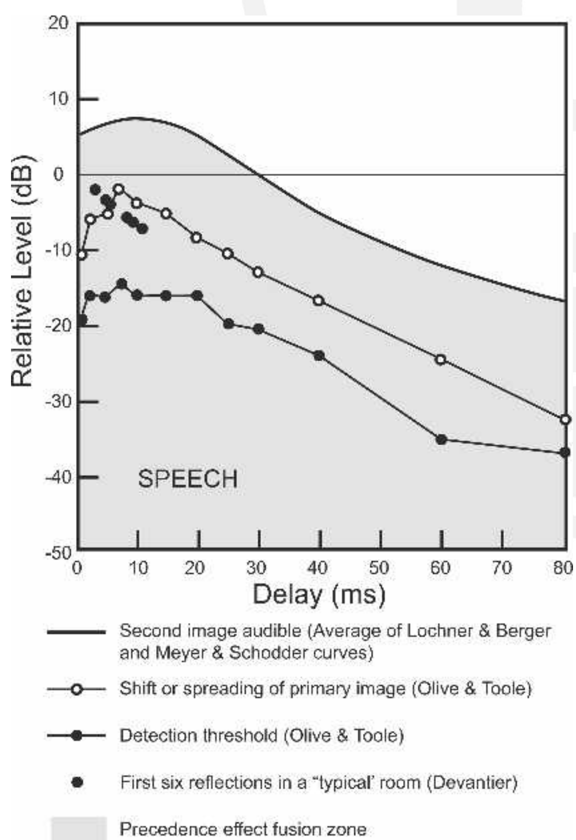

Fig. 7. Data from Fig. 6 relevant to sound reproduction in small rooms, with curves of [39] and [40] averaged. Shaded area—precedence effect fusion zone, defining levels and delays of reflections not perceived as second sound images. Also shown are the first six reflections in a "typical" listening room [48], indicating that all are well within the fusion zone.

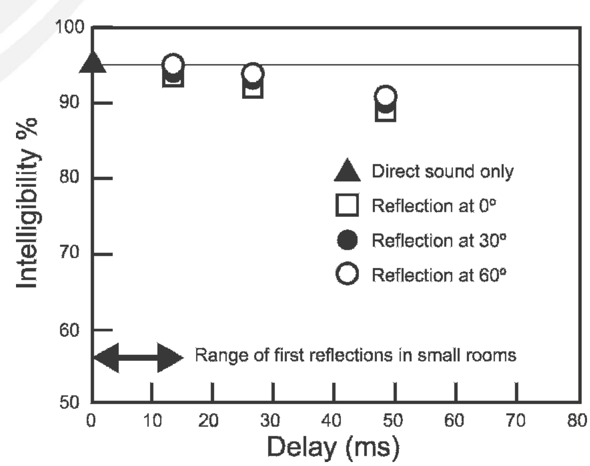

Fig. 9. Speech intelligibility measured for direct sound only at 0°, and with single reflections added, arriving from different horizontal angles relative to direct sound [47]. Reflections and direct sound at the same levels (anechoic listening conditions).

The average ceiling bounce occurs at 4.9 ms, attenuated by 3.6 dB. Looking at the overall evidence from these studies it seems clear that, in listening rooms, some individual reflections have a negligible effect on speech intelligibility, and others improve it, with the improvement increasing as the delay is reduced.

3.4 Effects of Single Reflections on Listener Preferences of Sound Quality

It is accepted that a reflective sound field is flattering to the sound of music. We like to listen in a reverberant space, rather than outdoors. The question is at what point does this positive attribute begin? Ando has provided some answers. Fig. 10 shows levels for a single delayed sound that listeners reported as enhancing the sound of classical music. Since the early reflections in real rooms are so low in level, the result suggests that we really need multichannel audio to provide added stronger and later reflections for our listening pleasure [49]. Fig. 11 shows comparable results for speech, and, again, the natural reflections fall short of preferred levels [50], [51]. Note that the second-image curve is just avoided, indicating that the "preferred" reflections were all within the precedence effect fusion zone. The inevitable conclusion is that, in natural listening, room reflections are not problems. In fact, they will need augmentation by recorded reflections before listeners are fully gratified.

3.4.1 Detailed Look at what Contributes to Preference in This Context

Reflections have a positive contribution to listener preferences, but some reflections are more desirable than others. It comes down to which of them are most effective at generating a sense of spaciousness which, in turn, follows from low interaural cross correlation. The greater the differences in sounds at the two ears, the greater the sense of spaciousness, and the higher the listener preference ratings. Fig. 12 shows that a single reflection arriving from the front is least preferred, whereas those from about 40–90° to the side are more desirable [50]. This is in the region of the first sidewall reflection from the front loudspeakers, and includes side-located surround loudspeakers in a multichannel playback system. Front–back symmetry in binaural hearing implies that these desirable effects will exist for sounds arriving at angles of 40–140°. Overall, sounds arriving from the sides, with substantial front and back tolerances, are beneficial. In concert halls the importance of these sounds is indicated by a measure of the lateral energy fraction LF, which correlates well with a sense of envelopment. In both concert halls and listening rooms reflected sounds arriving from the front or rear do not contribute to a positive impression. Fig. 12 shows diminished preference for reflected sounds arriving within about 20–30° of the median plane.

Looking into the time domain, Fig. 13 shows that lateral reflections arriving with delays greater than about 2–3 ms are effective at reducing the interaural cross correlation and therefore contribute to increased preference [50]. Beyond about 4–10 ms there is little change. In practical audio systems this suggests that reflecting surfaces very close to a loudspeaker are least beneficial. Of course, a lack of positive contribution is not necessarily a negative contribution, since many room reflections and many recorded reflections coexist in real multichannel listening situations. Yet another topic for investigation and, clearly, program is a variable.

4 EFFECT OF MULTIPLE REFLECTIONS—MOVING CLOSER TO REALITY

Working with a single reflection allows for intensely analytical investigations but, inevitably, the tests must include others in order to be realistic.

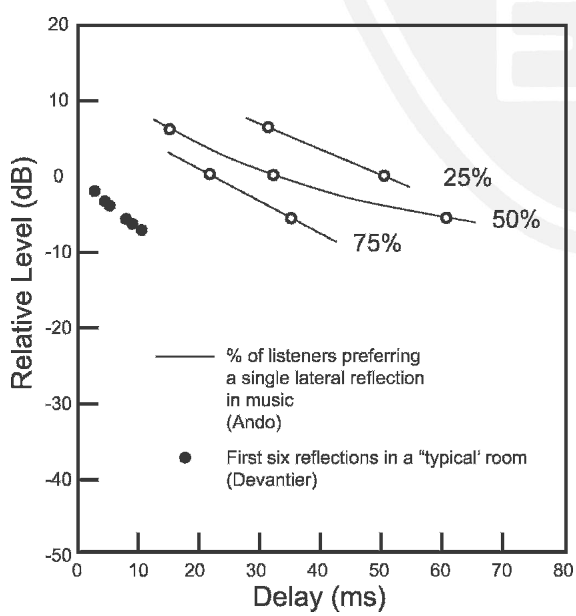

Fig. 10. Percentage of listeners who reported a preference for music in the presence of a single lateral reflection at 30° from direct sound [49]. Shown for comparison are the first six reflections in a typical listening room [48].

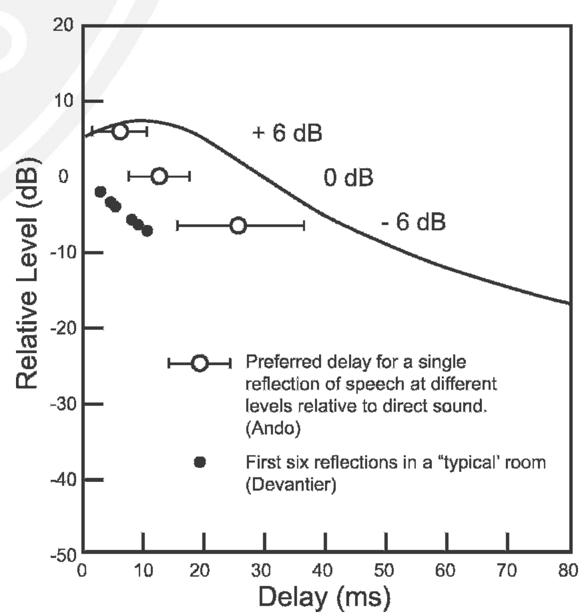

Fig. 11. With speech, preferred delay for a single reflection at a horizontal angle of 36°. Results are shown for three sound levels relative to direct sound. (Adapted from [50] Fig. 7 and [51] Fig. 42.)

4.1 A Reflection among Other Reflections

A long-standing belief, in the area of control-room design, is that early reflections from monitor loudspeakers must be attenuated in order to allow those in the recordings to be audible. Consequently embodied in several standards, and published designs, are schemes to deflect, diffuse, or absorb at least the first reflections from a loudspeaker.

Olive and Toole appear to have been the first to test this idea [32]. In a continuation of the experiments discussed earlier (Section 3.1, Figs. 6 and 7), which examined the audibility of a single lateral reflection in an anechoic chamber, other experiments were conducted using the same physical arrangement, first in a typical small room (IEC 268-13, 1985), in which first reflections had been attenuated, and second in the same room with most absorption removed (RT = 0.4 s) see Fig. 14. The large changes in the level of reflected sound had only a modest (1–5-dB) effect on the absolute threshold or the image-shift threshold of an additional lateral reflection occurring within about 30 ms of the direct sound. At longer delays the threshold shifts were up to about 20 dB, a clear response to higher level late-reflected sounds in the increasingly live rooms.

