The Measurement and Calibration of Sound Reproducing Systems

FLOYD E. TOOLE, *AES Life Fellow*

(soundnwine@sbcglobal.net)

Retired, Consultant to Harman International

The steady-state amplitude response measured at listening locations in a room is a widely accepted indicator of sound quality in sound reproduction systems. Room equalization schemes adjust the room curve to match a target believing that this ensures good and consistent sound. This paper examines both small and large venues, home theaters to cinemas, seeking a calibration methodology that could be applied throughout the audio industry. At present most of the industry follows a common philosophy but movie sound is a problematic exception. Some changes to current practice are indicated.

1 INTRODUCTION

For decades it has been widely accepted that a steady-state amplitude response measured with an omnidirectional microphone at the listening location in a room is an important indicator of how an audio system will sound. Such measurements have come to be known as generic "room curves," or more specific "house curves." That belief has a long history in professional audio, and now it has penetrated consumer audio with stand-alone products and receivers incorporating automated measurement and equalization capabilities. The implication is that by making in-situ measurements and manipulating the input signal so that the room curve matches a predetermined target shape, imperfections in (unspecified) loudspeakers and (unspecified) rooms are measured and repaired. It is an enticing marketing story.

In professional audio outside of the movie-sound domain the traditional pink-noise/real-time analyzer (RTA) process of measuring steady-state amplitude response has been superseded. Sound reinforcement system design and setup now routinely includes time-windowed measurements that can capture the direct sound, with selectable amounts of early and late reflected sounds. Generally the goal is well described as a variation on flat and smooth direct sound. Although final adjustments may involve a subjective opinion, measurements provide the foundation for defining system performance.

Sound quality is of fundamental importance in all aspects of the audio industry but perhaps most in sound *reproduction.* The timbral/spectral character of monitor loudspeakers and rooms affect the art as it is created in recording studios and dubbing stages, and again as it is reproduced for customers wherever they may be when a "play" button

is pressed. Significant uniformity throughout the process is needed if customers are to hear what the artists created. This is the "circle of confusion" shown in Fig. 1. For the system to function sensibly, mixing and mastering engineers need to experience sound that resembles what their customers will hear.

Acknowledging that audio systems in widespread use are not necessarily very good, audio professionals in the music side of the business have often used "bad" loudspeakers to check their mixes. The problem is that loudspeakers can be "bad" in countless ways. The dominant characteristics of small low cost audio devices are a lack of bass and reduced sound output capability—a high-pass filter in the playback signal path is a practical way to simulate that. The author's book [1] (chapters 2 and 18) illustrates the past and present situation in consumer and professional-monitor loudspeakers. Flat on-axis frequency response is clearly the engineering objective for most of these systems. Those that deviate significantly earn lower ratings in double-blind subjective evaluations. Although there is more to be considered, a flat direct sound delivered to listeners is the basis for most reproduced sound.

In terms of sound quality, the best examples of consumer loudspeakers are indistinguishable from the best studio monitors, although there are less than exemplary performers in both categories. The goal of this paper is to identify the key variables in sound reproducing systems that can lead to a calibration process for monitoring conditions during the creative process, as well as for reproduction systems for audiences of all sizes.

The following discussions will embrace both large and small venue sound systems: cinemas for public exhibitions, home theaters and stereo systems for private entertainment,
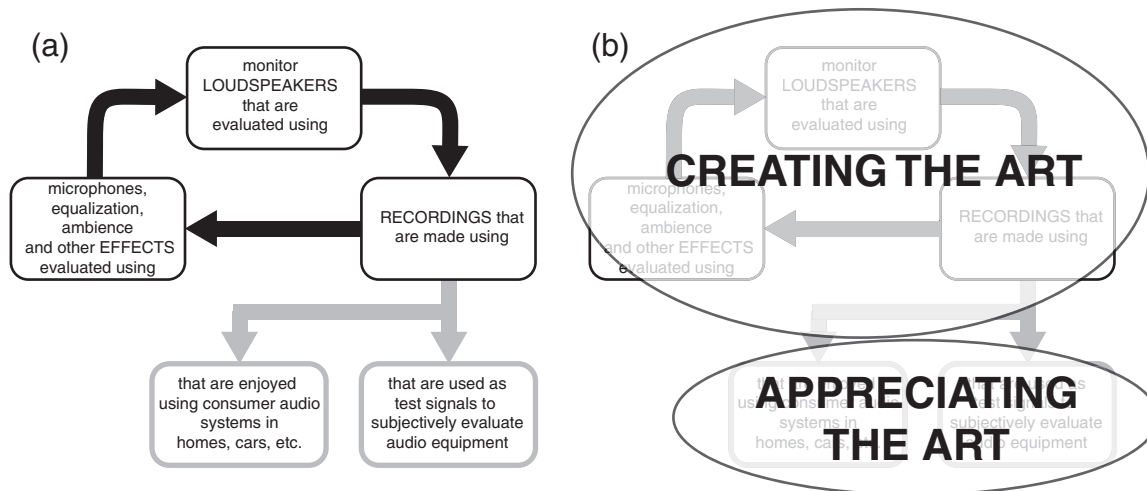
Fig. 1. The "circle of confusion" that underlies the audio industry. There is a requirement for mixing and mastering engineers to hear what their customers are likely to hear—the two domains shown in (b) need to have fundamental similarities. From [1] Figure 2.3.

dubbing stages for movie sound track creation, and recording control rooms for music production. The physics of sound creation and propagation are the same in all of these, as are the perceptual processes of listeners, so substantial commonality is anticipated.

2   SOUND FIELDS IN ROOMS

Sounds arriving at a listening location include characteristics that are traceable to the sound source (e.g., spectrum/amplitude response and directivity) and to the room through which the sound is communicated (the type and location of acoustical materials and the effects they have on the reflected sound field). Listeners identify aspects of sound quality (timbre), and also, because of binaural hearing, spatial attributes: localization, imaging, envelopment, etc. Some evidence suggests that the timbral and spatial perceptual dimensions may be comparable in importance, possibly not completely separable (discussed in [1] sections 8.2.1 and 20.1). Sound cues contributing to spatial perceptions are not revealed in steady-state amplitude-response measurements and the lack of directional discrimination makes time-windowed measurements ambiguous. Therefore, it is necessary to understand the spectral and directional properties of the sound sources as well as the sound reflecting behavior of listening venues in order to anticipate acoustical and perceptual events [2].

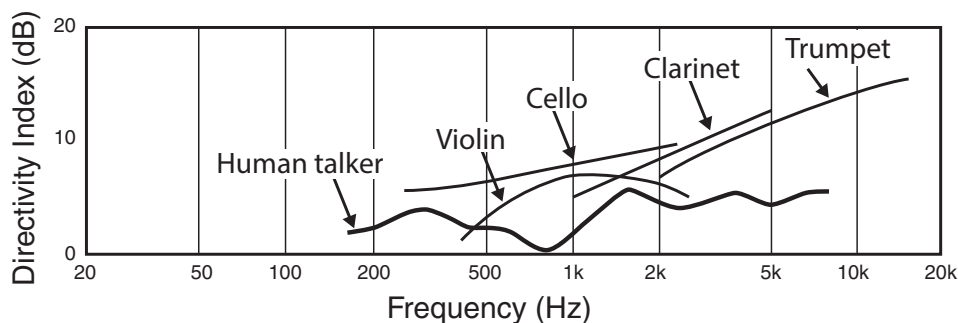### 2.1 Sound Source Directivity as a Function of Frequency

The audible spectrum, 20 Hz to 20 kHz, has wavelengths ranging from about 17 m to 17 mm. As a result the dispersion of radiated sounds changes with frequency depending on the size of the sound-radiating surface. It matters not whether we are considering voices, musical instruments, or loudspeakers. Low frequencies from most sources radiate essentially omnidirectionally because the wavelengths are long compared to the size of the source—the directivity index (DI) is near zero. The DI can be interpreted as the difference in dB between the on-axis sound and the total radiated sound power, which, in a room, is related to the difference between the direct sound and the reflected sound field. As frequency increases, so does the directivity of most sources. The higher the DI, the higher is the level of direct sound relative to later arriving reflections. Therefore, as a rule, humans are exposed to more energetic reflected sound fields at low frequencies than at high frequencies, whether we are at a live concert performance, carrying on a conversation in a corridor or listening to loudspeakers in a room. Fig. 2 shows examples of some musical instruments, voice, and loudspeakers.

Some musical instruments, especially stringed instruments, exhibit very complicated radiation patterns, so these are simplifications; Meyer shows much more detailed data in his book [3]. It is important to note that they fall into a similar range of directivities as the loudspeakers. Very clearly shown in the loudspeaker data is the importance of the size of the low frequency energy source, with the directivity index curve progressively flattening as the radiating area—the size and/or number of woofers—shrinks. There is also a reduction in DI as the angular dispersion of the high frequency horn expands, reducing further with the small dome tweeter in the domestic loudspeaker.

Whatever the shape of the spectrum of the direct sound from musical instruments, or loudspeakers reproducing recordings of those musical instruments, the steady-state sound field in a normal room will exhibit a version of that spectrum that rises at lower frequencies. Highly reflective concert halls attempt to preserve the limited sound output of musical instruments and voices in reflections, delivering as much of it as possible to listeners, while not masking temporal details in the music, and still creating a pleasant sense of envelopment. It is a difficult acoustical balancing act. Sound reproduction spaces are much less reflective: the principal cues to space and envelopment are in the

(a) Musical instruments and voice

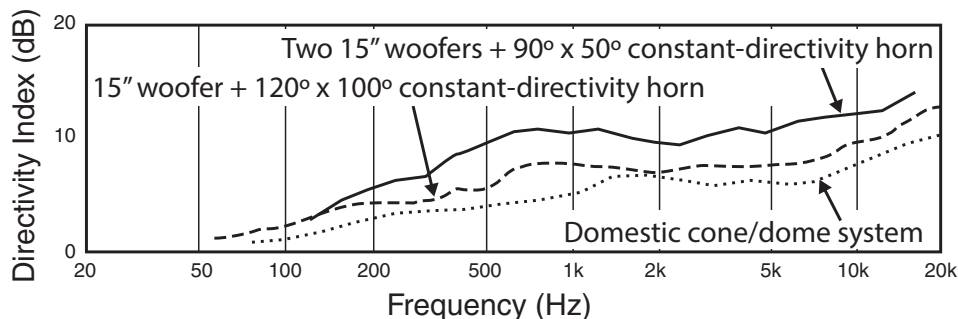(b) Loudspeakers: cinema, monitoring, domestic

Fig. 2. (a) The simplified directivity indices (DI) for some musical instruments [3]. The voice data are from [1] Figure 10.3. (b) Directivity indices for a large cinema loudspeaker system with a double-woofer bass unit and a 90° x 50° horn, a single woofer studio monitor system with a 120° x 100° horn, and a three-way domestic cone/dome system.

multichannel recordings, and a volume control changes loudness at will. Loudspeakers designed with flat on-axis frequency responses, so as to accurately reproduce the initial timbral signature of the recorded sounds, will therefore exhibit a rising sound power output at lower frequencies. The only exceptions would be arrays designed to maintain high DI at low frequencies.

2.2 Room Reflections as a Function of Frequency

The substantial low-frequency energy in movie sound tracks and music requires significant sound isolation to prevent these sounds from intruding on neighboring rooms and properties. High sound transmission loss at low frequencies requires room boundaries that are massive and stiff, with large spaces between layers. With little membrane absorption for bass sounds inside the space, reflections are more energetic and reverberation times increase. Low-frequency absorption in the quantities needed to be effective in large venues is expensive and, therefore, is frequently not provided. Fig. 3(a) shows the reverberation times of the cinemas and dubbing stages evaluated in the SMPTE B-chain report [4]. There is a wide range at low frequencies with only the "reference" cinema (B) and the two dubbing stages (E and F) showing evidence of substantial low-frequency absorption. Obviously these venues justified the additional expense of this acoustical treatment.

In cinemas, seats cover a high proportion of the interior surface area, and the directional loudspeakers are aimed at
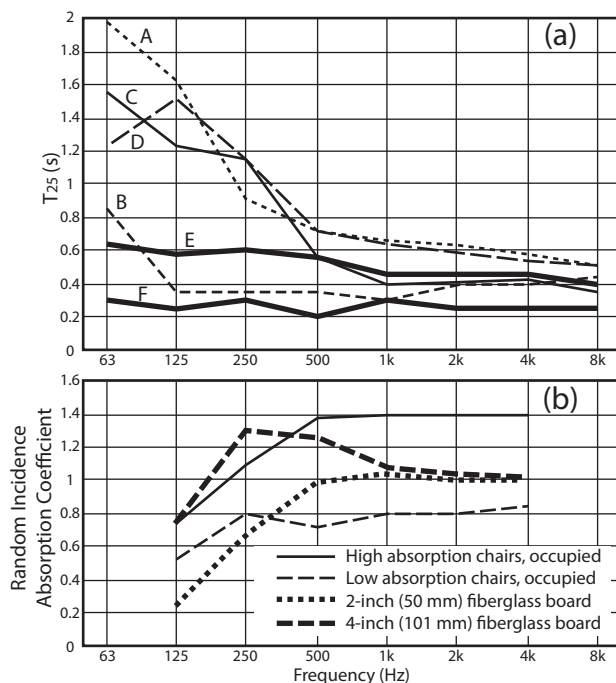

Fig. 3. (a) The reverberation times in four cinemas (A – D) and two dubbing stages (E and F) from [4]. Fig. 3(b) The random incidence absorption coefficients for two kinds of theater seats, from [5] and for two thicknesses of Fiberglas® 700 series insulation from www.owenscorning.com.

those seats. Fig. 3(b) shows that even generously uphol-stered chairs offer little sound absorption at low frequencies (there is only a small difference between occupied and unoccupied absorption) [5]. Fibrous absorbing materials commonly used on internal surfaces also provide little assistance at low frequencies. Anecdotal evidence suggests that 2-inch (50 mm) material is widely used, which is supported by the rapidly rising reverberation times below 500 Hz seen for cinemas A, C, and D in Fig. 3(a). Most of the middle-and high-frequency sound from the loudspeakers is absorbed on first contact, leaving little energy to participate in reflections. Diffuse reverberation is impossible. Only the expensive application of low-frequency absorbers prevents the build-up of reflected sound energy at low frequencies, and this is not common in mass-market cinemas.

The special problem of room resonances and standing waves in small rooms is strong encouragement to damp these modes. Rooms in wood frame houses have some naturally-occurring low-frequency membrane absorption [6]. Otherwise the absorption provided by carpets, drapes, and upholstered furniture is similar to the curves in Fig. 3(b). Custom designed home theaters and recording control rooms are usually acoustically isolated, requiring massive boundaries. They then exhibit problematic standing waves that need attenuating.

To summarize, with live sound and reproduced sound using typical loudspeakers in typical enclosed spaces we can generally anticipate that as frequencies decrease the room sound field will (a) contain more reflections, and (b) increase in sound level as the reflected energy accumulates. This is natural acoustics at work. It is possible, indeed probable, that these attributes are recognized by the human perceptual process and somehow incorporated into our expectations.

2.3 Sound Fields in a Specific Room with a Specific Loudspeaker

This example is from the author's early investigations in 1985–86. The loudspeaker was a three-way design with a flat on-axis amplitude response but with significantly non-flat off-axis performance. The venue was the prototype listening room on which the original IEC 268-13 [7] was based, described in the appendix of [8]. The reverberation time was about 0.3 s at frequencies above 300 Hz, rising at lower frequencies to a maximum of 0.8 s at 40 Hz. All anechoic measurements were made at 2 m in a chamber with 1 m wedges, calibrated to 30 Hz using a free-field (tower) reference. The following data come from [9] Part 2, section 7.2.

Fig. 4(a) shows anechoic loudspeaker measurements in which the excellent on-axis performance is clear, as is the undesirably irregular off-axis performance. Similar undulations can be anticipated in the total sound power output, which is dominated by off-axis radiated energy.

Fig. 4(b) shows the principal components of the sound fields as they were estimated to be at the prime listening location. In addition to the direct sound (the on-axis response) there is an energy sum of first reflections from the floor,

ceiling, and sidewalls derived from the appropriate anechoic off-axis responses, assuming spectrally perfect reflections and including inverse-square-law attenuation. There is also an estimate of the contribution made by the total radiated sound power, calculated from a weighted sum of measurements made on 360° equatorial and polar orbits around the loudspeaker, modified according to the frequency dependent absorption in the room revealed in reverberation-time measurements.

The first observation is that if one wishes to anticipate how this loudspeaker might sound in a room, it is necessary to pay attention to sound power at low frequencies—it is the highest curve in Fig. 4(b)—and the direct sound at highest frequencies for the same reason. In the middle frequencies all three components contribute significantly so all three need to be measured. The on-axis curve by itself is insufficient data. Full 360° data, appropriately processed, is important information. An energy sum of the measurements contributing to these curves yields an estimate of a steady-state room curve, shown shifted upwards by 10 dB.

The loudspeaker was then placed at three possible stereo-left locations in the listening room and averaged measurements at six seats yielded the curves shown in Fig. 4(c) with the prediction from Fig. 4(b) superimposed. Above about 400 Hz the curves are essentially identical and the prediction aligns well with the measurements. Below this the effects of standing waves in the room dominate, and the locations of the loudspeaker and the listener/microphone determine the acoustical coupling at different frequencies. The predicted curve provides an estimate of the upper limit of the steady-state sound levels, but destructive interference in the standing waves substantially reduces overall bass energy. Below the transition/Schroeder frequency, around 300 Hz here, the room is the dominant factor; above it, the loudspeaker is substantially in control.