In a large anechoic-chamber simulation of a room of similar size, Bech investigated the audibility of single reflections in the presence of 16 other reflections, plus a simulated "reverberant" sound field beginning at 22 ms [52]. One of his results is directly comparable with these data. The figure caption in Bech's paper describes the response criterion as "a change in spatial aspects," which seems to match the image-shift or image-spreading criterion used by Olive and Toole. Fig. 15 shows the image-shift thresholds in the real IEC room for two subjects (the FT data are from [32], the SO data were previously unpublished), and the thresholds determined in the simulated room, an average of the three listeners from Bech's work [52]. The similarity of the results is remarkable, considering the very different physical circumstances of the tests. It suggests that listeners were responding to the same audible effect, and that the real and simulated rooms had similar acoustical properties.

The basic audible effects of early reflections in recordings, therefore, seem to be remarkably well preserved in the reflective sound fields of ordinary rooms. There may be reasons to attenuate early reflections within listening rooms, but this, it seems, is not one.

4.2 Multiple Reflections

Following the pattern set by studies involving single reflections, Lochner and Burger [45], Soulodre et al. [46], and Bradley et al. [53] found that multiple reflections also contribute to improved speech intelligibility. The most

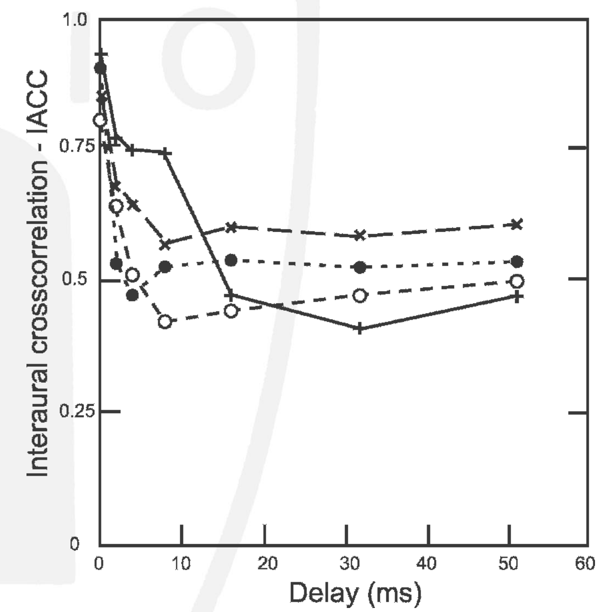

Fig. 13. Interaural cross correlation as a function of delay for single reflection at a horizontal angle of 36°, four musical motifs [50].

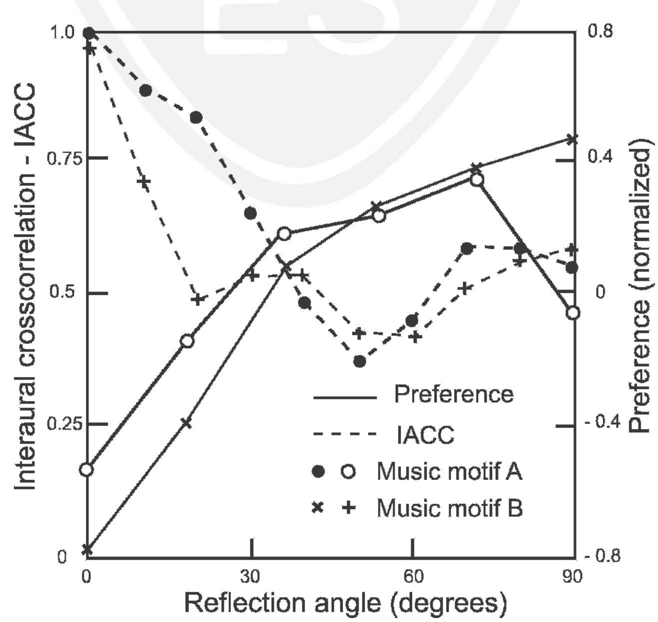

Fig. 12. Preference scores for two musical motifs as a function of horizontal angle of single reflection, with corresponding measured interaural cross correlation (IACC) [50].

elaborate of these experiments used an array of eight loudspeakers in an anechoic chamber to simulate early reflections and a reverberant decay for several different rooms [53]. The smallest was similar in size to a very large home theater or a screening room (390 m^3, 13 773 ft^3). The result was that early reflections (<50 ms) had the same desirable effect on speech intelligibility as increasing the level of the direct sound. The authors go on to point out that late reflections (including reverberation) are undesirable, but controlling them should not be the first priority, which is to maximize the total energy in the direct and early reflected speech sounds. Remarkably, even attenuating the direct sound had little effect on intelligibility in a sound field with sufficient early reflections. The findings were confirmed in subsequent tests employing "listening difficulty" ratings, which turn out to be more sensitive in indicating problems than conventional intelligibility or word recognition scores [54].

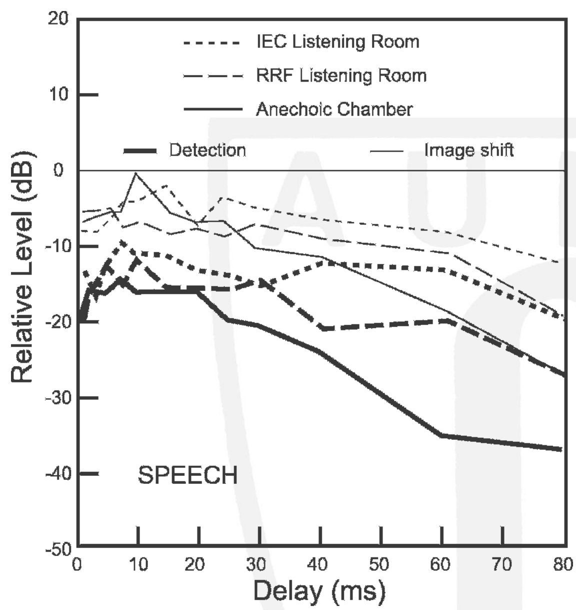

Fig. 14. Detection and image-shift thresholds as a function of delay for single lateral reflection ▬ anechoic chamber; ––– listening room in which all first reflections have been attenuated with 2-inch (50-mm) fiberglass board (RRF listening room); ---- same IEC listening room in highly reflective configuration [32].

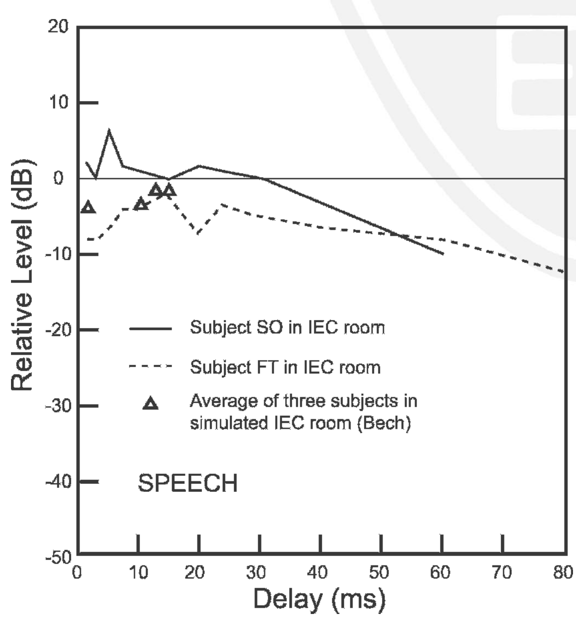

Fig. 15. Image-shift thresholds as a function of delay for single lateral reflection for two listeners in reflective IEC listening room (FT data from [32]) and for three listeners (averaged) in simulation of an IEC room using multiple loudspeakers in large anechoic chamber [52].

5 EFFECTS OF REFLECTIONS—A SUMMARY

5.1 Speech

Readers who have been keeping score will have noted a distinct absence of negative effects from reflections on any aspect of speech perception we have looked at. In fact, the effects range from neutral to positive. No single reflection has been shown to be a problem for speech reproduction in small rooms (see Table 1). Multiple early reflections contribute even more to intelligibility.

5.2 Other Sounds

Fig. 16 shows how different the shapes of the threshold curves can be for different kinds of sounds. The differen-

Table 1. Reflections in small rooms and perception of speech: report card.

	Problems	No Problems	Advantages
Localization	☐	☑	☐
Annoyance	☐	☑	☐
Intelligibility	☐	☐	☑
Preference	☐	☐	☑

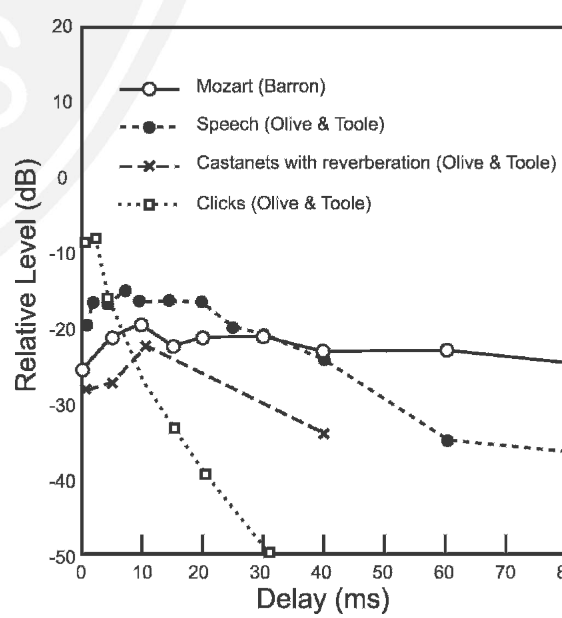

Fig. 16. Absolute detection thresholds for single lateral reflection in anechoic chamber for sounds of four different temporal structures: relatively continuous (Mozart), mixture of transient and continuous (speech), transients with reverberation (castanets in room), and isolated transients (electrical clicks) [32].

tiating element seems to be whether the sound is continuous, contains repetitions, or is purely transient in nature. In terms of the precedence effect fusion interval, it is clear that for continuous sound (pink noise) or highly sustained sounds (Mozart) the fusion interval is very long. At the other extreme, for isolated transient sounds it can be only a few milliseconds. Mixtures of transient and sustained sounds, such as speech or castanets with reverberation, are in between [32].

In technical terms, the degree of continuity might be more aptly described as the autocorrelation interval (colloquially described as the effective duration of the sound), a measure used by Ando in some of his investigations [50], [51]. Some casual tests suggest that the shapes of the threshold curves are continued in such families of curves as image shift or second image, as shown in Fig. 6. However, only a direct investigation of this would provide proof.

5.3 Sound Reproduction Presents More Opportunities for Research

All of the analyses of rooms thus far have focused on the effects that rooms have on sounds created within them, without confronting the essential factor unique to sound reproduction. It is necessary to distinguish between the factors related to the perceptions of timbre, direction, distance, and space generated by real sources of sound in listening rooms (such as the loudspeakers), and those associated with the timbre and (usually very different) illusions of direction, distance, and space that multichannel recordings attempt to create in those spaces through those same loudspeakers.

Individual reflections, in general, appear to be flattering to both music and speech sounds, and those occurring naturally in small rooms are, if anything, too low in level to have an optimal effect. This could be interpreted as providing justification for multichannel audio, to add more reflections, and implying that natural room reflections are not likely to detract from reproduced sounds incorporating reflections.

Indeed, numerous early reflections have a positive cumulative effect on speech intelligibility. In terms of image localization, the precedence effect appears to hold in the presence of many reflections, and our distance perception improves. Distortions of image size and position appear to be borderline issues. From the perspective of sound quality, multiple reflections reduce the effects of comb filtering (good) and enhance our perception of resonances (good for the music, and bad only if the resonances are in the loudspeakers). If managed properly, room reflections could contribute to our impressions of spaciousness. The audibility of an individual early reflection in a recording is not greatly influenced by collections of early reflections in a room, but we have no data on whether that recorded reflection, and others that accompany it, are perceived as contributing to the impressions of distance, for example, intended by the recording.

Data on these and other aspects of perception are incomplete. We need to know the circumstances under which desirable effects begin to diminish and neutral influences become audible degradations. At what point does our ability to judge the location and distance of the loudspeakers inhibit our ability to be persuaded that we are in the environment incorporated into a recording?

In a television news broadcast in which voices are discrete center-channel signals we may invoke plausibility to modify the precedence effect and localize to the moving lips on the screen, sometimes referred to as the ventriloquism effect. We may use naturally occurring early room reflections to judge the distance of the loudspeaker as corresponding closely to that of the screen. All is well. However, if that same center channel is reproducing a singer on a distant stage, or the sound of a car approaching from down a street, only recorded reflections can provide those distance cues. By what rules and measures is it possible to achieve both of these conflicting illusions? We know from experience that it can work, but could it work better, and more often?

In the end, we need new measures. One of these would be a replacement for the large-room measure, critical distance. In small listening rooms it seems to be important to have guidance about the optimal amplitude or energy relationships between direct and (collections of) early reflected sounds. In multichannel systems it is reasonable to think that this may differ by channel, depending on the primary function, such as front channels versus surround channels, and the application, movies versus music.

6 ADAPTATION

We humans adapt to the world around us in many, if not all, dimensions of perception—temperature, luminance level, ambient smells, colors, and sounds. When we take photographs under fluorescent or incandescent lighting or outdoors in the shade or direct sunshine, we immediately are aware of color balance shifts—greenish, orangish, bluish, and so on. Yet in daily life we automatically adjust for these and see each other and the things around us as if under constant illumination. We adapt to low and high light levels without thinking. There are limits—very colored lighting gets our attention, we cannot look into the sun, or see in the dark, but over a range of typical circumstances we do remarkably well at maintaining a comfortable normalcy. Most adaptation occurs on a moment-by-moment basis, and is a matter of comfort—bringing our perception of the environment to a more acceptable condition. In the extreme, adaptation, habituation, or acclimatization, whatever we call it, can be a matter of survival and a factor in evolution.

Earlier, in the contexts of precedence effect (angular localization), distance perception, and spectral compensation (timbre), it was stated that humans can track complex reflective patterns in rooms and adjust their processes to compensate for much that might otherwise be disruptive in our perceptions of where sounds come from, and of the true timbral signature of sound sources. This appears to be achieved at the cognitive level of perception—the result of data acquisition, processing, and decision making, involving notions of what is or is not plausible.

Some evidence of this, in the context of loudspeaker sound quality, comes from experiments in which three loudspeakers were evaluated subjectively in four different rooms [55]. In the first experiment listeners completed the evaluation of the three loudspeakers in one room before moving to the next one. Binaural recordings were made of each loudspeaker in each location in each room, and the tests were repeated, but this time with listeners hearing all of the sounds through headphones. All tests were double blind. In each room, three loudspeakers were evaluated in three locations for each of three programs. The whole process was repeated, resulting in 54 ratings for each of the 20 listeners. The result from a statistical perspective was that:

- "Loudspeaker" was highly significant: $p \leq 0.05$
- "Room" was not a significant factor
- Results of live and binaural tests were essentially the same.

A possible interpretation is that the listeners became familiar with—adapted to—the room they were in and, this done, were able to judge the relative merits of the loudspeakers accurately. Since they were given the opportunity to adapt to each of the four rooms, they were able to arrive at four very similar ratings of the relative qualities of the loudspeakers.

Then, using the same binaural recordings that so faithfully replicated the results of the live listening tests, another experiment was conducted. In this, the loudspeakers were compared with themselves and each other when located in each of the loudspeaker positions in each of the four rooms. Thus in this experiment the sound of the room was combined with the sound of the loudspeakers in randomized presentations that did not permit listeners to adapt. The result was that "room" became the highly significant variable ($p \leq 0.001$) and "loudspeaker" was not significant.

It appears therefore that we can acclimatize to our listening environment to such an extent that we are able to listen through it to appreciate qualities intrinsic to the sound sources themselves. It is as if we can separate the sound of a spectrum that is changing (the sounds from the different loudspeakers) from that which is fixed (the colorations added by the room itself for the specific listener and loudspeaker locations within it). This appears to be related to the spectral compensation effect noted by Watkins [36]–[38].

It is well not to overdramatize these results because, while the overall results were as stated, it does not mean that there were *no* interactions between individual loudspeakers and individual rooms. There were, and almost all seemed to be related to low-frequency performance. This thought-provoking observation has powerful implications on how we conduct listening tests, especially when direct comparisons are involved—are we listening to differences before or after adaptation? Are the differences being revealed in the listening test the same qualities of sound that would be perceived by us in a normal listening situation? Here, it seems, is another area ripe for research.

It seems safe to take away from this a message that listeners in comparative evaluations of loudspeakers in a listening room are able to "neutralize" audible effects of the room to a considerable extent. If residual effects of the room are predominantly at low frequencies, these differences, and also those in the reflected sound field, can be physically neutralized by employing a positional "shuffler" to bring active loudspeakers to the same location in the room [56].

There are everyday parallels to this. We carry on conversations in a vast range of acoustical environments, from cavelike to the near anechoic, and while we are certainly aware of the changes in acoustical ambience, the intrinsic timbral signatures of our voices remain amazingly stable. The excellence of tone in a fine musical instrument is recognizable in many different, including unfamiliar environments. Benade sums up the situation: "The physicist says that the signal path in a music room is the cause of great confusion, whereas the musician and his audience find that without the room, only music of the most elementary sort is possible! Clearly we have a paradox to resolve as we look for the features of the musical sound that give it sufficient robustness to survive its strenuous voyage to its listeners, and as we seek the features of the transmission process itself that permit a cleverly designed auditory system to deduce the nature of the source that produced the original sound" [57].

So we humans manage to compensate for many of the temporal and timbral variations contributed by rooms, and hear "through" them to appreciate certain essential qualities of sound sources within these spaces. Because adaptation takes time, even a little, there is the caveat to acousticians not to pay too much attention to what they hear while moving around—stop, or sit down and listen.

With this in mind, the concept of room correction becomes moot; how much and what really needs to change, and how much can the normal perceptual process accommodate? What do we have the option of changing, and what should we simply leave alone?

In spite of the incomplete state of this area of work, there remains one compelling result: when given a chance to compare, listeners sat down in four different rooms and reliably rated three loudspeakers in terms of sound quality. Now we need to understand what it is about those loudspeakers that caused some to be preferred to others. If that is possible, it suggests that by building those properties into a loudspeaker, one may have ensured that it will sound good in a wide variety of rooms; a dream come true.

7 ONE ROOM, TWO SOUND FIELDS

Fig. 17(a) shows frequency responses measured at the listening location for a loudspeaker in three different positions in a small room [58]. Below about 300 Hz the frequency response is dominated by loudspeaker position, whereas at higher frequencies the measurements follow a similar pattern. The acoustical explanation is the dominance of room modes and standing waves at low frequencies and of a complex and relatively stable array of reflected sounds at high frequencies. In between is a transition zone, the middle of which, in large rooms such as concert halls and auditoriums, would be defined as the

Schroeder frequency [59]. Calculation of the Schroeder frequency assumes meaningful reverberation times and a strongly diffuse sound field. As we know, in small rooms these are mismatched concepts, so the calculations are almost always in error [60]. However, the transition region is real, and it is necessary to take different approaches to dealing with acoustical phenomena above and below it.