Fig. 4(d) attempts to illustrate the sounds arriving at the listening location in a typical domestic listening room or home theater—different portions of the radiated sound dominate at different frequencies, determined by the frequency-dependent directivity of the loudspeaker and the reflective nature of the room. Obviously the frequencies at which the transitions occur will change with different loudspeaker designs and with different room acoustical configurations. In general, as loudspeakers become more directional and/or rooms become less reflective, the transitions move down in frequency.

These data illustrate some fundamentally important concepts:

- With sufficient anechoic data on a loudspeaker it is possible to predict with reasonable precision middle- and high-frequency acoustical events in a listening space with known properties.
- There is a difference between the spectrum of the direct sound arriving at a listener and that of the steady-state sound level that is achieved after reflected sounds arrive. The shape of a steady-state room curve is determined by the sound radiated
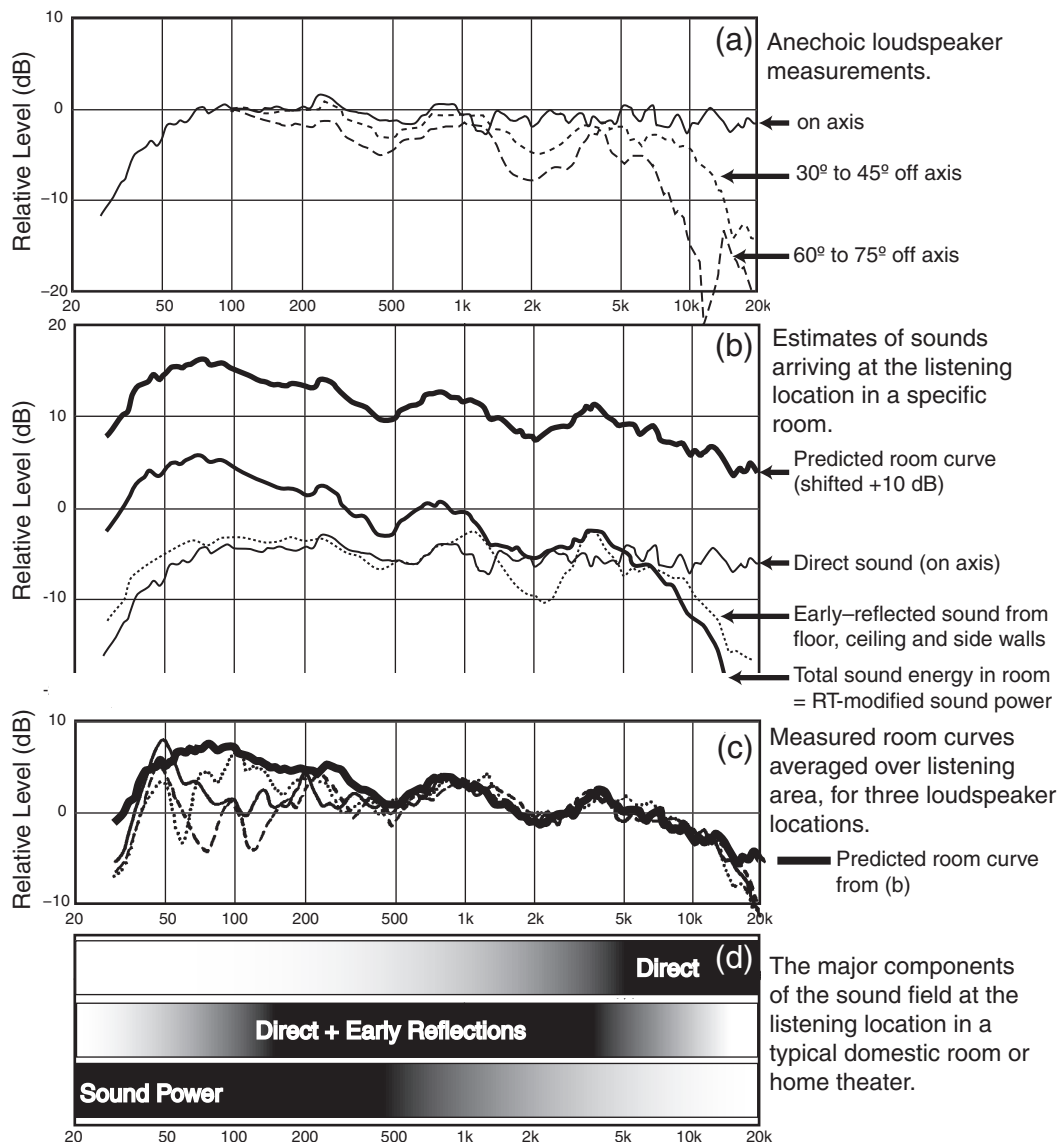
Fig. 4. Predictions of the direct and reflected sound fields in a room from a set of anechoic measurements on a loudspeaker. Descriptions are in the Figure and in the text. Data from [9] Part 2.

by the loudspeaker modified by the geometry and frequency-dependent reflectivity of the room. In an acoustically dead room, the room curve will be identical to the on-axis response of the loudspeaker. As reflections within the room increase, the room curve will rise towards the predicted room curve, as the off-axis sounds add to the result. The bass and midrange sound levels will build up over a short time interval, affecting what is measured and heard. At very high frequencies the direct sound becomes progressively dominant. Therefore with no knowledge of the loudspeaker, and no knowledge of the room acoustical properties, a steady-state room curve conveys ambiguous information.

- In normal rooms the on-axis frequency response is not the dominant physical factor. However, the direct sound has a high priority in perception, establishing a reference to which later arrivals are com-

pared in determining such important perceptions as precedence effect (localization), spatial effects, and timbre. In this example, the poor off-axis performance dominated the in-room measurements and in listening tests caused audible timbral degradation. Equalization of the room curve will destroy the only good performance in the loudspeaker—the on-axis/direct sound response. Equalization cannot change loudspeaker directivity; the remedy is a better loudspeaker. Adequate anechoic data on the loudspeaker would have revealed the problem in advance of measurements or listening.

- Below the transition/Schroeder frequency the room resonances and the associated standing waves are the dominant factors in what is measured and heard. These are unique to each room and are strongly location-dependent. Only on-site measurements can reveal what is happening and different loudspeaker

and listener locations will result in different bass sound quality and quantity.

- In domestic listening rooms and home theaters or in any acoustically well-damped room like a cinema, traditional diffuse reverberation is significantly absent. As discussed in [2] the sound fields in low-reverberation rooms are different from those described in classical acoustics.

Different sound sources in different rooms will change the pattern of sound fields shown in Fig. 4(d), moving the transition regions up or down the frequency scale, but the basic principles hold. Cinemas and dubbing stages vary, and the SMPTE B-chain report [4] gives a small sample of what can be expected with typical loudspeakers and acoustical treatment. As a crude generalization, at low frequencies there is evidence of significant room reflection and adjacent-boundary effects. From about 200 Hz to around 600–1000 Hz the energetic sound events happen within about the first 50 ms—listeners are exposed to direct plus a few early-reflected sounds. Above this, for the top three octaves or more, the direct sound is the dominant factor.

2.4 "Room Equalization" Is a Misnomer

It is a bold assertion that a single steady-state measurement in a room—a room curve—can reliably anticipate human response to a complex sound field. Such measurements take no account of the direction or timing of reflections within the sound field. Time-windowing the measurement is useful to separate events in the time domain, but these too ignore the directions from which sounds arrive. Human listeners respond to these cues, in some detail, and they exhibit skills in separating room sound from the timbral identity of loudspeakers, and in adapting to different circumstances. This is, after all, what happens at live, unamplified, musical events. This means that not everything measured is perceptually important, nor can our reaction to such sound fields be constant, we adapt (see [1] chapters 5–11, and section 11.3.1, and [2] section 2.5). The simple measurements therefore cannot be definitive.

Assuming that we had a credible target for a room curve, not all irregularities seen in it indicate the presence of a problem. If a problem is thought to exist, the visual cues do not lead unambiguously to a cause, and therefore to the appropriate remedy, as was seen in Fig. 4 and the accompanying discussion. The advent of DSP brought with it many explorations of what could be done when the starting point is a measurement in a room that is dominated by non-minimum-phase acoustical interference. Genereux [10] provides a good perspective on early efforts, and Fielder [11] provides a detailed analysis of other approaches. The results are mixed, with substantial limitations being found to what can be corrected. The underlying issue is one that Genereux stated clearly: "[W]e are not interested in removing all the effects of the room. Instead we wish to remove any audible colourations evident to the listener." The application of DSP in any of its many forms requires guidance from psy-

choacoustic research relating what we measure to what we hear and, historically, this has been significantly lacking.

Equalization is very limited in what it can "correct," yet the notion that changing the signal supplied to a sound system consisting of an unknown loudspeaker in an unknown room can "equalize" or "calibrate" a system is widespread. In the context of a practical application *where there is an audience of several listeners* conventional equalization cannot:

- Add or remove reflections
- Change reverberation time
- Reduce seat-to-seat variations in bass
- Correct frequency dependent directivity in loudspeakers
- Compensate for frequency dependent absorption in acoustical materials and furnishings. The exception is in the highly reflective sound field at very low frequencies.

It is essential to separate events above and below the transition/Schroeder frequency. Above it, at middle and high frequencies, constructive and destructive acoustical interference occurs when direct and reflected sounds combine at a microphone. If the frequency resolution of the measurement is sufficiently high—typically 1/3-octave or higher—the resulting peaks and dips can look alarming when seen in room curve. The tendency for a calibrating technician or automated equalization algorithm might be to attempt to smooth the curve. These are non-minimum-phase phenomena that are not correctable by minimum-phase equalization.

However, the direct and reflected sounds generally come from different directions, which a microphone cannot differentiate, but binaural hearing can. References [1, 2, 12–14] focus on several of the perceptual consequences of reflections. It turns out that in most small-to-medium-sized sound-reproduction spaces human listeners find these multi-directional reflected sounds to be mostly benign, even beneficial if the loudspeaker has relatively constant directivity. A common perception is spaciousness—information about the listening space, not timbre-damaging comb filtering. This is certainly true for recreational listening, but professionals may find that a less reflective space is preferred for mixing but perhaps not for mastering recordings [12]. These measurements are therefore misleading, and even if equalization were capable of removing such reflections, there is the decision of whether it is necessary. Overall, equalizing the spectral fluctuation "errors" found at *a few* measurement locations, can add new spectral variations to the total sound output of the loudspeaker that is radiated to *all* locations throughout the room. There is a significant risk of degrading the performance of good loudspeakers.

Below the transition/Schroeder frequency equalization has a role to play. In the upper-bass frequency range adjacent-boundary interactions affect the sound energy radiated into the room. Brought to the attention of the audio community by Allison [15, 16] and discussed in [1] chapter 12, these fluctuations can be corrected for by equalization,

using a spatially-averaged measurement to reveal the underlying curve. In the low-bass frequency range equalization can be very useful as a means of attenuating prominent room resonances at a single listening location. This is more successful in rooms with significant low-frequency absorption. Alternatively, with passive and active combinations of multiple subwoofers standing waves can be manipulated, seat-to-seat variations can be reduced, and global equalization can then be beneficial to several listeners (see [2,17–20], and [1] chapter 13). In-room measurements are therefore useful at low frequencies, differently in different sizes of rooms, and bearing in mind that the target performance for steady-state room curves may not be a horizontal straight line.

In conclusion, there are reasons to exercise great caution in the application of equalization based on conventional in-room measurements. However, it is definitely advantageous at lower frequencies, and later discussions will show that equalization based on anechoic data is also useful in the creation of superior loudspeakers.

3 ROOM CURVES: PREDICTIONS AND REALITIES

The objective set for this paper was to seek a single methodology for designing and calibrating sound reproducing systems of all kinds. The following discussions dwell on movie sound in cinemas and dubbing stages because performance standards exist [21, 22] and we wish to examine how they perform. The movie sound tracks created for cinema exhibition are also delivered to a wide audience through discs, television, and streaming, the playback systems for which do not follow those standards, meaning that compatibility with the audio world outside movie sound is a concern. Music uses these consumer delivery media, and occasionally penetrates the cinema world in the form of concert videos, so there is considerable overlap in the world of sound reproduction, all of which argues for a single all-embracing performance standard, modified as necessary to accommodate idiosyncrasies of the playback venues.

What should the target be based on: direct sound or steady-state sound? There is a choice, and the answer may be a combination of the two, employed over different portions of the frequency range. Only outdoors or in *very* dead rooms are the two identical. In the real world, rooms differ in their delivery of early reflections from loudspeakers by virtue of their geometries and acoustical treatment. So, before going any further, we must identify metrics that portray the important differences in direct and steady-state sound fields that will inevitably happen.

3.1 Measures of Room Reflections

Reverberation time (RT) had its origins in the significantly diffuse sound fields of concert halls and auditoriums. Desirable mid-frequency RTs, measured using an omnidirectional sound source in those spaces, range from 1 to 2 s, or so, depending on the intended musical program—the hall is part of the performance and an omnidirectional source is

assumed to be an adequate approximation of an orchestra for measurements. In cinemas and home theaters cues for directional and spatial effects, and a sense of envelopment, are embodied in multichannel recordings. The room is not required to contribute consequentially to the experience, but it should not degrade it.

RT estimates the time required for the sound level to fall by 60 dB, a much greater range than is perceptually relevant to sound reproduction. It is convenient that RTs found in cinemas [4] are much the same as those found in homes and recommended for home theaters [1, 2, 7, 8]: approximately 0.2 s to 0.5 s. A naturally diffuse sound field cannot exist, nor is it desired. Although measuring instruments give us a number for RT, what is measured is a sequence of progressively attenuated reflections—a Reflection Decay Time, or as suggested in [4] section 9.1, simply Decay Time. As was done there, the term RT will be used here because of its familiarity, but keep in mind what is actually being measured.

RT is undoubtedly useful for evaluating highly reflective venues, but there may be a better metric for our purposes: cumulative energy time—the time taken after the arrival of the direct sound for the sound field to rise to the steady-state level. This is a much shorter time than the corresponding RT and seems to more directly address the perceptual processes when brief sounds are involved. For example, Fig. 8(b) shows that in a cinema with a 2.5 s (2500 ms) RT at 50 Hz, a level within 2 dB of steady state is reached in 90 ms. At 500 Hz, RT is 800 ms and the cumulative energy time is 25 ms. These are enormous differences, and although RT is a related parameter, it is far removed from the temporal events that are likely to matter. Events very early in the RT decay data could be relevant but a new form of interpretation would be required.

Germane to this discussion are the findings of Bradley et al. [23] indicating that for speech intelligibility—a crucial consideration for movies—it is the early reflections that are the main contributors. They concluded that early reflection energy arriving within about the first 50 ms following the direct sound has the same effect on speech intelligibility scores as an equal increase in the direct sound energy. This was true for both normal and hearing-impaired listeners. They go on to say: "Although it is important to avoid excessive reverberant sound, adding large amounts of absorption to achieve very short reverberation times may degrade intelligibility due to reduced early reflection levels." They suggest a ratio of the energy within the first 50 ms of an impulse response to the energy associated with the direct sound as a new figure of merit for a room, called the early reflection benefit (ERB).

In addition to the increase in sound amplitude level as reflections arrive, the duration of the original sound is effectively extended. The result of the added duration is that loudness is increased and the detection threshold is reduced—we may be able to hear more than the physical sound levels alone would suggest. The reduction in threshold might be of the order of about 3 dB per doubling of signal duration over the range 20 to 100 ms, reducing for longer durations, for example [24]. On a related topic, it is

worth noting that early reflections also reduce the thresholds for detecting resonances in loudspeakers [25] making flawed loudspeakers more noticeable and revealing more timbral subtleties in music. For stereo recording and listening, early reflections can reduce the timbral degradation and speech intelligibility loss in the phantom center image. It helps to fill the large spectral dip around 2 kHz, created by stereo/interaural crosstalk ([26] and [1] section 9.3.1).

Leembruggen [27] showed several examples from the SMPTE report [4] in which cumulative energy data appeared to relate to details in frequency-domain measurements in ways that RT did not, and thereby, possibly to perceptions. In the context of concert halls, Bradley et al. [28] found that the perceived strength of bass sounds was related to early and late reflection arrivals and not significantly related to low-frequency reverberation time.

Multiple factors are at play in this situation but there are several persuasive reasons to pay attention to early reflections, and to consider alternatives to RT (and its familiar derivatives) as a criterion of acoustical performance in sound reproduction venues. In any event, both RT and cumulative energy measures are frequency dependent, confirmation that direct and steady-state sound fields have different spectra in most rooms.

3.2 Predictions of Sound Fields

The following expands upon the information in Fig. 2(b). In Fig. 5(a) the directivity index curves are inverted, thereby giving us an estimate of total radiated sound power when the loudspeakers are adjusted for flat axial amplitude response (compensated for screen loss in cinemas). Time-windowed and close-field measurements show that in cinemas and dubbing stages the sound field is dominated by direct sound above 1 to 2 kHz [4, 29–32]. Accordingly, the curves have been adjusted so that the high-frequency portions are aligned—the room has little influence on the spectral shape in this part of the frequency range other than contributing air attenuation as a function of distance. Below these frequencies, reflected sound from the ever-widening dispersion of the loudspeakers is the basis for the rapidly rising curve for the large two-woofer loudspeaker systems and a lower slope for the one-woofer system. The size of the sound source is the relevant parameter. The immediately following discussions will focus on the two-woofer cinema system.