As rooms get larger, the transition frequency drops. In concert halls it is at the bottom of the useful frequency range, and low-frequency room resonances cease to be problems. As rooms shrink, the transition frequency rises, explaining why, in cars, the cabin is the dominant factor over much of the frequency range.

7.1 Above the Transition Frequency

Above the transition frequency we hear a combination of the direct sound and many reflections. Steady-state measurements in a room are not definitive because the physical interaction of sounds at a microphone is very different from the perceptions arising in two ears and a brain. Visually alarming irregularities in measurements are frequently not heard (see Section 2.5). This is the frequency region within which all of the previous discussions of reflections apply, and where the bulk of adaptation, precedence effect, and the like occurs. The idea expressed in Section 6—that our judgments of sound quality are somewhat independent of the complications of room reflections—is challenging. We will look closely at this, because the consequences are profound.

Let us begin by attempting to understand the nature of the sounds arriving at the ears of listeners in rooms. Fig. 17(b) shows a synthesis of sounds arriving at a listening position in a room, calculated from many anechoic measurements made on horizontal and vertical orbits around a loudspeaker [58]. This loudspeaker had good on-axis behavior, but deteriorated off axis. It did not have constant, or even smoothly changing directivity. The result is that energy in the early and late reflected sounds exhibited undulating frequency responses. The energy sum of all three curves yields an estimate of what might be measured in a small room. Above the transition frequency it turns out to be an excellent fit to the curves shown in Fig. 17(a). From anechoic measurements on a loudspeaker, a room curve has been predicted.

The room used here was the prototype IEC 268-13 (1985) listening room, which in essential respects was intended to be representative of typical domestic listening spaces. It had carpet on the floor, drapes, furniture, bookcases, tables, and so on. Since this example, numerous other measurements have been made with different loud-

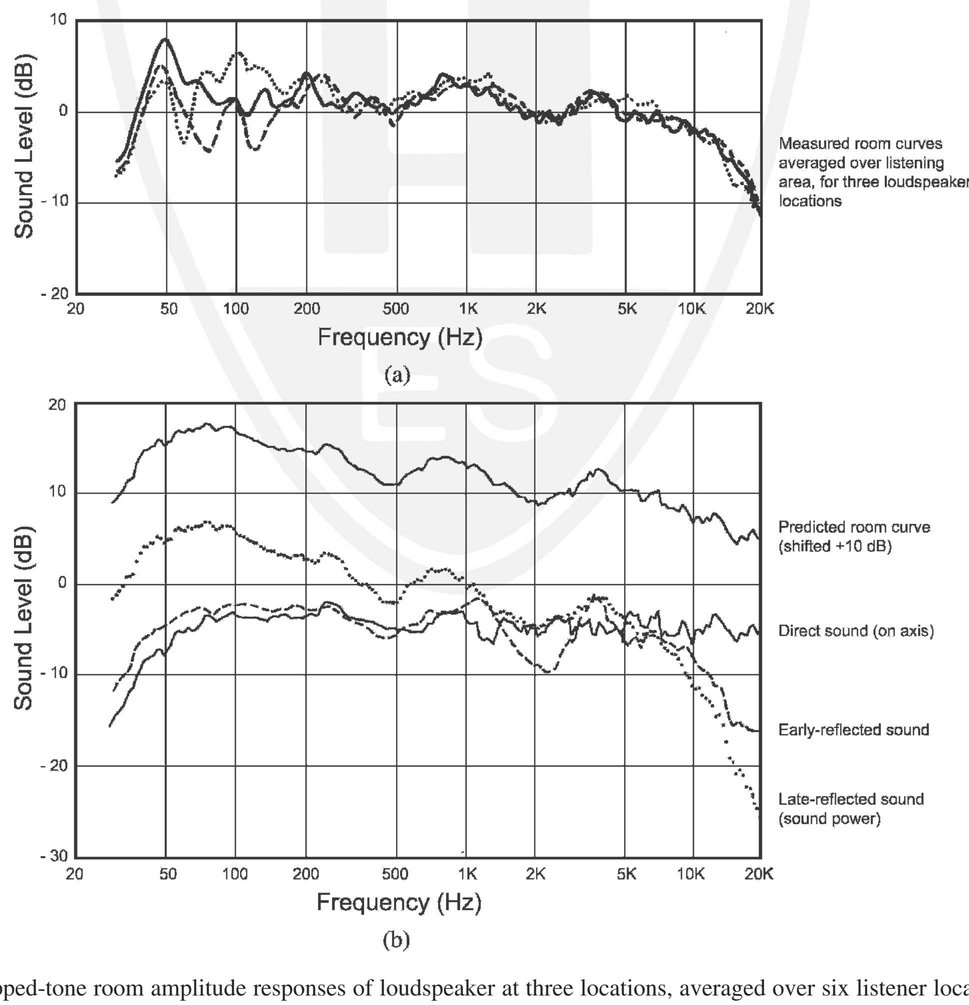

Fig. 17. (a) Stepped-tone room amplitude responses of loudspeaker at three locations, averaged over six listener locations (no spectral smoothing) in prototype IEC 268-13 (1985) room. (b) Estimates of direct, early reflected, and late reflected sounds at one loudspeaker–listener location used in (a), calculated from anechoic measurements on loudspeaker using geometry and acoustical properties of this specific room. Top curve—energy summation of lower three curves shifted +10 dB for clarity. No spectral smoothing [58].

speakers in different normal living spaces. The predictions have been very close approximations to the real measurements [48]. While interesting, this can only be of long-term value if it correlates with subjective evaluations. Because a room curve is a nonanalytical combination of all sounds from all directions, at all times, it would be surprising if it held the secrets to good sound.

From anechoic measurements on a loudspeaker system it is possible to anticipate the direct, early reflected, and late reflected sounds that arrive at a listening position in a typical small room. If those measured sounds are the ones that the ear/brain system relies on to form a timbral identity, then it should be possible to employ that collection of anechoic data in a model to predict subjective impressions of sound quality—a sound quality, or preference, rating. This, of course, is the ultimate goal of psychoacoustic research, to be able to anticipate a subjective response from quantitative data. In the context of loudspeaker sound quality, it is something that few have attempted.

7.1.1 Correlations between Subjective and Objective Domains

For over 20 years the author and his colleagues have conducted comprehensive anechoic measurements on many loudspeakers, and have examined the results of double-blind listening tests performed on these products. The results have been gratifyingly similar: loudspeakers exhibiting certain generally recognizable measured characteristics consistently achieved high scores in subjective evaluations [58]. As subjectively interpreted, a smooth, flat, wide-band axial frequency response, combined with similarly well-behaved off-axis responses, up to and including sound power, appeared to be the desirable pattern. What was missing was a mathematical process by which the technical data could be converted into a figure of merit, an estimate of a subjective rating of sound quality. In 2004 Olive took up the challenge and, using subjective and objective data from 70 loudspeakers, developed a model to perform the conversion [61], [62].

Earlier attempts had been based on basic measurements such as sound power or room curves, usually with restricted (such as one-third octave) frequency resolution. Using more recent psychoacoustic knowledge, the new models examined much smaller details in the raw measurements and, having access to more information, they could apply different weightings to the direct, early reflected, and late reflected sounds. When all of the new perspectives were included in the analysis, the result was a correlation of 0.86 between the calculated subjective rating and the real subjective rating. So it seems that we truly are measuring quantities that are important to our subjective tastes. It is not an accident.

As impressive as this is, it should be noted that there was a significant source of variation in the subjective data. The 70 loudspeakers were evaluated in 19 different listening tests, conducted over a period of many months. In each test, only three or four products were compared, so, inevitably, there was a certain amount of drift and elasticity in the subjective scales used by the listeners. In other words, depending on what other products it is being compared to, and how long it has been since it had previously been auditioned, the rating of any individual loudspeaker could move up or down the rating scale. The movement is usually not large, but it is a change that the statistical analysis regards as uncertainty about the rating, reducing the correlation.

To overcome this, all loudspeakers must be evaluated in one continuous test, with each product being compared to every other product. When this was done with a group of 13 bookshelf loudspeakers, the correlation improved to 0.995—near perfection. The fact that the loudspeakers being compared were of similar physical configuration was an advantage, but that does not detract from the importance of the result. It is clear that there is a way to translate anechoic data from loudspeakers into very reasonable predictions of subjective ratings as they occur in a normal listening room.

And there is more. The excellent correlations mentioned came from a model that had access to a complete library of anechoic data—70 individual high-resolution frequency response curves at different angles surrounding the loudspeaker. With less data the correlations were less good. High-resolution data (1/20 octave) were consistently better than one-third-octave data. No single curve, anechoic or in room, alone was adequate, although the axial response figured prominently in all of the successful models, perhaps because it is the event that triggers perceptual processes like the precedence effect, and how one perceives later arrivals. Early in this paper it was noted that reverberation is not a dominant factor in what we hear in small rooms, and here it is no surprise to find that the sound power output from a loudspeaker is, alone, an imperfect predictor of sound quality, especially when, as is commonly done, it is one-third-octave filtered.

7.1.2 Attenuating, Reflecting, and Scattering Indirect Sounds

Although reflections appear not to be great problems, it is reasonable to think that there must be a level above which the good attributes are diminished and negative attributes grow. Obviously an empty room is not a comfortable listening environment, even for conversation. The furnishings and paraphernalia of life tend to bring normal living spaces into familiar acoustical territory. Custom listening spaces need to be treated. In all rooms absorption, scattering or diffusion, and reflection occur, and devices to encourage each are commonly used by acousticians.

It appears that much of what we perceive in terms of sound quality can be predicted by the anechoic characterization of loudspeakers. Because most of these data pertain to sounds that reach listeners by indirect paths, it is proper to suggest that nothing in those indirect sound paths should alter the spectral balance. For example, a 1-inch (25.4-mm) layer of fiberglass board at the point of a strong first reflection is effective at removing sound energy above about 1 kHz. From the perspective of the loudspeaker, the off-axis response of the tweeter has just been greatly attenuated—it will sound duller and less good. Obviously if the purpose of the absorbing material is to attenuate the reflection, the material should be equally effective at all frequencies. Given the duplex nature of sound fields in small rooms, it seems reasonable to expect similar performance at all frequencies above the transition region.

In their examination of the audibility of reflections, Olive and Toole looked at detection thresholds as high frequencies were progressively eliminated from the reflected sounds, as they might be by frequency-selective absorbers. They found that only small to moderate threshold elevations occurred for low-pass filter cutoff frequencies down to about 500 Hz, where the investigation ended. Removing the high frequencies alone is not sufficient to prevent audible effects [32].

Finally there are the indications that the precedence effect is maximally effective when the spectra of the direct and reflected sounds are similar [4], [18], [20]. If the spectrum of a reflection is different from that of the direct sound, the probability that it will be heard as a separate spatial event is increased—not a good thing.

Thus from the perspectives of maintaining the excellence in sound quality of good loudspeakers, rendering an unwanted reflection inaudible, and preserving the effectiveness of the precedence effect, there are reasons not to alter the spectrum of reflected sounds. One is free to redirect them with reflectors or diffusers, or to absorb them with lossy acoustical devices, but in each case, the process should not alter the spectrum of the sound above some frequency toward the lower side of the transition region in a small room. It seems reasonable to propose, therefore, that all acoustical devices used in listening rooms—reflectors, diffusers, and absorbers—should be uniformly effective above about 200 Hz. For resistive absorbers this means thicknesses of 3 inches (76 mm) or more.

7.2 Below the Transition Frequency

Among the factors contributing to positive subjective ratings it is hard to ignore the fact that about 30% of the overall rating is contributed by the low-frequency performance of the loudspeaker [62]. All of the listening tests in the Olive study were done in the same room, which was equipped for positional substitution of the loudspeakers [56], and where listeners had ample time to adapt to its personality. All of this helps to neutralize the room as a factor in the evaluations. To achieve comparably good subjective ratings in different rooms, it may be necessary to find a way to ensure the delivery of similarly good bass to all listeners in all rooms. Of course, achieving such consistency is a desirable objective for the entire audio industry, professional and consumer. Let us see how far this idea can be taken.

At frequencies below the transition zone, investigating what is heard must involve measurements made within the room, with the loudspeaker at its intended location and measuring at the intended listening location. All else is of academic interest. Also because of standing waves, different listening locations will experience different low-frequency responses. At very low frequencies, wavelengths are long compared to room dimensions and periods are long compared to transit times within it, so what is heard is like listening through a complex filter. At subwoofer frequencies, at least, the behavior of room resonances is essentially minimum phase [63]–[65], especially for those with amplitudes rising above the average spectrum level. This suggests that what we hear can substantially be predicted by steady-state frequency-response measurements. It also means that both time- and frequency-domain correction is possible with minimum-phase parametric filters.

7.2.1 Room Modes, Room Dimensions, Ideal Rooms, and So On

All rooms exhibit resonant modes. Even nonrectangular rooms have modes, but they are difficult to predict. So most acousticians tend to prefer working with rectangular spaces. We can change the frequencies of modes by adjusting the room dimensions, and alter the frequency distribution of modes by adjusting the room proportions: length to width to height. A lot of effort was put into finding optimum dimensional ratios for reverberation chambers, where the sound power output of mechanical devices was measured and it was important to have a uniform distribution of the resonance frequencies.

These concepts migrated into the audio field, and certain room dimensional ratios have been promoted as having desirable characteristics for listening. In normal rooms the benefits apply only to low frequencies. Bolt, who is well known for his “blob”—a graphical outline identifying recommended room ratios—makes this clear in the accompanying, but rarely seen, “range of validity” graph. This shows that in an 85-m^3 (3000-ft^3) room the optimum ratios are effective from about 40 to 120 Hz [66]. This merges nicely with the common experience that above the low-bass region the regularity of standing-wave patterns is upset by furniture, openings, and protrusions in the wall surface so that predictions of standing-wave activity outside the bass region are unreliable. In fact, even within the low-bass region wall flexure can introduce phase shifts in reflected sound sufficient to make the “acoustic” dimension at a modal frequency substantially different from the physical dimension.

But there is a practical problem with the concept of “optimum” room dimensions. It is that, to experience the benefits, all of the modes must be excited simultaneously (sound source located at a three-boundary corner) and, of course, the listener must be able to hear all of the modes (head in an opposite three-boundary corner). This is simply ridiculous. Any departure from this loudspeaker location means that all of the modes are not equally energized, and any departure from this listening position means that all of the modes are not equally audible. The tidy predictions come to nothing. Multiple loudspeakers are a further unanticipated complication. So it is not that the idea of optimum room ratios is wrong, it is simply that, as originally conceived, it is irrelevant in our business of sound reproduction.

With modifications the idea can be made to work. However, to do so is not simple because it is necessary to take into account how many loudspeakers there are, where they are, how many listeners there are, and where they are seated.

In stereo it was common to think single-mindedly of a sweet spot, and to arrange for everything to be optimum for a single listener. At low frequencies an equalizer can be used to reduce the audible excesses of objectionable room resonances, thus delivering respectable bass to a single listener. However, the existence of the standing waves between and among the room boundaries ensures that other seats experience different bass.

Delivering similarly good bass to several listeners simultaneously means that the energy in room resonances must be physically attenuated, reducing the point-to-point variations in sound pressure. Conventional acoustics attacks the problem with absorption, damping the resonances by draining energy from the offending modes, which will result in reduced pressure maxima and elevated minima. Low-frequency absorption is difficult, usually requiring bulky, expensive, and unattractive devices, most of which are hostile to even progressive concepts of interior decor. Custom installations can employ wall constructions that absorb low frequencies invisibly, and at low cost [67].

However, there are two other methods to reduce the effects of modes.

- Locate the subwoofer at or near a pressure minimum in the offending standing wave. A subwoofer is a "pressure" source, and it will couple inefficiently when located at a pressure minimum (velocity maximum).
- Use two subwoofers to cancel the standing wave. This takes advantage of the fact that lobes of a standing wave on opposite sides of a null have opposing polarity; as the sound pressure is rising on one side, it is falling on the other. Two subwoofers connected in parallel, one on each side of the null, will destructively drive the mode, reducing its effects.

Examples of these solutions can be found in [68] (see Figs. 14–16). As elegant as these methods are, they address only a single or at most a small number of modes, and they require a level of analysis that is beyond most lay persons. We need general solutions.

7.2.2 Sound-Field Management—Taking Control of the Room at Low Frequencies

The first approach to a general solution is to restrict the application to simple rectangular spaces and then, within those spaces, to identify which arrangements of some number of subwoofers result in the least seat-to-seat variation at low frequencies. Fig. 18 shows some of the preferred subwoofer arrangements resulting from Welti's investigation [69]. All of the best arrangements have even numbers of subwoofers. In order to maximize mode cancellation and to exercise control in two dimensions, groups of four subwoofers are superior. There appears to be no advantage to using more than four.

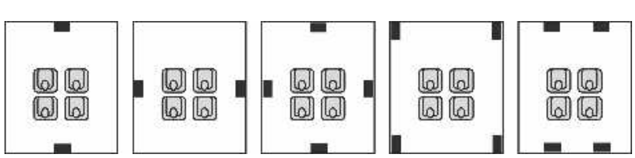

Fig. 18. Some subwoofer arrangements that yielded improved uniformity in low-frequency amplitude response (reduced seat-to-seat variation) over listening areas of small rectangular rooms [69].

The next step is to expand the solution to include non-rectangular and asymmetrical rooms. For this, measurements are needed, as well as signal processing in the feeds to the subwoofers. The results can be most gratifying [70]. Fig. 19 shows results obtained recently in the listening room of an audio reviewer. Now he can share a good bass experience with at least four other listeners. Note that when the effects of room modes have been attenuated, the need for aggressive global equalization is removed. Some rooms need none.

Even though some loudspeaker arrangements are better than others in rectangular rooms, the best performance of all is achieved when the room dimensions are optimized for a given arrangement. The most recent extension of this work completes the exercise by providing guidance to finding optimum room ratios for various arrangements of multiple subwoofers and listeners [71]. It turns out that, for some arrangements, the choices of dimensions are very generous, permitting even the much-scorned square room.

Are there disadvantages to any of this? Nothing serious, it seems. As with any subwoofer system, the low-pass filtering must be such that the sound output is attenuated rapidly above the crossover frequency (80 Hz). Excessive output, distortion products, or noises at higher frequencies increase the risk that listeners will localize the subwoofers.