Fig. 5(b) shows a simplification of the two-woofer data in Fig. 5(a) with 10 m of high-frequency air attenuation slightly rolling off the flat direct sound for realism. Perceptually, humans may or may not instinctively understand that long distance propagation results in high-frequency loss. This is entirely possible for real sound sources, but is it true for movies where the anticipated distance is likely to be related to the image on the screen—that is, if we truly "suspend disbelief"? It is an interesting question.

Below 1 kHz the curves show what could happen if different amounts of the radiated energy were reflected by the room boundaries and delivered to listeners. The "typically reflective" curve is the author's interpretation of the RT

and sound absorption data shown in Fig. 3, specifically for cinemas A, C, and D. No precision is claimed for this estimate. The shape of the curve is merely intended to reflect the fact that these spaces have an abundance of absorption from high frequencies down to about 500 Hz, diminishing at lower frequencies (Fig. 3(b)). The difference between the direct and steady-state curves is therefore smaller over this frequency range than in a highly reflective venue. Rooms with different properties could yield curves that fall anywhere in the shaded space.

SMPTE ST 202 [21], ISO 2969 [22], and a few other audio industry recommendations require systems to be adjusted to a flat steady-state frequency response target below about 1 kHz. Fig. 5(c) shows that in these instances a flat direct sound has been lost in anything but very dead rooms. The high-frequency portion of the response does not change because listeners are in a dominantly direct sound field.

Does this mean that the system sounds deficient in bass or lacks transient attack? The amount of bass energy accumulation after the arrival of the direct sound will depend on the low-frequency reflectivity of the cinema, which varies considerably. It could also depend on the program being auditioned, with possible differences between sustained and transient bass sounds. The energy accumulation times are short ($<$200 ms), but the amplitude differences can be large ($\leq$ 10 dB). If any of this is significantly audible, it means that dubbing stages and cinemas should have similar loudspeakers and acoustics—see the circle of confusion, Fig. 1. If this is not possible or practical, an intermediate target curve might be chosen (within the shaded area) and the necessary electronic equalization applied. It is not possible to decide this with certainty without specific psychoacoustic research investigating how strong the audible effects are and whether simple equalization is a suitable compensation. It is conceivable that human listeners consider bass rises following a flat direct sound as an innocent component of room sound that is accommodated by adaptation.

Before moving on, it is important to note that the sound output and DI of woofers will be modified by their proximity to a wall behind or beside or a floor under them. Installation in a baffle wall creates a 2π solid angle. These are all "adjacent-boundary" effects that modify radiated sound power at lower frequencies. As discussed earlier, they are measurable by spatial averaging over the audience area, and equalization is a suitable remedy. However, many cinemas have no baffle walls, or partial baffle walls, or insufficient acoustical treatment in the space behind the screen or partial baffle wall, creating audible consequences that are not likely to be revealed by steady-state in-room or loudspeaker anechoic data. Only acoustically-informed visual inspection of the site can reveal such problems.

3.3 B-Chain Sound Systems in Cinemas and the X-Curve

Before examining current data, let us look at some from the past. Fig. 6 shows steady-state room curves for cinema systems having flat on-axis frequency responses. The 1961 Snow curve was corrected by an amount to yield a flat
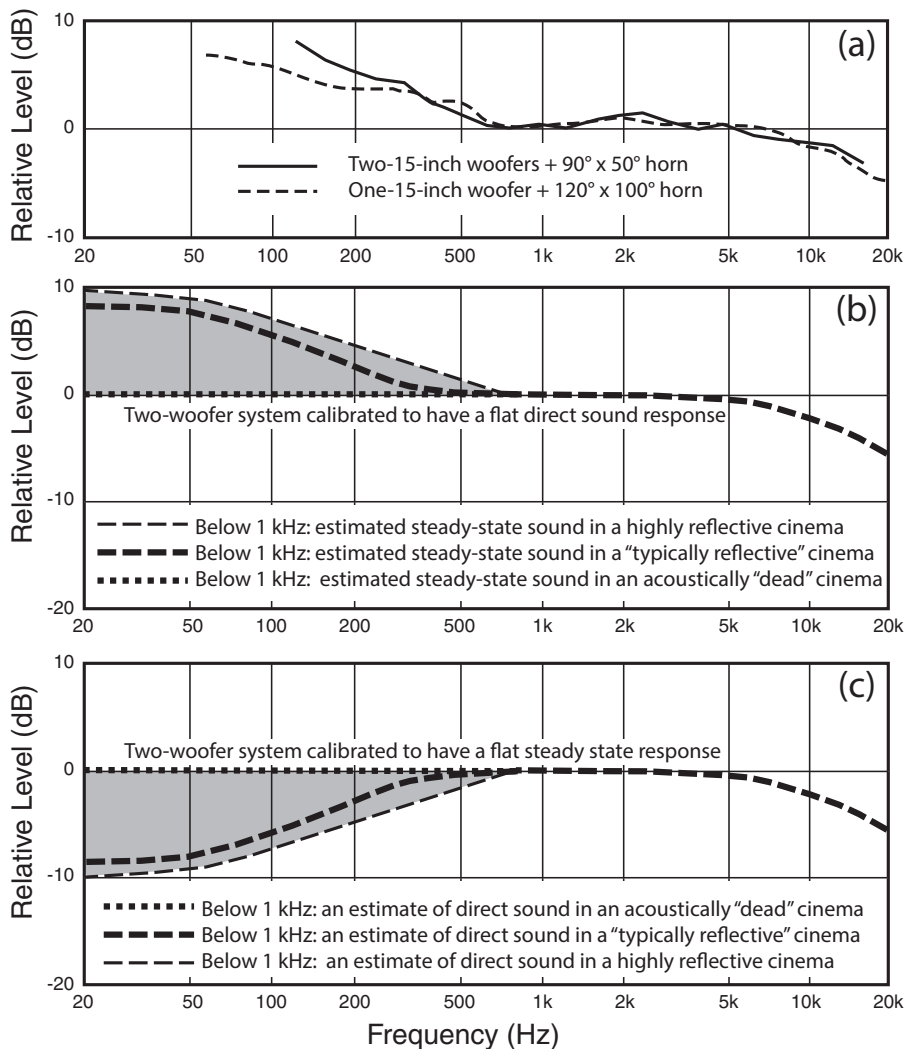
Fig. 5 (a) The estimated sound power radiated by two of the loudspeaker systems described in Fig. 2(b) assuming flat on-axis amplitude response. Fig. 5(b) A two-woofer cinema system with a flat direct sound response on the audience side of the screen radiates sound into three different venues. 10 m of high-frequency air attenuation is included. Fig. 5(c) The consequences to the direct sound of equalizing a sound system to a flat steady-state room curve room at low frequencies.
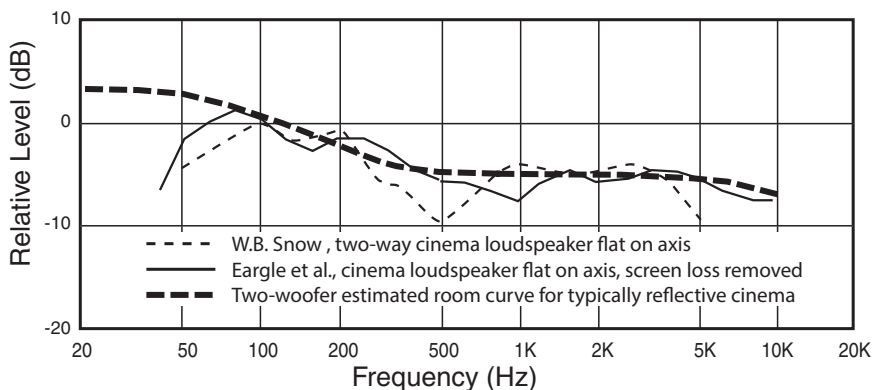

Fig. 6. Steady-state room curves for large venue cinema loudspeakers with basically flat direct-sound frequency responses. The curves are compared to the two-woofer "typically reflective" room curve prediction from Fig. 5(b).
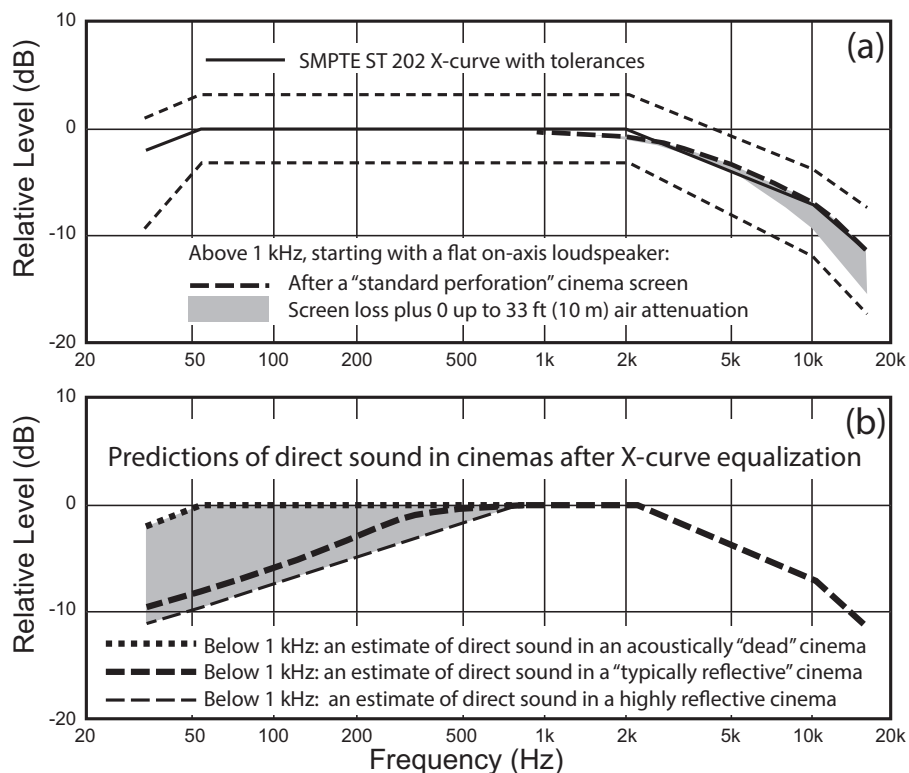
Fig. 7 (a) The SMPTE X-curve [21] compared at frequencies above 1 kHz to the direct sound from a flat on-axis loudspeaker radiating through a standard-perforated cinema screen. The shaded area shows additional high-frequency rolloff due to air attenuation that increases with distance to the listeners. Fig. 7(b) A steady-state X-curve above 2 kHz means that the high-frequency rolloff is also in the direct sound. Achieving a flat steady-state response below about 1 kHz means that the low frequency direct sound may be rolled off. Shown here are predictions for the direct sound in cinemas with different reflective characteristics, employing a two-woofer loudspeaker system—adapted from Fig. 5(c). Cinemas with other reflective characteristics will yield direct sound curves that fall in the shaded area.

direct-sound response (2π on-axis data were shown) [33] and the 1985 Eargle et al. [34] curve was corrected by an amount to remove the screen loss that attenuated the highest frequencies in the direct sound. Both are compared to the "typically reflective" room curve in Fig. 5(b). The anticipated low-to-mid frequency tilt is seen, as is the flat, direct-sound dominated curve above about 1 kHz. Also visible are possible problems with the mid-frequency sound power constancy of the older cinema loudspeaker system.

Measurements of this kind are scarce in the archival literature. Unfortunately there is little public documentation of the important experiments, in the early 1970s, where the cinema sound X-curve target was decided, Fig. 7(a). The resulting standard, which exists in SMPTE [21] and ISO [22] versions, was the most recent in a series of evolutionary changes influenced by many factors, as explained by Allen [35]. At that time, measurement capability was limited, with pink noise and real-time analyzers not being able to reveal time-domain information and, very likely, incomplete information was available about loudspeaker performance. Sound tracks were also significantly compromised compared to current digital technology. The past trend had been towards a wider bandwidth, flatter target curve, but a residual high-frequency rolloff remained in the last iteration. At

that time the thinking was that "if a room is tuned with pink noise as a test signal to have a 3 dB per octave slope from 2 kHz, the first-arrival signal will be closer to flat than the 3 dB per octave seen on an analyzer would suggest." [35]. This morphed into a common belief that seeing the X-curve high-frequency rolloff in steady-state measurements at the 2/3 listening distance ensured a flat direct sound. We now know that is not true with today's loudspeakers in today's dubbing stages and cinemas—direct and steady-state sounds above about 1 kHz are essentially identical [4, 29–32]. In other words, the standardized rolloff results in a dulling of the sound. Further thought led to extensions of the notion in which the slope of the curve varied with room (audience) size. As will be seen, this too is an idea in need of reconsideration.

Loudspeakers are normally designed to radiate a flat on-axis sound, meaning that achieving the flat X-curve target below 2 kHz requires reducing the bass radiated by the loudspeakers (see Fig. 6). The high-frequency rolloff is automatically achieved by the attenuation of common cinema-perforated screens. The X-curve in Fig. 7(a) has been the internationally recognized performance target for movie sound dubbing stages and cinemas for decades. It is used nowhere else.

Fig. 7(b) shows the low-to-mid-frequency direct-sound predictions of Fig. 5(b) added to the high-frequency rolloff required by the SMPTE ST 202 X-curve target. Not only is the direct sound rolled off at high frequencies, it is rolled off at low frequencies as well. The amount of the rolloff depends on the reflectivity of the venue and the directivity of the woofers. If a flat direct sound was the objective in calibrating cinemas, the attempt failed.

3.4 Real-World Measurements Provide Confirmation

The SMPTE 2014 B-chain report [4] combines data, analysis, and explanations that make it a benchmark in this field. Some of this information has been selected in what follows, beginning with two venues that exemplify the range of reflectivity that exists. As shown in Fig. 3(a), Cinema A is a relatively reflective large cinema, while dubbing stage F is acoustically very well damped over the entire frequency range.

Fig. 8(a) confirms the predictions of Fig. 7(b) in X-curve-calibrated Cinema A. The direct sound is estimated by two measures: a 50 ms time-windowed measurement spatially averaged over five positions around the 2/3-distance into the audience, and a close-field microphone array at about 5 m from the screen located in the direct-sound path from the loudspeaker to that location. These both exhibit substantial rolloff of the direct sound at low frequencies and closely match the steady-state rolloff at high frequencies. The fact that the 50 ms curve is slightly above the close-field curve is possibly related to the inclusion of some early-reflected energy within the 50 ms measurement window.

That these direct-sound measurements meander around the predicted curves for the two-woofer installed system is a reassurance that the prediction is based on a realistic interpretation of the physical events. The indication is that this cinema ranges from "typically" to "highly" reflective" that the RT and cumulative energy times shown in Fig. 8(b) confirm. Fig. 8(c) shows the equalization performed by the calibrators. As discussed earlier and predicted in Fig. 5(b), a flat on-axis loudspeaker will yield a rising steady-state low frequency curve in a reflective room. This curve shows attenuation by equalization to meet the X-curve requirement of a flat steady-state room curve at low frequencies. A broad hump centered on 500 Hz is puzzling. It is not likely to be a loudspeaker problem, but, as will be seen later, it might be related to seat interactions. If so, equalization is a questionable action. The author attended these measurements, which involved listening to soundtracks at the end of the exercise. In response to comments about unnatural-sounding, timbrally-colored, voices the equalization was turned off and voice quality significantly improved. This is opinion, not fact, but research is probably indicated. Above 2 kHz almost no equalization is done up to a 4 dB low-Q bump around 10 kHz. The loudspeaker appears to be quite flat on axis and the X-curve high-frequency rolloff is almost perfectly achieved by the perforated-screen loss, as shown in Fig. 7(a).

Dubbing Stage F [4] was designed to have a very low and constant reverberation time at all frequencies, which creates a very different listening situation. With reflected sounds substantially attenuated, Fig. 9(a) shows that the direct sound dominates over most of the frequency range, with some reflected sound gain showing only below about 50 Hz. The directivity of the loudspeaker is almost irrelevant. Calibrators let the high frequencies fall below the X curve.

Seeking data from other sources, Fig. 10 shows measurements done by Holman [31] in what was described as a modern stadium-seating facility. The large difference between direct and steady-state curves indicates a relatively reflective venue—there was no reverberation-time data, but the author commented that there was "not . . much low-frequency absorption." The direct sound measurements wander around the predictions from Fig. 7(b). A prominent peak and dip between 150 and 300 Hz suggest problems in the venue: acoustics or equalization, or both? Calibrators let high-frequency levels fall below the X curve, and left the low frequencies elevated, creating a bass-to midrange downward tilt.

A noteworthy study by Fielder [30] combines data from 50 front loudspeakers in 18 cinemas, Fig. 11. Looking below 1 kHz in Fig. 11(a) reveals that the "typical 500-seat cinema" in his survey exhibits a direct sound curve that is very close to the "typically-reflective" prediction of Fig. 7(b) (the bold dashed line).

Fig. 11(b) shows an informative superimposition of all 50 front loudspeakers in the 18 cinemas. These curves show the spectra of the direct sound in these venues, revealing just how varied the playback of film sound is across a selection of loudspeakers and venues. They basically fill in the shaded area of possibilities shown in Fig. 7(b). There was no information about the loudspeakers. The lower boundary is the two-woofer sound power prediction from Fig. 7(c), identified as the "completely reflective cinema," implying that there is little additional "gain" associated with these rooms, mainly losses due to absorption.