A second issue relates to the fact that in order for these systems to function fully, the bass must be monophonic below the crossover frequency. Most of the bass in common program material is highly correlated or monophonic to begin with and bass-management systems are commonplace, but some have argued that it is necessary to preserve at least two-channel playback down to some very low frequency. Experimental evidence thus far has not been encouraging to supporters of this notion (see [72] and references therein). Audible differences appear to be near or below the threshold of detection, even when experienced listeners are exposed to isolated low-frequency sounds. Another recent investigation concludes that the audible effects benefiting from channel separation relate to frequencies above about 80 Hz [73]. (In their conclusion, the authors identify a "cutoff-frequency boundary between 50 Hz and 63 Hz," these being the center frequencies of the octave bands of noise used as signals. However, when the upper frequency limits of the bands are taken into account, the numbers change to about 71 and 89 Hz, the average of which is 80 Hz.) More investigations would be needed to evaluate the relative merits of good sounding bass in several seats versus impressions of space in those recordings with "stereo bass" in those seats in which it might be audible.

8 WITHIN THE TRANSITION REGION

It is convenient to speak of a transition frequency, but the reality is that there is a gradual transition through a range of frequencies, where influences of resonances and discrete reflections combine in differing proportions [Fig. 17(a)]. It is in this region that the problem of adjacent-boundary interactions commonly arises. This occurs when the loudspeakers are less than a wavelength from one or more room boundaries. Then, depending on the distance from each boundary, a systematic acoustic interference causes fluctuations in the sound power radiated into the room [74], [75]. The effects can be seen as an underlying variation in the frequency response measured in the room, modulating the amplitude of both room modes and reflections. The phenomenon can be seen by performing a spatial average of frequency-response curves measured at several locations over the listening area or by special near-field measures of the sound power radiated by the loudspeaker [76]. In either case the remedy can be to minimize the fluctuations by an appropriate placement of the loudspeaker relative to the boundaries [75] or by equalization [76]. The seat-to-seat variations addressed in Section 7 remain an issue within this frequency range.

9 DISCUSSION AND SUMMARY

The ultimate application of knowledge of this kind is in providing guidance in the design of loudspeakers and listening spaces for sound reproduction in professional (the monitoring of sound recordings) or consumer (recreational listening) domains.

Rationally, professional circumstances need to be more severely regulated than those used for entertainment. However, evidence from recording industry standards and publications by acoustical practitioners over the years indicates that the professionals are as arbitrary in some of their decisions as are consumer-audio enthusiasts. Fortunately music has survived. Perhaps it is adaptation at work.

If we are looking for hard recommendations, it is evident that all of the necessary facts are not yet available. New room acoustics measures need to be defined and the tolerances specified. There are questions that will exercise researchers for years to come. However, there is useful

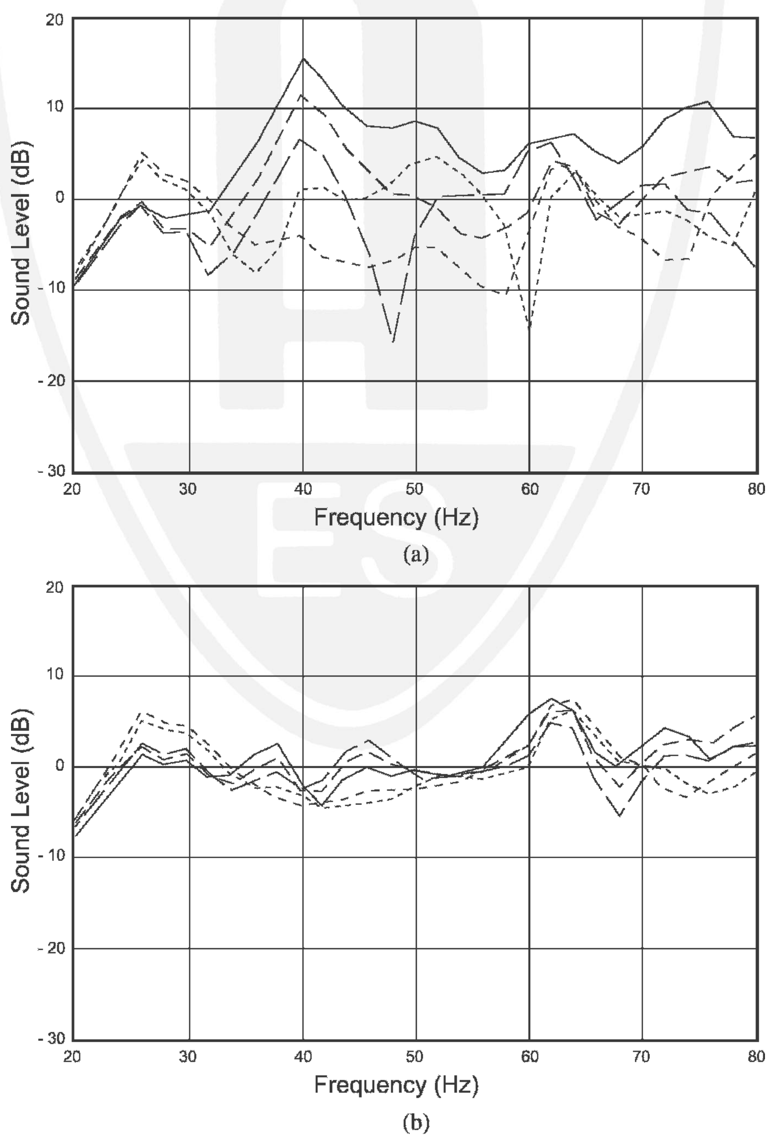

Fig. 19. (a) Measurements of low-frequency amplitude response at five listening locations in nonrectangular domestic room for single subwoofer located in front-left corner. (b) Measurements at same listening locations for four subwoofers, one in each corner, optimized and signal processed as described in [70]. No global equalization.

guidance from the research that has been done, indications that we can relax in certain quarters and focus our efforts on matters of greater perceptual importance. Here is a simplistic summary of where we stand.

9.1 Diffuse-Field Concepts

- A "small" room in acoustical terms can have a large floor area. It is one with significant absorbing material on the room boundaries, and sound-absorbing and scattering objects (such as furniture) within it having dimensions that are significant fractions of the ceiling height. Thus the sound from a source may be absorbed on the first encounter or it may be redirected, perhaps many times, eventually into absorption, so that late reflected energy diminishes with distance from the source. Because humans provide the scaling reference in determining the size of furnishings, "small" rooms are likely to have low ceilings, certainly compared to auditoriums.
- Small rooms are not Sabine spaces, so concepts developed in large performance spaces, especially those with high ceilings, apply imperfectly or not at all.
- Reverberation time is a minor factor in normally furnished small rooms so long as it is not excessive. As a measurement it is not wrong, it just does not reveal what we need to know. It is a descriptor of the room as a whole when we need to have information about how the loudspeaker(s) interact with certain features of the room.
- Critical distance is not a useful concept in small rooms. A new measure related to the relative strengths of direct and early reflected sounds is needed.

9.2 The Sound Field

We lack a method of easily identifying and calculating the transition/critical/Schroeder frequency in small rooms. Nevertheless there is such a transition, and we must deal with acoustical events above and below it in very different ways.

9.2.1 Above the Transition Frequency

- Persuasive evidence points to several beneficial and few negative effects of early reflections. However, sound reproduction brings some conflicting requirements, and more research is required to identify what control of overall reflections is appropriate. That research should take into account the normal multichannel loudspeaker configurations and the primary roles played by each of the channels.
- A room with abundant reflections is not likely to exhibit audible evidence of comb filtering from any single reflection.
- Multiple reflections improve the audibility of timbral cues from resonances in the structure of musical and vocal sounds.
- Early reflections improve speech intelligibility.
- Early lateral reflections increase our preference for the sound of music and speech. Individual reflections in small rooms may be too low in level to have the optimum effect, thus providing opportunities for multichannel sound.
- Since low interaural cross correlation is related to listener preference in certain circumstances, it is possible that asymmetrical diffusion, favoring reflections along the lateral axis, may be a good thing in listening rooms for movies and traditional styles of music recordings.
- Reflections from central portions of the front and back walls have the least positive contributions to what we hear. Attenuating them may be advantageous.
- Comprehensive high-resolution anechoic frequency-response data on loudspeakers contain sufficient information to permit remarkably good predictions of subjective preference ratings based on listening in a normal room. Single measures, such as the on-axis frequency response, sound-power response or steady-state in-room curves, are less reliable.
- Steady-state in-room measurements may be indicative of certain problems that are audible, but they are of little use in assigning corrective measures. One cannot separate problems in loudspeakers from problems in rooms, and each requires different solutions. For example, a dip in a room curve could be caused by destructive interference from a strong reflection or standing wave, a dip in the frequency response of the loudspeaker, or a reduction in the dispersion of the loudspeaker. Some of these problems require acoustical or electroacoustical treatment, and others can be corrected by equalization. Equalization schemes based only on room curves involve a risk that the wrong corrective measure will be applied to a problem.
- Any device inserted into a reflected sound path—reflector, absorber, or diffuser—should perform uniformly well at all frequencies above the transition frequency region, say, 200–300 Hz. This is in order to preserve the spectral balance of the loudspeakers, to uniformly attenuate the full spectrum of reflections, and to ensure that the precedence effect is maximally effective.

9.2.2 Below the Transition Frequency

- The modal misbehavior of rooms can be treated by passive or active acoustical methods. It is a problem over which we have considerable control.
- This is very good news, since about 30% of our subjective assessment of overall sound quality is associated with bass performance.
- Optimum room dimensional ratios exist, but only if the loudspeaker and listener locations are known in advance. Generic "good" listening room ratios are a myth.
- Multiple subwoofers, with or without active signal processing, provide options for achieving more uniformly good bass at several listening locations in small rooms. The need for equalization is reduced.
- Equalization is the final touch, and, properly done, it works because low-frequency room resonances behave as minimum-phase systems.
- The numbers and positions of low-frequency drivers required for optimal low-frequency performance may not

be compatible with the locations of the multiple loudspeakers required for optimum directional and spatial effects in surround-sound systems. Therefore it seems that full-range loudspeakers in multichannel systems may not be capable of delivering the best possible overall sound. Subwoofer–satellite systems that had their origins in low-cost systems may, in an evolved form, be the optimum configuration.