3.5 The Controversial X-Curve

There are two basic issues with the X curve:

- The first is the high-frequency rolloff that has been the subject of much debate and misunderstanding. Because of technical advances in all aspects of audio recording and reproduction, from digital audio through power amplifiers and loudspeakers, the circumstances of its origins no longer pertain. At this time there seems to be no justification for the rolloff. Because of the directional loudspeakers and cinema acoustics, this rolloff exists almost identically in the direct sound and in the steady-state sound—it is a rolloff in a target curve, not a consequence of natural room acoustics as has been a common misconception. The direct sound arriving at listeners is not flat.

- The second issue is the flat steady-state target below 2 kHz. Fig. 11(b) shows that, depending on
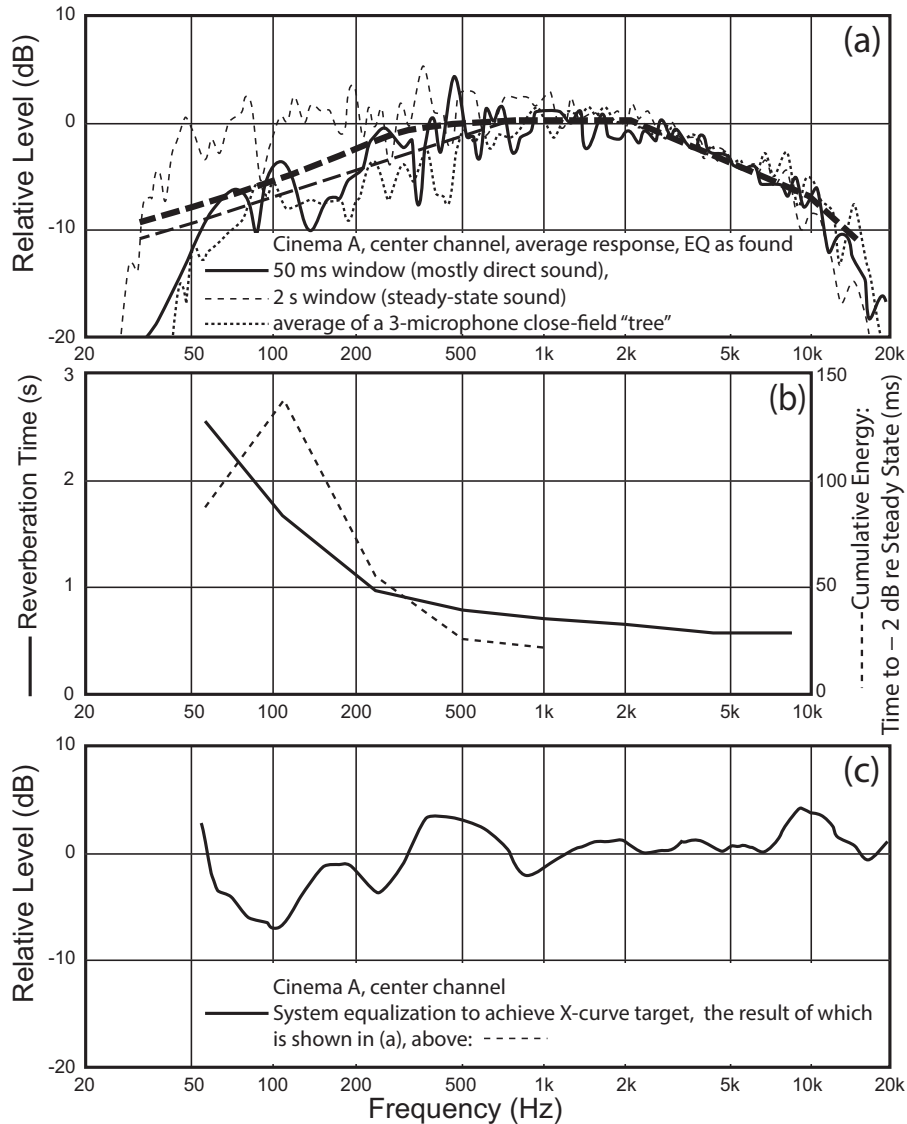
Fig. 8 (a) Cinema A provides an example of a relatively reflective cinema from the SMPTE report [4]. These data show that a system calibrated to the ST 202 X-curve (the thin dashed curve) delivers a direct sound spectrum that is far from flat (the solid curve from time-windowed data at 2/3 distance from the screen and the dotted curve from close-field data). The heavy dashed-line predictions of the bass rolloff from Fig. 7(b) are good fits. Fig. 8(b) Reverberation time (solid line) is high at low frequencies predicting a substantial growth of reflected bass energy after the direct sound arrives at the listening locations. The dotted line is the time required for the energy to accumulate to a sound level 2 dB below the steady-state level (the choice of this criterion was the author's decision). Fig. 8(c) shows the system equalization used to achieve the X-curve calibration.

the loudspeaker directivity and room absorption at middle-to-low frequencies, the direct sound will exhibit varying amounts of bass deficiency. Intuitively, it seems as though the leading edge of a bass transient might lack "substance" or "impact" in some situations. Again, the direct sound arriving at listeners is not flat.

These conditions exist *only* in X-curve calibrated dubbing stages where sound tracks are created and in X-curve calibrated cinemas where they are reproduced—nowhere else in the audio world is it applied. It is part of a closed system, and within it the "circle of confusion" (Fig. 1) should theoretically not be a problem. But it is because speakers

and rooms are not identical throughout the system. Anecdotal evidence indicates that some facilities simply ignore the X-curve requirement, or alter it to meet the owner's desires after calibration. Others maintain two equalizer settings; one for calibration, one for everyday use. The Newell data [32] and the SMPTE B-chain report [4], and some data shown here, provide evidence that inconsistent calibrations are an additional issue.

If mixers and cinema audiences are to hear a neutrally-balanced spectrum from a neutral original recording or microphone pickup, a significant high-frequency boost, at most approximating the inverse of the X-curve rolloff, needs to be incorporated into the soundtrack. This is an additional task for mixers—a subjectively-judged
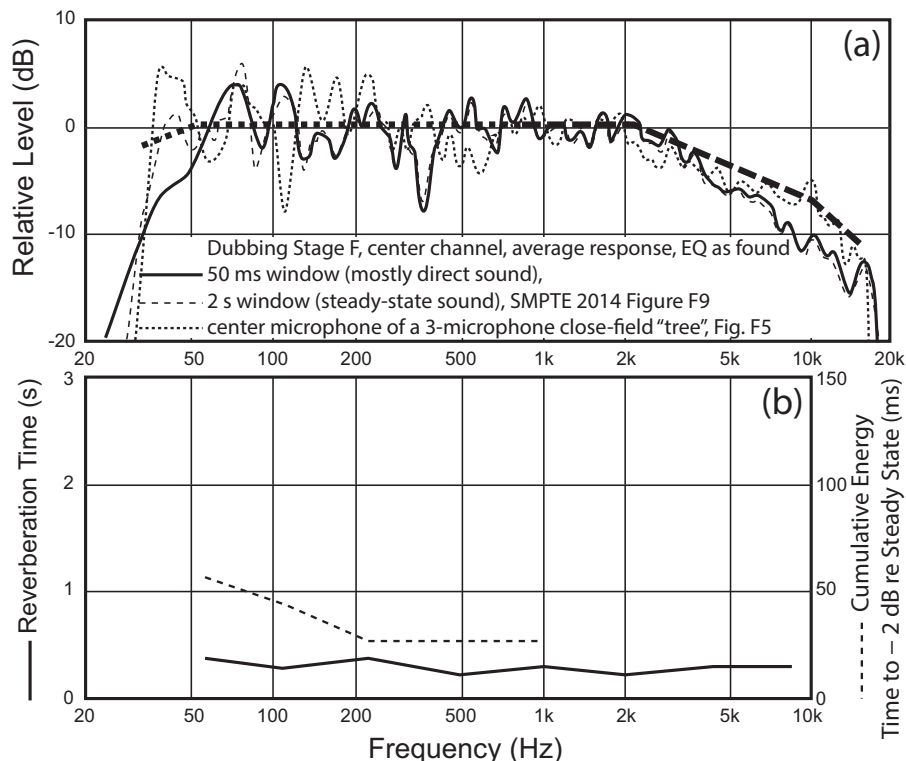
Fig. 9. (a) Using the format of Fig. 8, these data are from Dubbing Stage F [4]. This low RT room is obviously absorbing most of the reflected sound so the direct and steady-state sounds are of comparable level at all frequencies above about 50 Hz. The "dead" cinema prediction from Fig. 7(b) shows good agreement. In Fig. 9(b) it can be seen that the cumulative energy times correlate with the RT values—both are low.

equalization to compensate for a quantified technical problem. A recent survey of film sound mixers indicates that not all dubbing stages employ the full X-curve rolloff (providing satisfaction for the mixers but not the audiences), and not all mixers boost the high frequencies (thereby delivering insufficient high frequencies to audiences) [36]. In fact, from this report and other anecdotal evidence it seems that some experienced mixers accept the rolloff as normal for film sound and do little or no compensation. None of this incorporates the influence of hearing loss, an occupational hazard in the audio industry, [1] section 19.1.2.

It is not clear that the issue of direct-sound bass variation was widely recognized. Yet, the Fielder data [30], Fig. 11, and other evidence shown in this paper indicates that it is large enough not to be ignored, given the powerful influence of bass on overall spectral balance. Mixers very likely manipulate programs to exhibit a pleasing bass balance in the dubbing stages, which will very likely be acoustically different from each other, and different from cinemas. The result is permanently imbedded in the soundtrack mix as part of the art (Fig. 1).
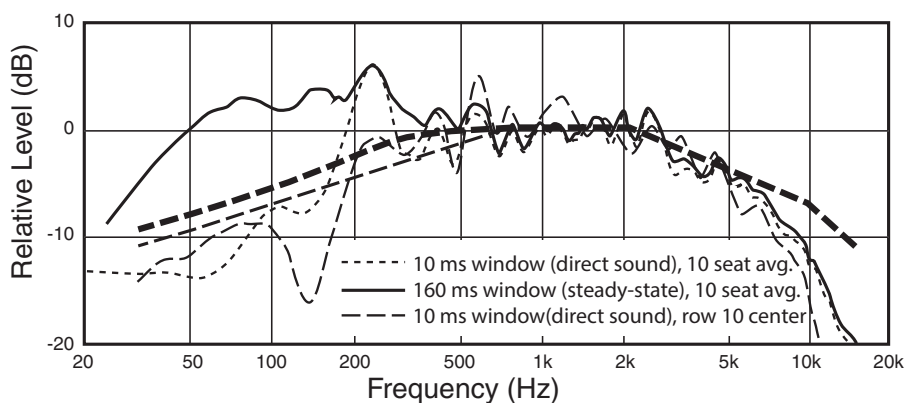

Fig. 10. Steady-state and direct sound data from an X-curve calibrated stadium-seating cinema from Holman [31]. The heavy dashed lines are the "typically" and "highly" reflective predictions from Fig. 7(b).
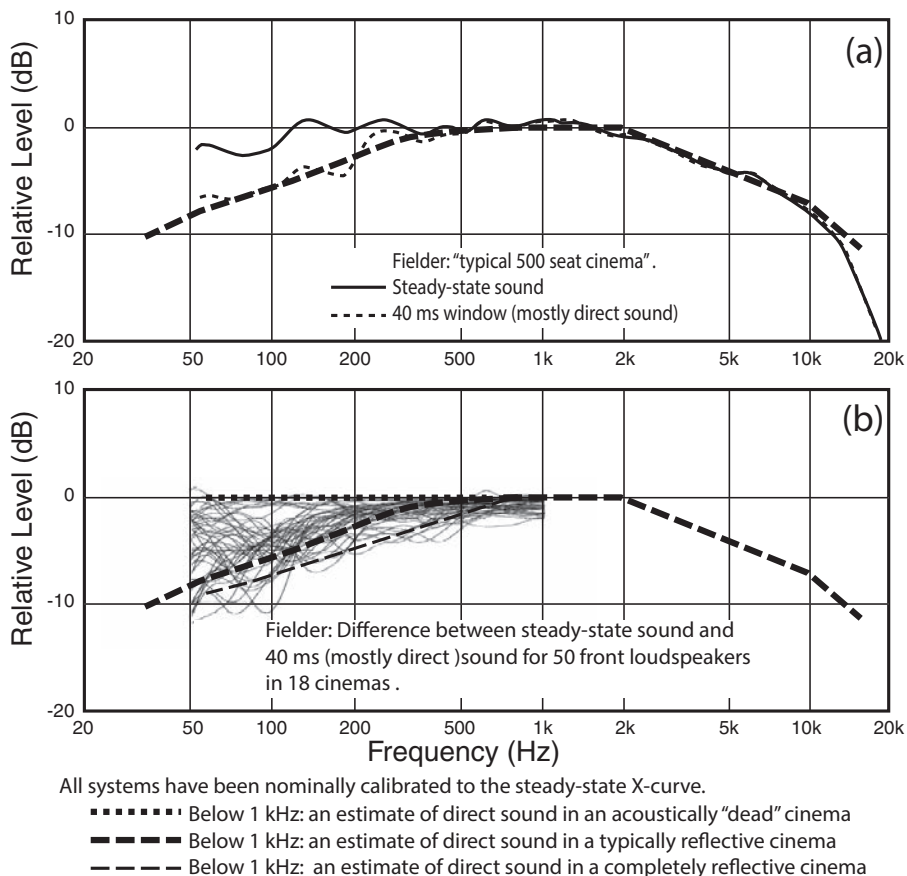
Fig. 11.   (a) Fielder [30] shows steady-state and direct sound for a "typical 500 seat cinema." The heavy dashed line is the typically reflective" prediction from Fig. 7(b). In Fig. 11(b) we see all 50 measurements from which (a) was derived. It is clear that various speakers in various rooms create many options for direct-to-steady-state sound relationships. Below 1 kHz the predictive curves from Fig. 7(b) are superimposed.

Holman proposed a fixed high-frequency-only re-equalization for home reproduction of film sound tracks that had been balanced on X-curve calibrated monitors [37]. Clearly this cannot work reliably in the present situation. In contrast, Allen [35] states: "translation of a film soundtrack from the theatrical version to a consumer release on DVD or TV, is much more likely to need adjustment for the home listening environment (dynamics, limiting, etc.) than for any need for re-equalization." Given the preceding discussion and the large difference between X-curve calibrated facilities and home theaters, this seems most unlikely. A consumer-adjustable "treble" tone control seems appropriate, and given the variations seen in low-frequency levels, a "bass" control as well. The author can attest to having used both to advantage over the years.

Outside of X-curve calibrated film sound facilities, all sound reproduction systems including portable, television, stereo, and home theater audio systems employ nominally flat direct-sound loudspeakers ([1] Figure 2.5, and chapter 18)—even headphones are designed with the expectation of playing program monitored using flat direct-sound loudspeakers [38]. The dubbing stage/cinema/X-curve combination is an isolated system in the audio industry.

3.6 Looking at Alternative Calibration Methods and Targets

From the beginning, loudspeakers intended for sound reproduction have been designed with the goal of a flat on-axis frequency response so that the first sound to arrive is an accurate representation of the spectrum of the recorded sound. Double-blind subjective evaluations of loudspeakers conducted by the author and his colleagues for 35 years have shown consistent preference for those having flat and smooth on-axis frequency response, accompanied by well-behaved off-axis response—i.e., a smoothly changing or constant directivity index ([1] chapter 18). Changing the room has little effect on this observation ([1] section 11.3.1).

Based on this, a generalizable performance standard for sound reproduction systems would require a loudspeaker that radiates a flat, smooth, direct sound, and that has gradually changing or constant directivity. In cinemas this requirement must be met on the audience side of the screen.

A logical action is to simply eliminate the high-frequency rolloff in film sound systems, thereby ensuring compatibility with the rest of the audio world, and use metadata to activate an X-curve rolloff in cinemas for playback of legacy films. Cinemas providing musical concert programs will benefit. High-frequency drivers might need to work harder
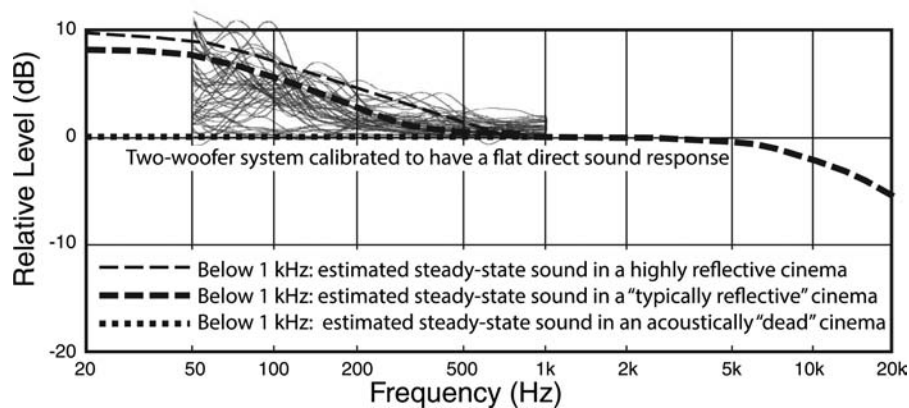
Fig. 12.   The prediction of Fig. 5(b) has been augmented by data from Fielder [30] to show examples of how much the steady-state sound levels can vary using loudspeakers that are flat on-axis, but which may differ in the size of bass units, and used in rooms having differing amounts of low-frequency absorption.

to compensate for lossy screens, but no harder than when reproducing soundtracks in which the mixer has restored neutral spectral balance by fully compensating for the X-curve rolloff. In practice, this may or may not be done [36]. There are also concerns about compression driver overload in cinemas with marginal loudspeaker systems, but the preceding statement also applies so this is not a new concern.