9.2.3 Within the Transition Frequency Region

Fluctuations in the sound power output of the loudspeaker(s) caused by adjacent boundary interference can be measured as a spatial average of frequency response over the listening area, and the effect minimized by positional adjustments, or the appropriate correction can be applied by equalization.

9.3 Adaptation

We adapt to several aspects of the rooms we listen in, allowing us to hear through them to identify sound qualities intrinsic to the source itself, and to identify the correct direction and distance of the source in spite of a massively complicated sound field. We need to have measures of the limits of this adaptation, at what points and in what ways our perceptual processes can use some help. The following are a few salient points to ponder.

- Voices, musical instruments, and other sounds are instantly recognizable in many rooms and through seriously flawed communication channels. We seem to be able to separate a spectrum that is changing from one that is fixed. What range of spectral variation can we adapt to, and at what level, deviation, and so on, is it necessary to intervene manually?
- Once we adapt to the room, subtle differences in quality among a group of loudspeakers are recognizable, and the distinctions are retained when the comparison is done in other rooms.
- Some of the things we hear while moving around in a room may drift into inaudibility when we sit down.
- The fact that we can accurately judge the distance of the loudspeakers in a room seems as though it should be a detriment to creating illusions of great distance in the playback of recordings. What are the rules relating distance perception for loudspeakers within a room to that for recorded sources with their own sets of multichannel reflections? Are there features of loudspeaker performance or room acoustic treatment that could make this better or worse?
- This is a topic area ripe for research. Adaptation influences almost everything of value in what we hear, whether we are in a professional or a recreational listening situation or in a subjective scientific experiment. In the latter instance, adaptation can be a great advantage or a problem. To be sure, it cannot be ignored.

10 ACKNOWLEDGMENT

This paper is the compilation of data from many sources and the result of discussions with several persons who volunteered their time to review and comment on the manuscript. In addition to the Harman International Corporate R&D Group, Sean Olive, Allan Devantier, Todd Welti, and Don Keele, the author is grateful to Richard Small, John Bradley, Gilbert Soulodre, Marshall Buck, and Brad Gover for their insights.

11 REFERENCES

[1] L. Beranek, *Concert and Opera Halls, How They Sound* (Acoustical Society of America, New York, 1996).

[2] M. Forsyth, *Buildings for Music* (MIT Press, Cambridge, MA, 1985).

[3] C. L. S. Gilford, "The Acoustic Design of Talks Studios and Listening Rooms," *Proc. IEE,* vol. 106, pp. 245–258 (1959); reprinted in *J. Audio Eng. Soc.,* vol. 27, pp. 17–31 (1979 Jan./Feb.).

[4] A. H. Benade, "From Instrument to Ear in a Room: Direct or via Recording," *J. Audio Eng. Soc.,* vol. 33, pp. 218–233 (1985 Apr.).

[5] J. Meyer, "The Sound of the Orchestra," *J. Audio Eng. Soc.,* vol. 41, pp. 203–213 (1993 Apr.).

[6] J. Blauert, *Spatial Hearing—The Psychophysics of Human Sound Localization,* rev. ed. (MIT Press, Cambridge, MA, 1997).

[7] F. E. Toole and S. E. Olive, "The Modification of Timbre by Resonances: Perception and Measurement," *J. Audio Eng. Soc.,* vol. 36, pp. 122–142 (1988 Mar.).

[8] L. Beranek, *Acoustics* (Acoustical Society of America, New York, 1986).

[9] T. J. Schultz, "Improved Relationship between Sound Power Level and Sound Pressure Level in Domestic and Office Spaces," Rep. 5290, American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), prepared by Bolt, Beranek and Newman, Inc. (1983).

[10] M. Hodgson, "Experimental Evaluation of Simplified Models for Predicting Noise Levels in Industrial Workrooms," *J. Acoust. Soc. Am.,* vol. 103, pp. 1933–1939 (1998).

[11] M. Hodgson, "Measurements of the Influence of Fittings and Roof Pitch on the Sound Field in Panel-Roof Factories," *Appli. Acoust.,* vol. 16, pp. 369–391 (1983).

[12] V. M. A. Peutz, "The Sound Energy Density Distribution in a Room," presented at the 6th International Congress on Acoustics (Tokyo, Japan, 1968), *paper E*-5-2.

[13] B. N. Gover, J. G. Ryan, and M. R. Stinson, "Measurements of Directional Properties of Reverberant Sound Fields in Rooms Using a Spherical Microphone Array," *J. Acoust. Soc. Am.,* vol. 116, pp. 2138–2148 (2004).

[14] P. D'Antonio and D. Eger, "T_{60}—How Do I Measure Thee, Let Me Count the Ways," presented at the 81st Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 34, p. 1022 (1986 Dec.), preprint 2368.

[15] H. Kuttruff, "Sound Fields in Small Rooms," presented at the AES 15th Conference (1998).

[16] E. R. Geddes, *Premium Home Theater: Design and Construction* (GedLee LLC, Novi, MI, 2002), www.gedlee.com.

[17] D. Jones, "A Review of the Pertinent Measurements and Equations for Small Room Acoustics," *J. Acoust. Soc. Am.,* vol. 113, p. 2273 (Abstract), (2003). Personal communication; presentation text from author.

[18] R. Y. Litovsky, H. S. Colburn, W. A. Yost, and S. J. Guzman, "The Precedence Effect," *J. Acoust. Soc. Am.,* vol. 106, pp. 1633–1654 (1999).

[19] R. Y. Litovsky and B. G. Shinn-Cunningham, "Investigation of the Relationship among Three Common Measures of Precedence: Fusion, Localization Dominance, and Discrimination Suppression," *J. Acoust. Soc. Am.,* vol. 109, pp. 346–358 (2001).

[20] J. Blauert and P. L. Divenyi, "Spectral Selectivity in Binaural Contralateral Inhibition," *Acustica,* vol. 66, pp. 267–274 (1988).

[21] T. Djelani and J. Blauert, "Investigations into the Build-up and Breakdown of the Precedence Effect," *Acta Acustica—Acustica,* vol. 87, pp. 253–261 (2001).

[22] B. Rakerd and W. M. Hartmann, "Localization of Sound in Rooms, II: The Effects of a Single Reflecting Surface," *J. Acoust. Soc. Am.,* vol. 78, pp. 524–533 (1985).

[23] S. H. Nielsen, "Auditory Distance Perception in Different Rooms," *J. Audio Eng. Soc.,* vol. 41, pp. 755–770 (1993 Oct.).

[24] A. W. Bronkhorst and T. Houtgast, "Auditory Distance Perception in Rooms," *Nature,* vol. 397, pp. 517–520 (1999).

[25] P. Zahorik, "Assessing Auditory Distance Perception Using Virtual Acoustics," *J. Acoust. Soc. Am.,* vol. 111, pp. 1832–1846 (2002).

[26] R. S. Pellegrini, "Perception-Based Design of Virtual Rooms for Sound Reproduction," presented at the AES 22nd International Conference, Copenhagen, Denmark (2002).

[27] B. G. Shinn-Cunningham, "Localizing Sound in Rooms," in *Proc. ACM SIGRAPH and EUROGRAPHICS Campfire: Acoustic Rendering for Virtual Environments* (Snowbird, Utah, 2001), http://cns.bu/edu/~shinn/pages/RecentPapers.html.

[28] M. Schoolmaster, N. Kopo, and B. G. Shinn-Cunningham, "Effects of Reverberation and Experience on Distance Perception in Simulated Environments" (Abstract), *J. Acoust. Soc. Am.,* vol. 113, p. 2285 (2003).

[29] M. Schoolmaster, N. Kopo, and B. G. Shinn-Cunningham, "Auditory Distance Perception in Fixed and Varying Simulated Acoustic Environments" (Abstract), *J. Acoust. Soc. Am.,* vol. 115, p. 2459 (2004).

[30] K. Kurozumi and K. Ohgushi, "The Relationship between the Cross-Correlation Coefficient of Two-Channel Acoustic Signals and Sound Image Quality," *J. Acoust. Soc. Am.,* vol. 74, pp. 1726–1733 (1983).

[31] H. Haas, "The Influence of a Single Echo on the Audibility of Speech," Ph.D. dissertation, University of Göttingen, Göttingen, Germany (1949); translation reprinted in *J. Audio Eng. Soc.,* vol. 20, pp. 146–159 (1972 Mar.).

[32] S. E. Olive and F. E. Toole, "The Detection of Reflections in Typical Rooms," *J. Audio Eng. Soc.,* vol. 37, pp. 539–553 (1989 July/Aug.).

[33] J. S. Bradley, R. D. Reich, and S. D. Norcross, "On the Combined Effects of Early- and Late-Arriving Sound on Spatial Impression in Concert Halls," *J. Acoust. Soc. Am.,* vol. 108, pp. 651–661 (2000).

[34] P. M. Zurek, "Measurements of Binaural Echo Suppression," *J. Acoust. Soc. Am.,* vol. 66, pp. 1750–1757 (1979).

[35] M. Barron, "The Subjective Effects of First Reflections in Concert Halls—The Need for Lateral Reflections," *J. Sound Vib.,* vol. 15, pp. 475–494 (1971).