Data on screen loss has long existed, e.g., [34], and has recently been revisited [39–42]. An examination of the spectral/temporal content of demanding program indicated that power output requirements for compression drivers would be little changed and within the capabilities of contemporary products [43]. Screen materials with higher acoustical transparency continue to be developed and compression drivers continue to improve. Existing sound systems with limited output capabilities can be protected by appropriately adjusted on-site high-frequency compression/limiting algorithms, leaving the majority of the sound untouched. If a new performance target is set, the technology will follow if it is necessary.

The situation at low frequencies is very different because listeners are in a stronger reflected sound field. If we choose a flat direct-sound criterion, what does it mean for the sound fields in rooms? Fielder's data on the difference between the direct and steady-state sound fields created by 50 loudspeakers in 18 cinemas provides this insight [30]. Fig. 12 shows the Fielder data inverted and superimposed on Fig. 4(b) to illustrate how much the steady-state sound field varies if the loudspeakers have flat on-axis (direct sound) frequency responses. For most circumstances there will be a rising bass response and the slope of the tilt varies. Only outdoors or in very dead rooms will the difference not exist. If a steady-state room curve is to be employed as a target, is a tilted curve below about 1 kHz what is needed? If so, what tilt?

4   SMALL VENUE SOUND SYSTEMS

Rooms of a size appropriate for stereo or TV entertainment, home theaters, home studios, and recording control

rooms do not require massive, highly directional loudspeakers. The low-frequency sources are smaller, exhibiting lower directivity indexes, Fig. 2(b). Mid and high frequencies tend to be radiated by wider-dispersion horns or small cones and domes. These wide dispersion sources generate more reflected sounds over a wider frequency range than is seen in cinemas. Consequently, as was seen in the example of Fig. 4, predictions of acoustical events must change.

Employing the logic used in Fig. 5(b) for these smaller loudspeakers, Fig. 13(a) begins with seven inverted DIs—estimates of total radiated sound power for loudspeakers adjusted to flat on-axis frequency responses. These seven loudspeakers had all achieved very high subjective ratings in double-blind, positional-substitution listening tests (see [44] and [1] section 17.7). All had uncommonly flat and smooth on-axis responses as well as similar smoothly-changing DIs, confirming the excellence of the designs and explaining the high subjective ratings [45, 46]. The author has estimated an average curve through the collection, extrapolated to lower frequencies, assuming a 6 to 7 dB maximum DI, as seen in Fig. 5(b). In the display 10 dB below this, this curve provides an upper limit to anticipated steady-state room curves—assuming that all of the radiated energy is delivered to the listening positions by reflections. The horizontal line is the "perfect" direct sound, which is the only sound in rooms with negligible reflections. In normally reflective small rooms the direct sound is strongly influential only at the very highest frequencies, Fig. 4. Therefore this line is adjusted to meet the sound power curve at the high-frequency end of the scale, around 8 to 10 kHz. The shaded area between these is where steady-state curves in "real" rooms might fall. The author has drawn an intermediate heavy dashed line that could represent a "typically reflective" room.

Fig. 13(b) shows six-seat spatially averaged high-resolution steady-state room curves for three highly-rated loudspeakers and the result of a very large survey of home-theater satellite loudspeakers (80 Hz high-pass filtered for multichannel bass management) [47]. At low frequencies the huge variations caused by room standing waves are
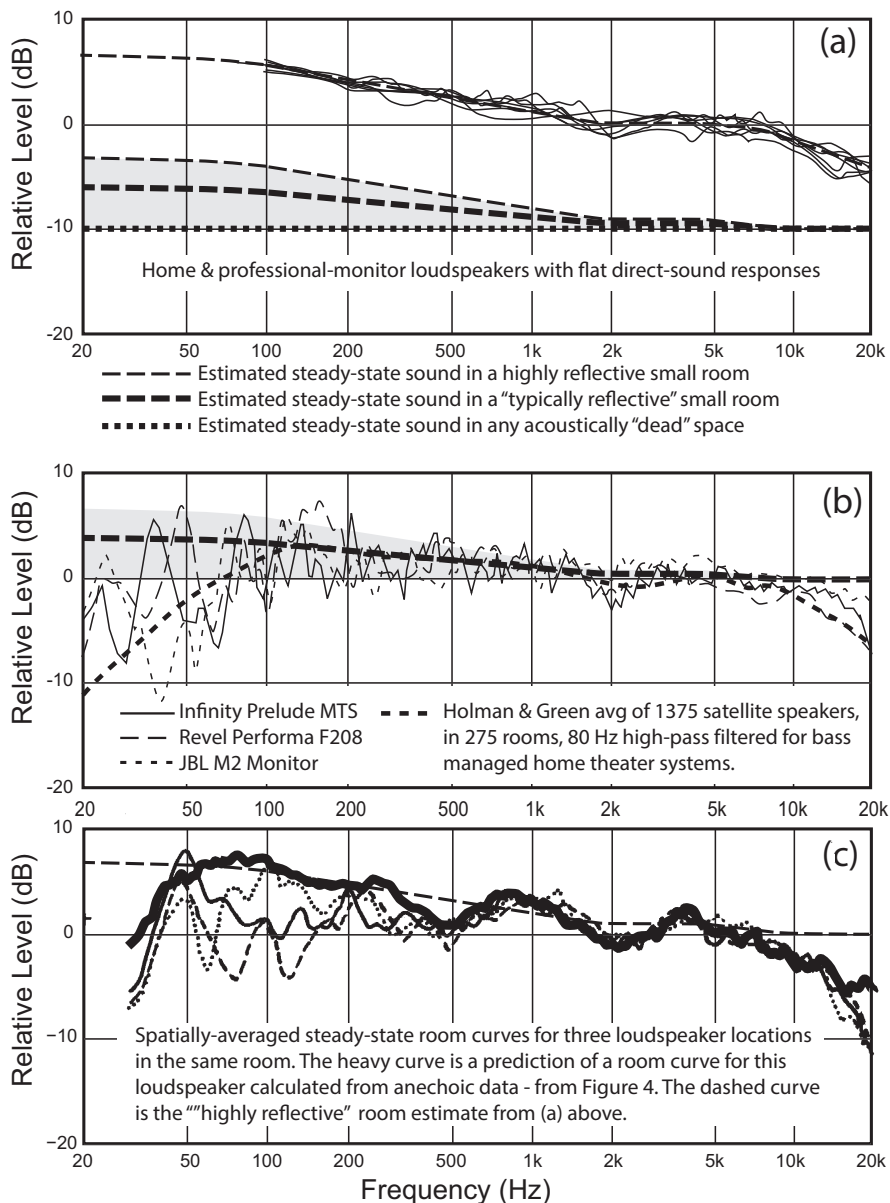
Fig. 13 (a) Inverted directivity indices for the three loudspeakers in (b) and four others of comparable sound quality are superimposed. An author-created average curve (light-dashed) is shown. Below this is a prediction of the range of steady-state room curves that might occur in rooms of different acoustical properties. Fig. 13(b) shows six-seat average room curves for three highly rated loudspeakers and the average of a large survey. The prediction of (a) is superimposed. Fig. 13(c) shows a comparison between a calculated prediction for a specific, problematic, loudspeaker and averaged room curves for that loudspeaker at three different locations in the same room (From [1] and [9, part 2]).

very evident. Superimposing the shaded area from (a) along with the "typical room" prediction reveals good agreement above about 300 Hz even without the benefit of spectral smoothing.

Fig. 13(c) is from the author's early work, shown in Fig. 4, where it was demonstrated that a calculated prediction agrees very well with the in-room measurements above about 300 Hz. Room mode problems dominate at lower frequencies. The predicted curve from (a) is superimposed. The large undulations from 300 Hz to about 5 kHz match the inconsistent off-axis performance of this loudspeaker shown in Fig. 4(a). DI varied significantly with frequency. Even though the on-axis response was very flat, subjective evaluations indicated significant coloration.

Early-reflected energy influences steady-state room responses at frequencies up to 5 kHz or higher, which is a major change from the sound field in cinemas that transitions to direct-sound dominance above about 600 Hz. The wider dispersion of the home theater loudspeakers is clearly a major factor. Constant or at least smoothly changing directivity is important so that reflected energy does not contradict the spectrum of (and sound quality signaled by) the direct sound. At low frequencies the calculated prediction falls close to the upper (highly reflective) limit of the shaded area. The disparity at high frequencies is possibly due to the large directional tweeter in this unit that discriminated against listening locations away from the prime axis, as is indicated in Fig. 4(a).
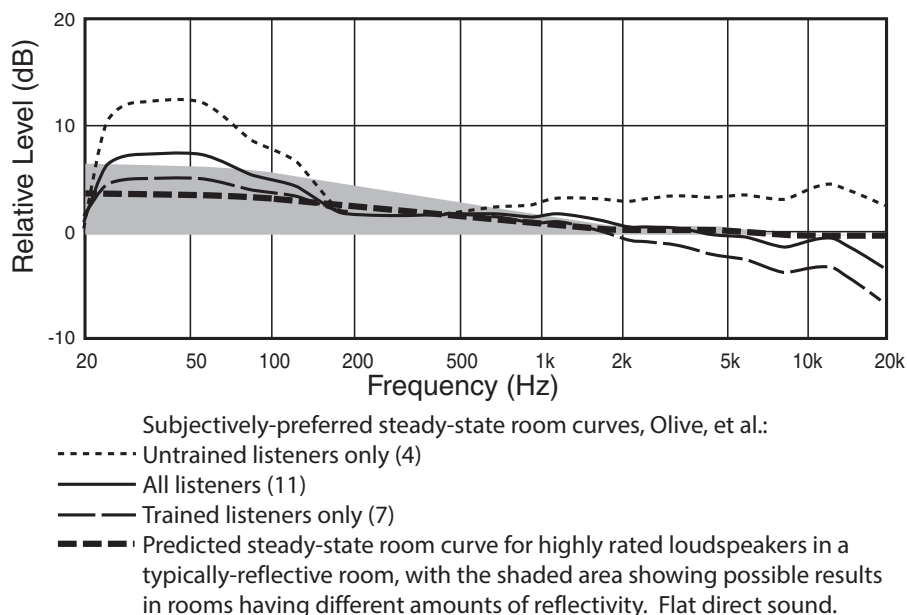
Fig. 14. Subjectively preferred steady-state room curve targets in a typical domestic listening room [49], from Olive et al. [48]. The prediction of Fig. 13(a) is superimposed.

Summarizing, it is well known and is confirmed here that the foremost distinction of small rooms is the dominance of room modes/standing waves at frequencies below about 200 to 300 Hz, the "transition frequency" (the Schroeder frequency in large reflective rooms) discussed in [2], section 7. Above this frequency, anechoic data can be used to anticipate steady-state room curves with useful accuracy. Early reflected sound is influential up to much higher frequencies than in cinemas making frequency-dependent directivity data very important information. In loudspeaker design the anechoic evidence of poorly controlled off-axis response is seen in the radiated sound power and the steady-state room curves and is heard by two ears and a brain. Equalization cannot repair faulty loudspeaker directivity.

4.1 Some Evidence of Listener Preferences in Home Theaters

Over the years a few investigators have attempted to identify advantageous room curve targets for small rooms. However the studies that the author is aware of have been compromised by a lack of adequate loudspeaker measurements and/or information about the room acoustics. No double-blind listening tests appear to have been done so there are no trustworthy subjective evaluations. Consequently, the resulting targets can be challenged.

Research by Olive et al. [48] was distinctive in that the loudspeaker used was anechoically characterized, the room described [49], and high-resolution room curves measured. In the double-blind tests, listeners made bass and treble balance adjustments to a loudspeaker that had been equalized to a flat smooth steady-state room curve. The loudspeaker had previously received high ratings in independent double-blind comparison tests, without equalization. Three tests

were done, with the bass or treble adjusted separately with the other parameter randomly fixed, and a test in which both controls were available, starting from random settings. It was a classic method-of-adjustment experiment. For each program selection, listeners made adjustments to yield the most preferred result.

In Fig. 14 the author has modified the original data to separately show the result of evaluations by trained and untrained listeners. This is compared to the small room prediction from Fig. 13(a). The "all listeners" average curve is close to the predicted target, except at low frequencies where it is apparent that the strongly expressed preferences of inexperienced listeners significantly elevated the average curve. In fact, the target variations at both ends of the spectrum are substantial, with untrained listeners simply choosing "more of everything." An unanswered question is whether this was related to overall loudness—more research is needed. However, most of us have seen evidence of such more-bass, more-treble listener preferences in the "as found" tone control settings in numerous rental and loaner cars.

More data would be enlightening, but this amount is sufficient to indicate that a single target curve is not likely to satisfy all listeners. Add to this the program variations created by the "circle of confusion" and there is a strong argument for incorporating easily accessible bass and treble tone controls in playback equipment. The first task for such controls would be to allow users to optimize the spectral balance of their loudspeakers in their rooms, and, on an ongoing basis, to compensate for spectral imbalances as they appear in movies and music.

The attenuated high frequencies preferred by the trained listeners stands in contrast to the preferences exhibited by those same listeners in numerous double-blind
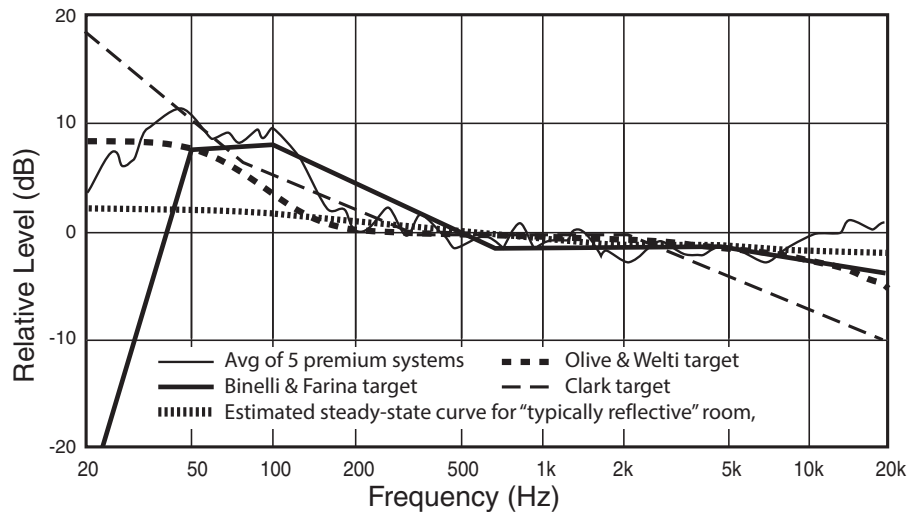
Fig. 15. Three experimentally determined targets for car audio systems compared to the average of premium audio systems in five contemporary luxury automobiles. All are compared to the average steady-state room curve in a "typically reflective" small room shown in Fig. 13(a).

multiple-comparison loudspeaker evaluations. In those tests, it is the flat on-axis loudspeakers that are most highly rated (those that perform close to the predicted curve in Fig. 14). Is this a consequence of the different experimental methods: the different listener tasks? In one, listeners adjusted the bass and/or treble balance in a single loudspeaker model; in the other they rated spectral balances and other attributes in randomized comparisons of different products. It is a subtle but important difference awaiting an explanation.

A significant observation is that the results indicate a general preference for a steady-state spectrum that rises toward the low frequencies. In previous experiments Olive et al. [50] compared five different room correction products, finding that the most preferred product followed a curve similar to the "all listeners" curve shown here—those having less bass were not rated as highly. The author has followed developments in "room correction/equalization" products over the years. To date there is some evidence of agreement that the target curve should exhibit a downward slope over at least a portion of the frequency range. Most combine flat, tilted, and curved portions. None have yet involved a target that rises towards high frequencies. In some schemes the customer is able to choose among several target options—the trial and error approach.

4.2 Car Audio

Consumers spend considerable time listening to music in their cars through audio systems that nowadays can be quite good. These systems can only be measured after installation in the acoustically complex automobile cabin. Fig. 15 shows results from U.S. subjective evaluations by Olive and Welti [51] and Clark [52], an Italian study by Binelli and Farina [53], and measurements in five multi-national luxury cars. All boost the bass to compete with the substantial road, aerodynamic, and mechanical noise at low frequencies (in some vehicles this can vary with speed and/or background noise). There is substantial agreement through the middle

frequencies, including the predicted small-room curve from Fig. 13(a). Apparently, designers wanted cars to sound like good home systems. Differences in how the in-cabin automotive data are gathered lead to some disagreements in curve shape at very high frequencies. The perceived sounds are likely to be more similar than these measurements indicate.

5 CINEMAS REVISITED

The SMPTE report [4] documents performance of some important west-coast USA cinemas and dubbing stages, all of which had been calibrated to the X-curve. Fig. 16 shows steady-state measurements of these facilities at the reference locations (a) and averaged over all measurement locations (b). Both are compared to the objectives set by SMPTE ST 202 [21]. The author has attempted a visual normalization through the middle frequencies.

The results show significant variations at all frequencies, especially in the single-seat curves (a). These contain undulations suggesting the effects of acoustical interference, probably including seat interactions in the middle frequencies. The latter is the likely consequence of placing the microphones at ear height, which in many cinemas is close to the high seat backs.