[36] A. J. Watkins, "Central, Auditory Mechanism of Perceptual Compensation for Spectral-Envelope Distortion," *J. Acoust. Soc. Am.,* vol. 90, pp. 2942–2955 (1991).

[37] A. J. Watkins and S. J. Makin, "Some Effects of Filtered Contexts on the Perception of Vowels and Fricatives," *J. Acoust. Soc. Am.,* vol. 99, pp. 588–594 (1996).

[38] A. J. Watkins, "The Influence of Early Reflections on the Identification and Lateralization of Vowels," *J. Acoust. Soc. Am.,* vol. 106, pp. 2933–2944 (1999).

[39] J. P. A. Lochner and J. F. Burger, "The Subjective Masking of Short Time Delayed Echoes by Their Primary Sounds and Their Contribution to the Intelligibility of Speech," *Acustica,* vol. 8, pp. 1–10 (1958).

[40] E. Meyer and G. R. Schodder, "On the Influence of Reflected Sound on Directional Localization and Loudness of Speech," *Nachr. Akad. Wiss. Göttingen,* (Math. Phys. Klasse IIa), vol. 6, pp. 31–42 (1952).

[41] A. Devantier, personal communication (a reconfiguration of the data in W. T. Chu and A. C. C. Warnock, "Detailed Directivity of Sound Fields around Human Talkers," National Research Council Canada, Institute for Research in Construction Rep. IRC-RR-104 (2002), http://irc.nrc-cnrc.gc.ca/ircpubs).

[42] B. G. Shinn-Cunningham, "Acoustics and Perception of Sound in Everyday Environments," in *Proc. 3rd Int. Workshop on Spatial Media* (Aisu-Wakamatsu, Japan, 2003), http://cns.bu.edu/~shinn/pages/RecentPapers.html.

[43] R. H. Bolt and P. E. Doak, "A Tentative Criterion for the Short-Term Transient Response of Auditoriums," *J. Acoust. Soc. Am.,* vol. 22, pp. 507–509 (1950).

[44] R. W. Muncey, A. F. B. Nickson, and P. Dubout, "The Acceptability of Speech and Music with a Single Artificial Echo," *Acustica,* vol. 3, pp. 168–173 (1953).

[45] P. A. Lochner and J. F. Burger, "The Subjective Masking of Short Time Delayed Echoes by Their Primary Sounds and Their Contribution to the Intelligibility of Speech," *Acustica,* vol. 8, pp. 1–10 (1958).

[46] G. A. Soulodre, N. Popplewell, and J. S. Bradley, "Combined Effects of Early Reflections and Background Noise on Speech Intelligibility," *J. Sound Vib.,* vol. 135, pp. 123–133 (1989).

[47] T. Nakajima and Y. Ando, "Effects of a Single Reflection with Varied Horizontal Angle and Time Delay on Speech Intelligibility," *J. Acoust. Soc. Am.,* vol. 90, pp. 3173–3179 (1991).

[48] A. Devantier, "Characterizing the Amplitude Response of Loudspeaker Systems," presented at the 113th

Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 50, p. 954 (2002 Nov.), convention paper 5638.

[49] Y. Ando, *Concert Hall Acoustics* (Springer, Berlin, 1985).

[50] Y. Ando, "Subjective Preference in Relation to Objective Parameters of Music Sound Fields with a Single Echo," *J. Acoust. Soc. Am.,* vol. 62, pp. 1436–1441 (1977).

[51] Y. Ando, *Architectural Acoustics* (Springer, New York, 1998).

[52] S. Bech, "Spatial Aspects of Reproduced Sound in Small Rooms," *J. Acoust. Soc. Am.,* vol. 103, pp. 434–445 (1998).

[53] J. S. Bradley, H. Sato, and M. Picard, "On the Importance of Early Reflections for Speech in Rooms," *J. Acoust. Soc. Am.,* vol. 113, pp. 3233–3244 (2003).

[54] H. Sato, J. Bradley, and M. Masayuki, "Using Listening Difficulty Ratings of Conditions for Speech Communication in Rooms," *J. Acoust. Soc. Am.,* vol. 117, pp. 1157–1167 (2005).

[55] S. E. Olive, P. L. Schuck, S. L. Sally, and M. E. Bonneville, "The Variability of Loudspeaker Sound Quality among Four Domestic-Sized Rooms," presented at the 99th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 43, pp. 1088, 1089 (1995 Dec.), preprint 4092.

[56] S. E. Olive, B. Castro, and F. E. Toole, "A New Laboratory for Evaluating Multichannel Audio Components and Systems," presented at the 105th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 46, pp. 1032, 1033 (1998 Nov.), preprint 4842.

[57] A. H. Benade, "Wind Instruments in the Concert Hall," oral presentation at Parc de la Villette, Paris, France, as part of a series of lectures entitled "Acoustique, Musique, Espaces" (May 15 1984), personal communication.

[58] F. E. Toole, "Loudspeaker Measurements and Their Relationship to Listener Preferences," *J. Audio Eng. Soc.,* vol. 34, part 1, pp. 227–235 (1986 Apr.); part 2, pp. 323–348 (1986 May).

[59] M. R. Schroeder, "Statistical Parameters of the Frequency Response Curves of Large Rooms," *Acustica,* vol. 4, pp. 594–600 (1954); translated from German, *J. Audio Eng. Soc.,* vol. 35, pp. 299–306 (1987 May).

[60] A. Baskind and J. D. Polack, "Sound Power Radiated by Sources in Diffuse Field," presented at the 108th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 48, p. 361 (2000 Apr.), preprint 5146.

[61] S. Olive, "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part I—Listening Test Results," presented at the 116th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 52, p. 861 (2004 July/Aug.), convention paper 6113.

[62] S. Olive, "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part II—Development of the Model," presented at the 117th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 53, p. 83 (2005 Jan./Feb.), convention paper 6190.

[63] R. P. Genereux, "Signal Processing Considerations for Acoustic Environment Correction," presented at the AES UK DSP Conference (1992).

[64] P. G. Craven and M. A. Gerzon, "Practical Adaptive Room and Loudspeaker Equalizer for Hi-Fi Use," presented at the 92nd Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 40, p. 436 (1992 May), preprint 3346.

[65] P. Rubak and L. G. Johansen, "Design and Evaluation of Digital Filters Applied to Loudspeaker/Room Equalization," presented at the 108th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 48, p. 365 (2000 Apr.), preprint 5172.

[66] R. H. Bolt, "Note on Normal Frequency Statistics for Rectangular Rooms," *J. Acoust. Soc. Am.,* vol. 18, pp. 130–133 (1946).

[67] J. S. Bradley, "Sound Absorption of Gypsum Board Cavity Walls," *J. Audio Eng. Soc.,* vol. 45, pp. 253–259 (1997 Apr.).

[68] F. E. Toole, "Loudspeakers and Rooms for Stereophonic Sound Reproduction," presented at the AES 8th International Conference (Washington, DC, 1990).

[69] T. Welti, "How Many Subwoofers Are Enough," presented at the 112th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 50, p. 523 (2002 June), convention paper 5602.

[70] T. Welti and A. Devantier, "In-Room Low Frequency Optimization," presented at the 115th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 51, p. 1244 (2003 Dec.), convention paper 5942.

[71] T. Welti and A. Devantier, "Low-Frequency Optimization Using Multiple Subwoofers," *J. Audio Eng. Soc.,* vol. 54, pp. 347–365 (2006 May).

[72] T. S. Welti, "Subjective Comparison of Single Channel versus Two Channel Subwoofer Reproduction," presented at the 117th Convention of the Audio Engineering Society, San Francisco, CA, 2004 October 28–31, preprint 6322.

[73] W. L. Martens, J. Braasch, and W. Woszczyk, "Identification and Discrimination of Listener Envelopment Percepts Associated with Multiple Low-Frequency Signals in Multichannel Sound Reproduction," presented at the 117th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 53, pp. 98, 99 (2005 Jan./Feb.), convention paper 6229.

[74] R. V. Waterhouse, "Output of a Sound Source in a Reflecting Chamber and Other Reflecting Environments," *J. Acoust. Soc. Am.,* vol. 30, pp. 4–13 (1958).

[75] R. F. Allison, "The Influence of Room Boundaries on Loudspeaker Power Output," *J. Audio Eng. Soc.,* vol. 22, pp. 314–320 (1974 June).

[76] J. A. Pedersen, "Adjusting a Loudspeaker to Its Acoustic Environment—The ABC System," presented at the 115th Convention of the Audio Engineering Society, *J. Audio Eng. Soc.* (*Abstracts*), vol. 51, p. 1223 (2003 Dec.), convention paper 5880.

THE AUTHOR

Floyd E. Toole studied electrical engineering at the University of New Brunswick, Canada, and at the Imperial College of Science and Technology, University of London, where he received a Ph.D. degree.

He joined the National Research Council of Canada in 1965, reaching the position of senior research officer in the Acoustics and Signal Processing Group. In 1991 he joined Harman International Industries, Inc., as corporate vice president—Acoustical Engineering. In this position he works with all Harman International companies and directs the Harman Research and Development Group, a central resource for technology development and subjective measurements.

Dr. Toole's research has focused on the acoustics and psychoacoustics of sound reproduction in small rooms. Most notably he established methods for subjective and objective evaluations, which have been used to clarify the relationships between technical measurements of audio equipment and listeners' perceptions. All of this work was directed to improving engineering measurements, objectives for loudspeaker design and production control, and techniques for reducing variability at the loudspeaker–room–listener interface. For papers on these subjects he has received two AES Publications Awards and the AES Silver Medal. He is a fellow and past president of the Audio Engineering Society and a fellow of the Acoustical Society of America.