There is a trend toward boosted bass—the curves drift upward below about 200 Hz—and a trend to attenuated high frequencies in which the curves drift towards the lower tolerance limit. Both of these deviations are above audible thresholds for broadband (low-Q) spectral deviations [25]. These same kinds of deviations from the X-curve target can be seen in Fig. 17 from Newell et al. [32] who surveyed nine cinemas in Europe and in the steady-state Holman data in Fig. 10. One wonders if a final subjective tweaking led to them—did they simply sound better? Whatever the causes, the results are spectral balances that audibly deviate from what was intended in the standard.
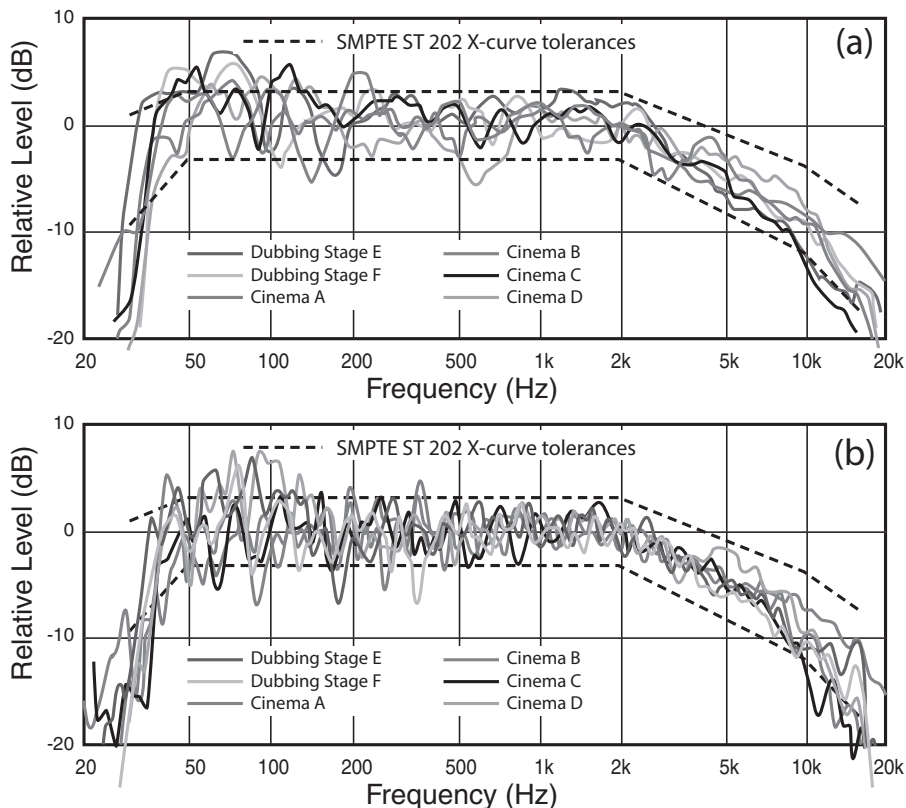
Fig. 16. Steady-state room curves for four cinemas and two dubbing stages from the SMPTE 2014 report [4] showing measurements at the reference locations (a) and averaged over all measurement locations (b), compared to the SMPTE ST 202 tolerances.

5.1 A Test of the Test

A study by Gedemer [54] provides data that differs from the other cinema measurements shown here in that the same loudspeaker was used in all venues, the anechoic performance of the loudspeaker was known beforehand, and no room equalization was used. The object was to compare steady-state cinema curves with the anechoic loudspeaker data, looking for useful relationships. The sound source was a studio monitor loudspeaker, a JBL Professional M2, with full anechoic documentation of the kind described in [1] Figure 18.6 and [55] (look ahead to Fig. 20). It was set up in front of the screen, at the center location, in several film sound venues. Measurements were made at many seating locations at ear heights and at differing microphone heights.
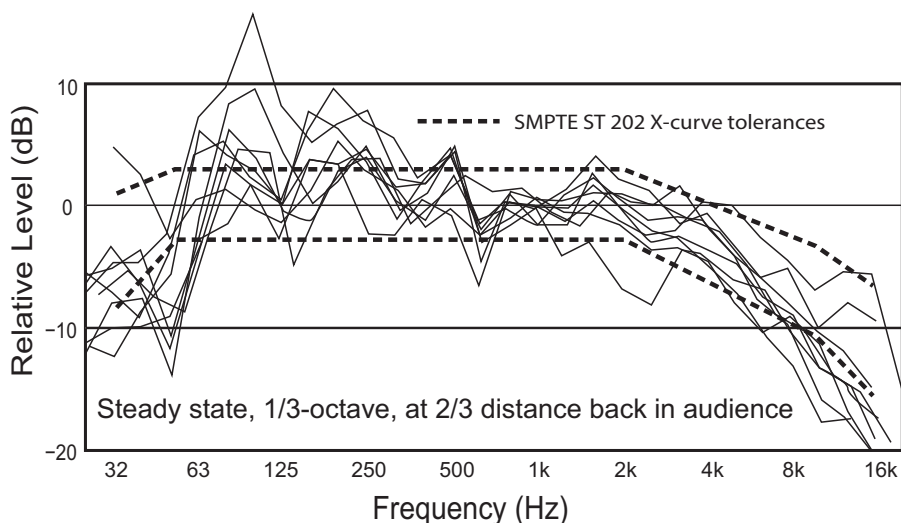

Fig. 17. 1/3-octave steady-state room curves measured at the reference location in nine European cinemas calibrated to the X-curve. From Newell et al. [32]. The author created these normalized point-to-point curves from the original histogram data.
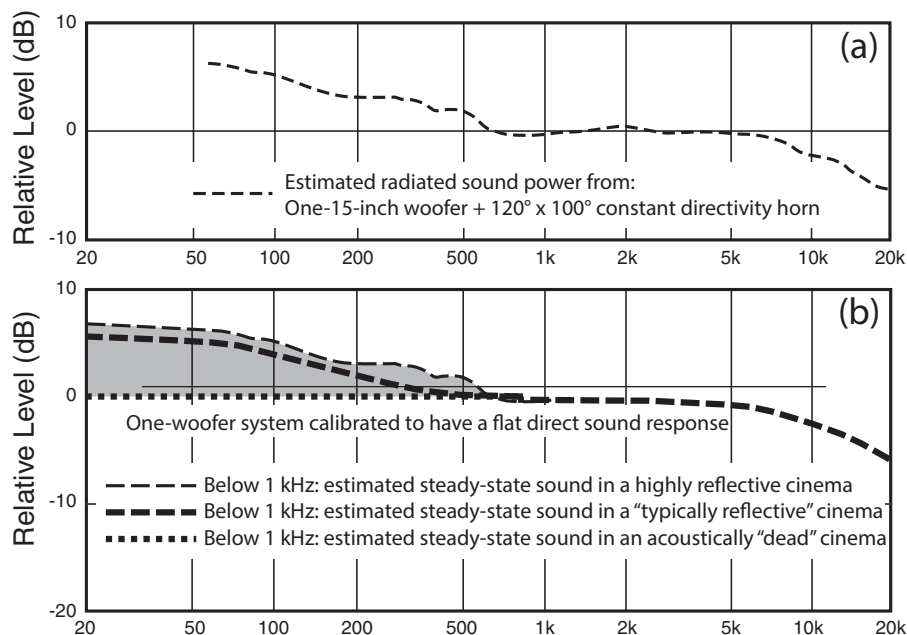
Fig. 18 (a) The inverted directivity index for the JBL M2, which is an estimate of the spectrum of the sound power output when it radiates a flat direct sound (which it was designed to do). Fig. 18(b) shows how this radiated energy might be modified by a "typically reflective" cinema (explained in the text associated with Fig. 5(b)). Also shown is the prediction for an acoustically "dead" venue. Other rooms will fall within the shaded area. The high frequency rolloff is air attenuation, as Fig. 5(b).

Using the same process described in Fig. 5(b) for the two-woofer cinema loudspeaker, Fig. 18 shows a prediction for the steady-state sound that might be measured using this one-woofer loudspeaker in venues of different acoustical properties. In the real world, steady-state responses may fall anywhere in the shaded area, depending on the room reflectivity.

The curves shown in Fig. 19(a) are averages of steady-state measurements made at every second seat in each of the facilities, except for cinema G where 109 locations were used. Above about 1 kHz the curves agree very well, indicating minimal room effects—the listeners are in the direct sound field. The small variation at the highest frequencies is due to different propagation distances in the very different sizes of venues. The curves were normalized to this frequency range. Below about 150 Hz the variations are of a kind expected for differing amounts of reflected energy, adjacent boundary effects and, in the smaller venues, room modes. The irregular behavior in the 150 to 800 Hz range was troubling. The well-known seat dip effect is a probable factor, but the variations are more complex than those seen in concert halls. With high-backed and stadium seating in some venues, additional seat interaction effects were suspected (the microphones were 1.0 to 1.2 m above the floor).

Fig. 19(b) shows that when the microphones were elevated to heights randomized over the range 1.3 to 2 m in three of the venues the curves congregate more closely to each other and to the prediction from Fig. 18(b). The obvious outliers above 300–400 Hz are from measurements made at ear height.

Cinemas exhibit an elaborated version of the concert-hall seat-dip effect, possibly because the cinema sound fields are considerably less reflective than those in concert halls. There is reason to doubt that these non-minimum-phase acoustical interference effects can be treated by equalization. There are also reasons to wonder how audibly important they are—think of the time one has spent in high-backed chairs and recliners, with a large acoustical obstacle behind the head. Once stationary in a chair, listeners usually are unaware of, or attribute innocent spatial descriptors to, measured variations suggesting repugnant comb-filtering phenomena. Thus, a topic for psychoacoustic investigation could be to determine whether ear-level measurements convey information of unique value—a "stand-up/sit-down" listening comparison. If the result is that it is sufficient to ensure the integrity of the direct sound, this can be ascertained using microphones in the propagation path, located away from the seats, or directly from loudspeaker and screen data.

5.2 The Effects of Room Size

The loudspeaker used in the Gedemer cinema experiments was also measured in a typical domestic listening room/home theater environment [49]. There it yielded a steady-state room curve very similar to two other subjectively highly rated cone-and-dome domestic loudspeakers, Fig. 13(b). The home-theater room curve for this loudspeaker is shown superimposed on the cinema data in Fig. 19(c). Above about 2 kHz the curves differ mainly in the reduced air attenuation at the short listening distances in the home theater. At low frequencies small-room standing waves dominate results.

Above about 150 Hz the cinema curves and the home theater curve congregate around a similar central tendency.
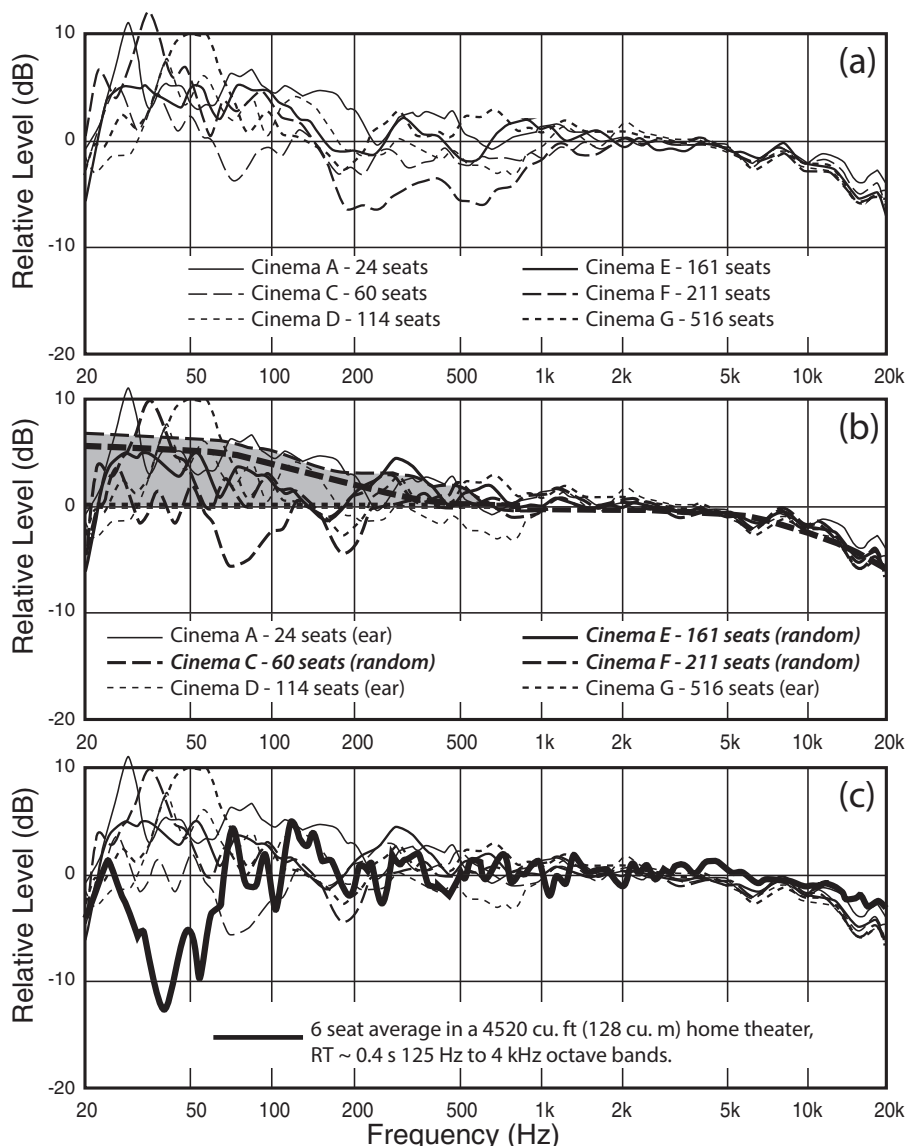
Fig. 19 A loudspeaker known to have a very flat on-axis frequency response radiates directly into six different film-sound venues. (a) shows the result of microphones at ear level. (b) shows the results when three of the cinemas are represented by measurements at random microphone heights. The prediction from Fig. 18(b) is superimposed. (c) shows these steady-state cinema measurements with a six-seat average of the same loudspeaker measured in a typical domestic listening room/home theater.

Even with a seat count that ranges through 6, 24, 60, 114, 161, 211, and 516 there is no obvious pattern of change beyond the small difference in air attenuation in the direct sound, which is a function of listening distance. This is in contrast to the advice of the SMPTE and ISO recommendations [21, 22] where, above 2 kHz, different high frequency target tilts are specified for rooms of different seating capacities. The large differences are at low frequencies where room reflectivity is a factor in cinemas and standing waves the dominating influence in the home theater. The unavoidable standing wave effects in small rooms are distinctive for different rooms, loudspeaker, and listener locations and so defy simple standardization, although, as mentioned earlier in Sec. 2.4, there are effective methods of attenuating their effects.

In summary, there is no evidence of need for an audience-size scaling correction in these measured data. The under-lying assumption is that this loudspeaker, which has been subjectively praised in small room evaluations, would be similarly praised in comparable double-blind subjective evaluations in the cinema spaces. There is no reason to think that this would not be the case, nevertheless a test of this assumption remains to be done.

6 CHARACTERIZING LOUDSPEAKER PERFORMANCE

Obviously, high quality sound reproduction requires good loudspeakers. But what, in technical terms, describes a good loudspeaker? Subjective evaluations done in small rooms show that highly rated loudspeakers exhibit good performance in anechoic measurements and vice versa [1, 2, 9, 45, 46]. We lack correspondingly thorough evaluations
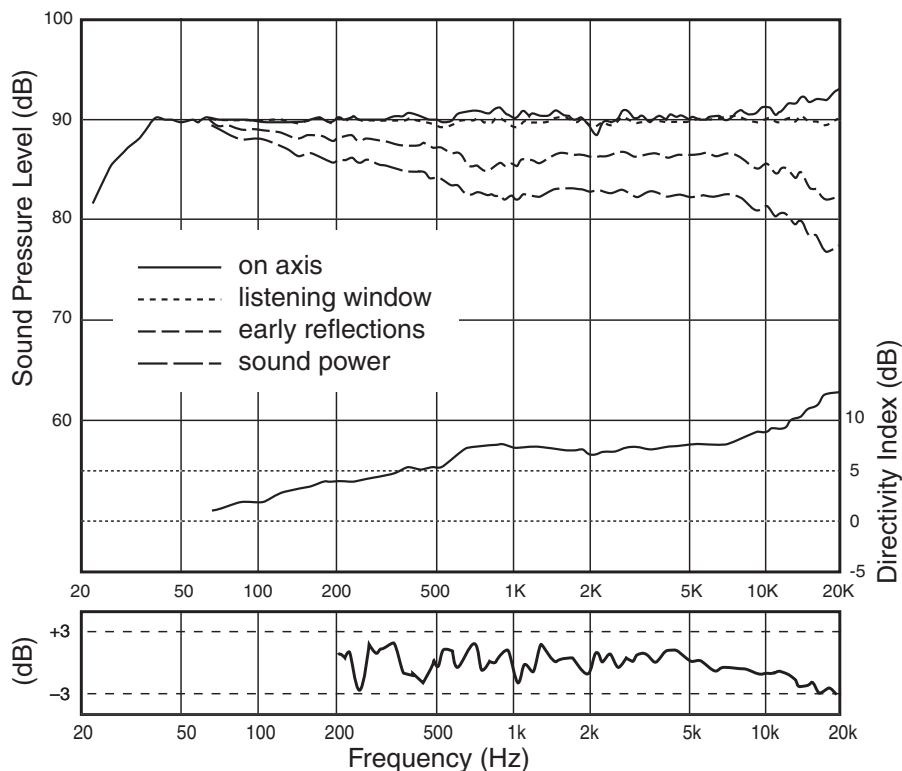
Fig. 20. An example of curves calculated from a data set of 70 1/20-octave anechoic measurements. Informally called the "spinorama," it contains sufficient information to anticipate the basic form of a room curve and to estimate subjective ratings in double-blind listening tests [45, 46]. The smoothness and uniformity of the curves are clues to excellence. This is an example of a very good floor-standing loudspeaker with dedicated digital electronics adjusted here to maximize performance within the "listening window," estimating the direct sound received by members of an audience.

in large venues but there is no reason to expect a different result.

So far we have considered loudspeaker performance from a simplistic perspective: the on-axis amplitude response, sound power output, and directivity index, the difference between the two. Loudspeaker sound quality is related to these data to be sure but also to the presence of timbre-modifying resonances. In fact, eliminating resonances from loudspeakers is essential to avoid timbral colorations that are monotonously added to all reproduced sounds. It is important, therefore, to be able to identify and evaluate the audibility of resonances. For this more information is needed [25].

6.1 Subjective and Objective Loudspeaker Measurements

Quantifying loudspeaker performance requires an anechoic space large enough to permit measurements in the acoustical far field—at least 2 m for domestic and monitor loudspeakers and much more for large cinema and sound reinforcement systems. Time-windowed measurements extend this capability to reflective spaces in which reflections can be excluded but with compromises in frequency resolution at low frequencies. Near-field scanning can compile enough spatially-sampled data to permit the prediction of 3D performance in the far field providing another alterna-

tive [56, 57]. The goal is to collect enough data about the sound radiated by the loudspeaker to predict what might arrive at a listener's ears in a room.

Fig. 20 shows a family of curves, derived from 70 1/20-octave resolution amplitude-response measurements at 10° increments on horizontal and vertical axes, informally called the "spinorama" ([1] section 18.2.2, ANSI/CEA-2034 [55], and [58]). The starting point, the reference, is an on-axis curve. The "listening window" curve is a spatial average of measurements within a ±30° horizontal by ±10° vertical window, intended to describe the direct sound for an audience. The "early reflections" curve is a spatial average of off-axis sounds that would contribute to the early reflections in typical home theaters (this would be modified for significantly different spaces). The sound power curve is shown as an amplitude response, having been normalized. The directivity index shown here is the difference between the sound power curve and the listening window curve, the latter being preferred to the traditional on-axis curve because it is less susceptible to small acoustical interference effects that often appear in the single-point on-axis data.

Below this is shown a six-seat averaged steady-state room curve measured in a good, typical, home theater situation [49], from Fig. 19(c). It is shown in 1/12-octave resolution to reveal detailed irregularities. As has been discussed, the low-frequency performance is dominated by standing waves—a room-dependent factor—so the curve is truncated

at 200 Hz. Fluctuations fall comfortably within a ±3 dB tolerance; traditional 1/3-octave smoothing would reduce the undulations even further.

In double-blind subjective evaluations in a typical room, loudspeakers exhibiting this caliber of anechoic performance typically are awarded ratings of 7 or more on a scale of 10 (e.g., see [1] Figure 18.14)—experienced listeners are reluctant to award much higher scores possibly because the evaluation also includes the recordings with their idiosyncrasies (Fig. 1). This has been the case from the author's earliest experiments [8, 9], others described in [1] chapters 17, 18, and 20, and powerfully reinforced by Olive's subjective/objective correlations [45, 46]. No "room equalization" is required except what may be appropriate to deal with site-dependent low-frequency adjacent-boundary effects and standing waves (see Sec. 2.4).

The importance of loudspeaker performance was persuasively shown in an elaborate test in which three closely-rated loudspeakers were evaluated in four very different rooms. The double-blind subjective ratings were essentially unchanged as the test was moved from room to room [59] (this is elaborated on in [2] section 6). Listeners were able to separate the timbral signatures of the loudspeakers from the timbral cues added by the rooms. Making the result even more remarkable, in evaluations of loudspeakers in different rooms, listeners were required to isolate and place into a specific context, the timbral cues of recordings. This can be interpreted as a case of perceptual streaming, the result of, as Bregman describes it: auditory scene analysis [60]. Clearly there is significant perceptual analysis, deduction, and adaptation going on. This cognitive activity apparently happens subconsciously in everyday listening; we hear and recognize the timbral identities and subtleties of the same sound sources but in different rooms. Two ears and a brain are massively more analytical and adaptable than an omnidirectional microphone and an analyzer.

To provide useful perspective, Fig. 21 shows similar data on a loudspeaker with serious problems [61]. The 1/20-octave resolution curves reveal medium-Q resonances at 600 Hz, 1200 Hz, and 16 kHz and a substantial low-Q resonance around 3.5 kHz. Because loudspeaker transducers are minimum-phase devices, we know that ringing at these frequencies will occur, but, as was found in [25], it is the spectral bump that is the most reliable indicator of audibility. The low-Q resonance will be the dominant audible problem because its amplitude is many times the detection threshold [25]. The other resonances, though, are likely above the threshold of detection for complex music. A crossover acoustical interference dip around 1.5 kHz is visible in the on-axis and other front-hemisphere measurements but is much attenuated in the sound power curve. It therefore shows up in the directivity-index curve.

Because they are minimum-phase phenomena in transducers the resonances can be attenuated by using matched parametric filters, thereby correcting both the amplitude and time-domain problems. When DSP is available in the signal path, this is a powerful use of equalization, but it can only be done based on high-resolution anechoic data allowing center frequency and Q to be identified. The non-

minimum-phase interference dip cannot be corrected by equalization. From low to middle frequencies the steady-state room curve is clearly influenced by the radiated sound power as has been seen in earlier examples. In double-blind subjective evaluations this loudspeaker, supplemented by an excellent subwoofer, scored 3.9/10—not good.

In this anechoic data one sees audible resonances that are not evident in the steady-state room curve. The dominant low-Q bump is seen because it contributes substantially to the shape of the sound power curve. Room equalization can therefore attenuate its effect, but the presence and significance of the others are masked. The point of this is to show that all of the information is in the anechoic data and only some of it is seen in the steady-state room curve. The conclusion is that it is strongly advantageous to begin with comprehensive high-resolution anechoic data from which problems in a room can be anticipated, at least above the transition/Schroeder frequency.

The generally well-behaved directivity suggests that the sound of this loudspeaker can be improved by equalization. However, the improvement is limited because of the medium-Q resonances that are not clearly revealed in the room curve and, thus, would not be addressed. If a loudspeaker exhibits a similar undulation in the room curve because of frequency-dependent directivity, as seen in Fig. 4, even such a limited remedy would not be available. The solution would be to substitute a better loudspeaker, a fact that would have been anticipated if comprehensive anechoic data had been available. The common specifications on loudspeakers are significantly uninformative. The same dilemma exists when frequency-dependent absorption in the room affects the steady-state room curve—the effect is seen but the cause is unknown and the remedy not obvious.

A final point: these data show that a ±3 dB tolerance on a steady-state room curve is merely a license to make sound. Determining whether it is good sound or not requires additional data. More research of this kind is necessary.

7 A "UNIFIED" APPROACH TO SOUND PRODUCTION AND REPRODUCTION

Nothing about the detailed acoustics of the spaces we live and listen in is constant, yet humans cope remarkably well, adapting to the varying sound fields while maintaining a stable perceptual impression of the sound sources. Changing the venue does not fundamentally alter the timbral character of voices and musical instruments. They are merely those voices and instruments in different venues. Sound production and reproduction are not treated differently. This is normal—natural acoustics at work. We seem to accept the varying bass boosts resulting from cumulative reflected sound as part of the character of the rooms, not so much as a description of the sound source. If a symphony orchestra sounds "thin," it is blamed on the hall, not the instruments or musicians. If a familiar male voice sounds unusually "rich," it is attributed to the character of the space within which the sound is communicated.

For the entire history of sound reproduction a flat and smooth on-axis frequency response has been the
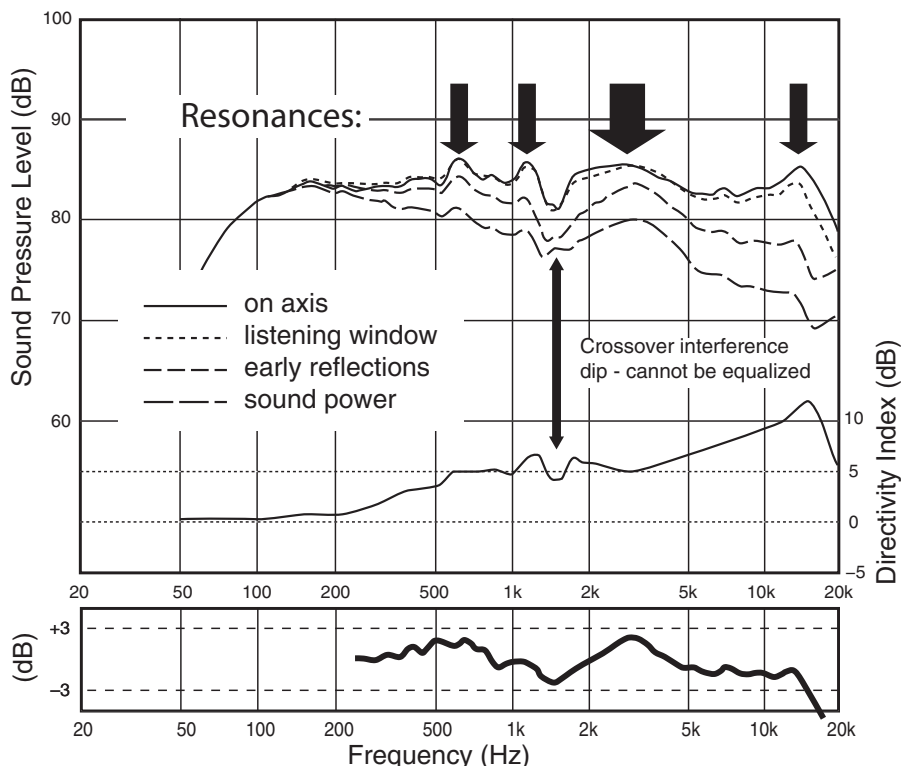
Fig. 21.   The upper box shows "spinorama" data on a flawed loudspeaker showing resonance peaks and an interference dip. The lower box shows a six-seat spatially averaged steady-state room curve. Data from [61].

performance target for loudspeaker designers. The logic was that whatever the spectrum of the original captured sound was, the loudspeaker should reproduce it as faithfully as possible. Later it was found that maintaining similarly smooth performance off axis resulted in even higher subjective ratings [1, 9, 45, 46]. Fig. 13(c) provides persuasive evidence of what can go wrong when loudspeaker engineers neglect this factor.

7.1 Is There a Single Performance Target for all Sound Reproduction Purposes?

In discussions to this point it has been indicated that measurements in a room are, by themselves, insufficient data upon which to base a prediction of sound quality. However, they are not without value, especially if certain information about the loudspeaker is available. The normal practice in interpreting an in-room measurement is to compare the measured data to a target curve, supposedly describing a desirable outcome. We know that there is the possibility of many such curves, depending on the time window of the measurement. Very short and very long time windows give us indications of, respectively, the direct sound and of the steady-state sound field after reflected sounds have made their contributions.

Fig. 22 brings together the prediction curves from this paper. Measurements in real venues have been in good agreement. In Fig. 22(a), (b), and (c) the shaded areas show the estimated range within which steady-state room curves may fall, depending on the reflectivity of the venue—high

reflectivity at the top, low reflectivity at the bottom. The loudspeakers radiated flat direct sounds in all three cases. The heavy dashed curves are the author's estimates for "typically reflective" rooms in each of the categories. Curve (d) is the result of an experiment in which listeners adjusted bass and treble controls to achieve a satisfying spectral balance with a selection of programs. Curve (e) is current SMPTE ST 202 target performance for steady-state sounds at the 2/3 distance in cinemas and dubbing stages [21]. The target sets a flat limit for the bass rise—i.e., the steady-state. Therefore, in this case, the gray area shows the range of variation to be expected in direct sound, as shown in Fig. 12(b).

To this must be added the LFE channel, a separate bass channel providing additional output for special effects at frequencies that overlap the bass region of the screen channels. There are no standard practices for dealing with the acoustical interactions created by the overlap [4]. It matters not whether the LFE sounds are different from or the same as the sounds in the screen channels; the mere coexistence of sound from both loudspeakers results in acoustical interference, the audible effect of which depends on the physical locations of the loudspeakers radiating the energy, and the locations of listeners in the room.

In this respect, home theaters using bass management, including complementary low-and high-pass filters, are in much better control of low-frequency acoustical events. Cinema sound could benefit from embracing this system in an elaborated form if necessary to accommodate the specific issues of multichannel audio in large venues.
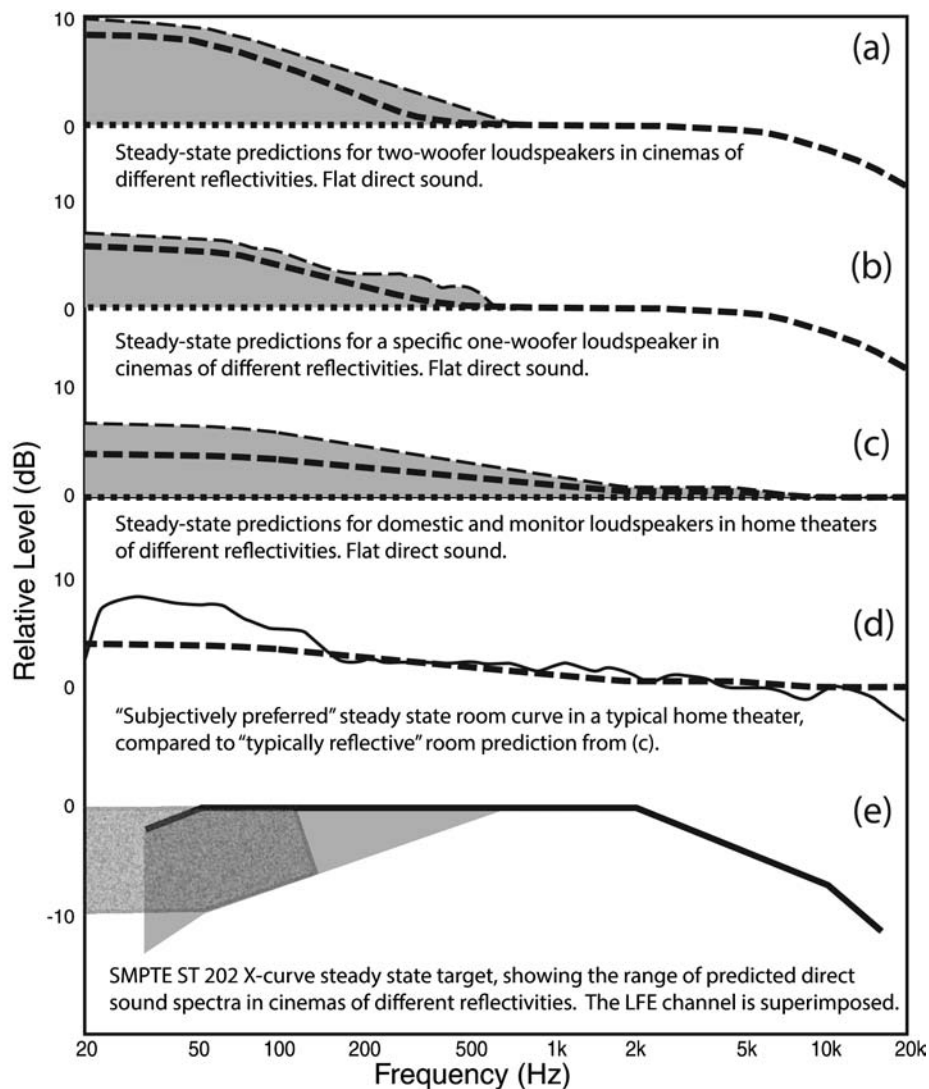
Fig. 22. (a) Predictions of steady-state sound from a two-woofer cinema loudspeaker radiating flat direct sound into cinemas; from Fig. 5(b). (b) Predictions of steady-state sound from a specific one-woofer monitor loudspeaker radiating flat direct sound into cinemas; from Fig. 18(b). (c) Predictions of steady-state sound from the one-woofer monitor loudspeaker and various full range consumer loudspeakers radiating flat direct sound into small to medium sized domestic rooms; from Fig. 13(a). (d) The "all listeners" preference curve from Fig. 14. (e) The SMPTE ST 202 steady-state target for dubbing stages and cinemas, from Fig. 7(a). The shaded area shows the range of direct sound curves for venues of different reflectivity, from Fig. 11(b). The LFE channel is superimposed showing the large spectral overlap with the screen channels.

The SMPTE/ISO method of calibration cannot deliver a flat direct sound. The high-frequency rolloff is identical in both the direct and steady-state sound fields. The low frequency direct sound will exhibit varying amounts of attenuation depending on the directivity of the woofers and the reflectivity of the venues—the shaded area. Only in dead rooms will the low-frequency direct sound be approximately flat. Nowhere is the bass allowed to rise above flat, which begs the question, were the erroneous bass boosts found in several of the cinema measurements shown earlier added because they sounded better?

Soundtracks mixed in a room calibrated as shown in Fig. 22(e) will incorporate bass and treble spectral balance adjustments that become part of the audio "art." A commonly overlooked aspect of spectral balance is that a deficiency in bass is often interpreted as an excess of treble and vice versa. When reproduced in acoustically-similar X-curve calibrated cinemas the spectral balance of the art should be preserved. However, when reproduced anywhere else in the audio domain, as represented by (a), (b), and (c)—i.e., well-designed sound reinforcement systems, recording control-rooms, home and car audio systems (including headphones)—the spectral/timbral cues cannot be the same. The art is not completely preserved (Fig. 1). Adaptable humans will still find information and entertainment, no doubt, but there is no escaping the fact that the sound quality is different. Movie soundtracks need to be compatible with the world outside cinemas, and cinemas may wish to play musical programs created outside the movie context. There is a problem.

There is no industry requirement for spectral corrections when soundtracks created for cinema presentations are

repurposed for other media. Consumers who are sensitive to timbral imbalances may find old-fashioned bass and treble tone controls to be beneficial.

Some professional music mixing rooms, dubbing stages, and a few cinemas are much more acoustically absorptive than is the norm in the rest of the audio world. A mix done in Dubbing Stage F (Fig. 9) will probably sound similar when auditioned in "reference" Cinema B, which has similar reflectivity (Fig. 3), but is likely to be different in typical cinemas represented by A, C, and D in Fig. 3 and Fig. 8. Is this a problem? Is there an effect on the perceived bass-to-midrange spectral balance when most of the reflected low-to-middle frequency sound energy is absorbed? It is a possible variable in the "circle of confusion" when movies or music recordings made in such spaces are played in normally reflective rooms.

It is tempting to simply "eliminate the room" when mixing. Some argue that it makes the job simpler. A recent paper provided evidence that acoustically dead monitoring environments may be preferred while mixing, but for mastering a more reflective space is preferred [12]. This agrees with other findings indicating that for recreational listening (which mastering should attempt to emulate) some room reflections are desirable, e.g., [1, 2, 13, 14]. In the creation of movie sound tracks mixing and mastering are simultaneous. This seems to be another worthy topic for psychoacoustic investigation.

7.2 Research Questions

In this paper several factors in sound reproduction have been addressed. Some seem to have supportable solutions, others are still a matter of opinion and judgment, including the author's. In the end, additional guidance will have to come from research. The following are some topics that are awaiting attention:

1. In normal rooms the spectra of the direct sound and the steady-state sound are different. What is the perceptual consequence of the bass rise (variables being time, amplitude, and frequency) due to room reflections? What is the detection threshold and, if necessary, can it be compensated for by electronic means?
2. To quantify room reflectivity, is it better to use the early decay of conventional reverberation time data, or to use a measure of early-reflection energy accumulation time? The latter would seem to have a more direct relationship to perceptions. Topics 1 and 2 are obviously related. In any case measurements should be made using the loudspeakers installed for the entertainment, not an omnidirectional source, such as a dodecahedron loudspeaker. These are for performance spaces, and even there they are a crude substitute for musical instruments and voices.
3. A flat, timbrally neutral, direct sound appears to be the logical objective. How best can it be achieved in practice? Anechoic loudspeaker data (corrected if

necessary for screen loss) and in-situ time-windowed measurements are possibilities.
4. What is the optimum frequency resolution for acoustical measurements? For anechoic descriptions of loudspeakers and for low-frequency measurements in rooms, high resolution (e.g., 1/20-octave) is needed to reveal and characterize audible problems, particularly resonances, and to permit correction by parametric equalization. If it is decided that steady-state in-room measurements are useful in standardized loudspeaker/room combinations, at middle and high frequencies much reduced frequency resolution might have advantages by revealing broad spectral trends without contamination by non-minimum-phase ripples that cannot be equalized. Tolerances on these measurements are another topic, and it seems that ±3 dB is too generous, embracing room curves for loudspeakers covering a wide range of perceived sound quality. This entire subject needs serious discussion.
5. At low frequencies adjacent boundaries (baffle wall, floor, nearby rear or side walls, etc.) contribute to what is measured and heard. These effects can be quantified by some number of in-room measurements and the appropriate equalization applied. A methodology is needed to address these phenomena, which are likely to exist below about 300 Hz.
6. Is it necessary to compensate for high frequency air attenuation over the propagation distances in cinemas or do humans expect it? Existing opinions relate primarily to reflective performance spaces. Does the low RT of a cinema and the expectations created by the image on a large movie screen make a difference?
7. It is evident that ear-level microphone locations experience elaborate seat dip and acoustical interference effects. It seems unlikely that these effects can be eliminated by equalization. Is it necessary to try, or do humans adapt to the local acoustical conditions? Are there better ways to characterize the direct sound?
8. If bass management is selected as an alternative to the existing LFE configuration, how many subsystems, e.g., screen, side, rear, elevation channels, are necessary to avoid distracting localizations? What are the optimum crossover frequencies and attenuation slopes?

8 DISCUSSION AND CONCLUSIONS

Steady-state room curves are important data but not complete data. It has been shown that with comprehensive anechoic data on the loudspeakers and with very basic knowledge of the acoustics of the playback venue, fundamental aspects of room curves can be predicted in many normal listening venues. The reverse is not true. Once the multidirectional sound from a loudspeaker has been launched into a three-dimensional somewhat reflective space we have lost the ability to thoroughly interrogate the sound source. Among other things, information about audible resonances

in loudspeakers may be obscured in room curves, meaning that they cannot be evaluated or corrected.

A binaural human listener is vastly more complex and capable than an omnidirectional microphone and analyzer, and we are not close to having a computer equivalent. In the meantime, it has been found that a comprehensive anechoic data set on a loudspeaker can be manipulated to predict subjective sound quality ratings from double-blind listening tests conducted in a normally-reflective medium-sized listening room [45, 46]. The correlations between the subjective ratings and the objective predictions are too high to ignore: 0.86 (for loudspeakers of varying low-frequency extension) to 0.995 (for loudspeakers of comparable bandwidth), $p = <0.0001$ in both cases. Bass performance alone accounted for about 30% of the overall subjective ratings—spectral balance matters.

This is a logical parallel to common experience in live unamplified concerts, where humans are able to substantially separate the sounds of instruments and voices from the sounds added by the different venues in which they perform, even though the venue is truly part of the performance. If the goal is good sound it is hard to escape the notion that the starting point for a calibration scheme is free-field data on the loudspeakers. A minimum data set would consist of descriptions of direct sound (on axis and/or listening window) and total radiated sound power and/or directivity index. The target for the direct sound should most likely be flat: neutral. As shown in Fig. 19(c), in "typical" cinemas and home listening rooms such loudspeakers should deliver predictably good sound over the mid-to-high frequency range with no intervention. No "calibration" may be necessary beyond compensating for screen loss in cinemas.

There is a difference between the level of the direct sound and that of the steady-state sound that increases at lower frequencies by an amount determined by the directivity of the loudspeaker and the reflectivity of the venue. Both direct and steady-state sounds are important to timbral perceptions, but the awkward reality is that in normal rooms both cannot be the same. Therefore, neither a steady-state room-curve nor a direct sound curve can be a definitive descriptor of timbre for all programs in all venues at all frequencies. The movie industry has decided that the steady-state sound should be flat below 2 kHz, most of the music/audio industry has decided that a flat direct sound is the norm. The latter has the advantage of agreeing with natural hearing of live (unamplified) acoustical events.

In small rooms, in-situ measurements are necessary at low frequencies because of standing-waves, and remedial measures will almost certainly be necessary. Multiple-subwoofer strategies are able to reduce the seat-to-seat variations, allowing equalization to be effective for more than one listener. Adjacent-boundary issues exist in all rooms. These will be revealed by spatially-averaged in-room measurements and they are responsive to equalization. In the end, a single metric that does everything remains elusive, meaning that the proper calibration of a sound-reproducing system in a large venue of unknown pedigree may require some time and a knowledgeable technical team. However,

it needs only to be done once, after which periodic routine checks employing simple measurements can ensure the continued functioning of all the elements. Standardizing an anechoic data set for loudspeakers would be a significant simplification to the entire process.

We are left with a combination of questions and answers, but at this time there is enough information to think about some universal guidelines for sound reproduction—anywhere it occurs. The starting point would be the delivery of an accurate, neutral, direct sound. The fact that there is a rise in bass sound level in the short (<150 ms) interval following the direct sound is a variable for which there is no practical control—it is room and source-directivity dependent. Humans may well regard this phenomenon as a normal component of room sound, in which case it is not an "error" in need of fixing but merely information about the listening environment. It means that a steady-state room curve should rise by some amount at low frequencies. It is a worthy topic for research.

There is no need for fundamentally different approaches for auditoriums, cinemas, stereo systems, home theaters, or cars—they each have understandable distinctions but they all exhibit basically similar performances—a theme with surprisingly small variations. We don't yet have all the pieces of the puzzle, but we have enough to imagine that some carefully constructed research can provide the missing links and validations. We do know, however, that some of the present recommendations and industry practices are not optimum and others are simply wrong.

9 REFERENCES

[1] F. E. Toole, *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms* (Focal Press, 2008).

[2] F. E. Toole, "Loudspeakers and Rooms for Sound Reproduction—A Scientific Review," *J. Audio Eng. Soc.*, vol. 54. pp. 451–476 (2006 Jun.).

[3] J. Meyer, *Acoustics and the Performance of Music*, Fifth Edition (Springer, New York, 2009).

[4] SMPTE TC-25CSS, "B-Chain Frequency and Temporal Response Analysis of Theaters and Dubbing Stages," *Soc. Motion Picture and Television Eng.* (2014).

[5] Y-J. Choi, J. S. Bradley, and D-U. Jeong, "Experimental Investigation of Chair Type, Row Spacing, Occupants, and Carpet on Theater Chair Absorption," *J. Acoust. Soc. Am.*, vol. 137, pp. 105–116 (2015). http://dx.doi.org/10.1121/1.4904532

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

[7] IEC-268-13, "Listening Tests on Loudspeakers," 11985).

[8] F. E. Toole, "Listening Tests—Turning Opinion into Fact," *J. Audio Eng. Soc.*, vol. 30, pp. 431–445 (1982 Jun.).

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

[10] R. Genereux, "Signal Processing Considerations for Acoustical Environment Correction," *Audio Eng. Soc. UK 7th Conference: Digital Signal Processing, Paper DSP-14* (1992 Sep.).

[11] L. D. Fielder, "Analysis of Traditional and Reverberation-Reducing Methods of Room Equalization," *J. Audio Eng. Soc.*, vol. 51, pp.3–26 (2003 Jan./Feb.).

[12] S. Tervo, P. Laukkanen, J. Pätynen, and T. Lokki, "Preferences of Critical Listening Environments Among Sound Engineers," *J. Audio Eng. Soc.*, vol. 62, pp. 300–314 (2014 May)

[13] 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 (1997). http://dx.doi.org/10.1121/1.381661

[14] S. Kishinaga, Y. Shimizu, S. Ando, and K. Yamaguchi, "On the Room Acoustic Design of Listening Rooms," presented at the *64th Convention of the Audio Engineering Socoity* (1979 Nov.), convention paper 1524.

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

[16] R. F. Allison, "The Sound Field in Home Listening Rooms, II," *J. Audio Eng. Soc.*, vol. 24, pp. 14–19 (1975 Feb.).

[17] T. Welti, "How Many Subwoofers Are Enough," presented at the *112th Convention of the Audio Engineering Society* (2002 Apr.), convention paper 5602.

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

[19] A. Celestinos and S. B. Nielsen, "Optimizing Placement and Equalization of Multiple Low Frequency Loudspeakers in Rooms," presented at the *119th Convention of the Audio Engineering Society* (2005 Oct.), convention paper 6545.

[20] A. Celestinos and S. B. Nielsen, "Low Frequency Sound Enhancement System for Rectangular Rooms Using Multiple Low Frequency Loudspeakers," presented at the *120th Convention of the Audio Engineering Society* (2006 May), convention paper 6688.

[21] SMPTE ST 202:2010, "For Motion Pictures—Dubbing Stages (Mixing Rooms), Screening Rooms and Indoor Theaters—B-Chain Electroacoustic Response" (2010).

[22] ISO 2969, "International Standard. Cinematography—B-Chain Electroacoustic Response of Motion-Picture Control Rooms and Indoor Theaters—Specifications and Measurements" (1987).

[23] 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). http://dx.doi.org/10.1121/1.1570439

[24] C. S. Watson and R. W. Gengel, "Signal Duration and Signal Frequency in Relation to Auditory Sensitivity," *J. Acoust. Soc. Am.*, vol. 46, pp. 989–997 (1969). http://dx.doi.org/10.1121/1.1911819

[25] 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.).

[26] B. G. Shirley, P. Kendrick, and C. Churchill, "The Effect of Stereo Crosstalk on Intelligibility: Comparison of a Phantom Stereo Image and a Central Loudspeaker Source," *J. Audio Eng. Soc.*, vol. 55, pp. 852–863 (2007 Oct.).

[27] G. Leembruggen, "Low Frequency Issues: Cinema/Home/Internet," presented at the *AES 57th International Conference: The Future of Audio Entertainment Technology: Cinema, Television and the Internet* (2015 Mar.), conference paper 1-4.

[28] J. S. Bradley, G. A. Soulodre, and S. Norcross, "Factors Influencing the Perception of Bass," *J. Acoust. Soc. Am*, vol. 101, p. 3135 (1997). http://dx.doi.org/10.1121/1.419017

[29] F. E. Toole, "Home Theaters to Cinemas: Sound Reproduction in Small and Large Rooms," presented as part of Sound for Pictures workshop at the *133rd Convention of the Audio Engineering Society* (2012 Oct.).

[30] L. D. Fielder, "Frequency Response versus Time-of–Arrival for Typical Cinemas," presented at the *2012 SMPTE Annual Technical Conference*.

[31] T. Holman, "Cinema Electro-Acoustic Quality Redux," *SMPTE Mot. Imag. J.*, vol. 116, no. 5–6, pp. 220–233 (May/June 2007).

[32] P. Newell, K. Holland, S. Torres-Guijarro, S. Castro, and E. Valdigem, "Cinema Sound: A New Look at Old Concepts, *Proc. Inst. of Acoust.*, vol. 32, pt. 5 (2010).

[33] W. B. Snow, "Loudspeaker Testing in Rooms," *J. Audio Eng. Soc.*, vol. 9, pp. 54–60 (1961 Jan.).

[34] J. Eargle, J. Bonner, and D. Ross, "The Academy's New State-of-the-Art Loudspeaker System," *SMPTE Mot. Imag. J.*, vol. 94, no. 6, pp. 667–675 (June 1985).

[35] I. Allen, "The X-Curve: Its Origins and History," *SMPTE Mot. Imag. J.*, vol. 115, no. 7–8, pp. 264–273 (July/Aug. 2006).

[36] L. A. Gedemer, "Evaluation of the SMPTE X-Curve Based on a Survey of Re-recording Mixers," presented at the *135th Convention of the Audio Engineering Society* (2013 Oct.), convention paper 8996.

[37] T. Holman, "A History of the X-Curve," *Surround Professional*, vol. 3 (Jan. 2000).

[38] S. E. Olive, T. Welti, and E. McMullin, "Listener Preference for Different Headphone Target Response Curves," presented at the *134th Convention of the Audio Engineering Society* (2013 May), convention paper 8867.

[39] P. Newell, "Cinema Calibration and the X-Curve," *J. Audio Eng. Soc.*, vol. 62, pp. 808–809 (2014 Nov.).

[40] B. Long, R. Schwenke, P. Soper, and G. Leembruggen, "Further Investigations into the Interactions Between Cinema Loudspeakers and Screens," *SMPTE Mot. Imag. J.*, pp. 46–62 (Nov./Dec. 2012).

[41] P. Newell, J. G. Garcia, and K. Holland, "An Investigation into the Acoustical Effects of Cinema Screens on Loudspeaker Performance," *Proc. Inst. of Acoustics*, vol. 35, Pt. 2 (2013).

[42] D. Elliott, K. Holland, and P. Newell, "The Audibility of Comb-Filtering Due to Perforated Cinema Screens," *Proc. Inst. of Acoustics*, vol. 36, Pt. 3 (2014).

[43] G. Leembruggen and P. Newell, "Equalisation of High Frequency Cinema Screen Losses" presented at the *AES 57*th *InternationalConference: The Future of Audio Entertainment Technology: Cinema, Television and the Internet* (2015), conference paper 1-4.

[44] S. E. Olive, B. Castro, and F. E. Toole "A New Laboratory for Evaluating Multichannel Audio Components and Systems," presented at the *105*th *Convention of the Audio Engineering Society* (1998 Sep.), convention paper 4842.

[45] S. E. Olive, "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 1 – Listening Test Results," presented at the *116*th *Convention of the Audio Engineering Society* (2004 May), convention paper 6113.

[46] S. E. Olive, "A Multiple Regression Model for Predicting Loudspeaker Preference Using Objective Measurements: Part 2 – Development of the Model," presented at the *117*th *Convention of the Audio Engineering Society* (2004 Oct.), convention paper 6190.

[47] R. Green and T. Holman, "First Results from a Large-Scale Measurement Program for Home Theaters," presented at the *129th Convention of the Audio Engineering Society* (2010 Nov.), convention paper 8310.

[48] S. E. Olive, T. Welti, and E. McMullin, "Listener Preferences for In-Room Loudspeaker and Headphone Target Responses," presented at the *135th Convention of the Audio Engineering Society* (2013 Oct.), convention paper 8994.

[49] S. E. Olive, "A New Reference Listening Room for Consumer, Professional and Automotive Audio Research," presented at the *126th Convention of the Audio Engineering Society* (2009 May), convention paper 7677.

[50] S. E. Olive, J. Jackson, A. Devantier, D. Hunt, and S. Hess, "The Subjective and Objective Evaluation of Room Correction Products," presented at the *127th Convention of the Audio Engineering Society* (2009 Oct.), convention paper 7960.

[51] S. E. Olive and T. Welti, "Validation of a Binaural Car Scanning System for Subjective Evaluation of Automotive Audio Systems," presented at the *AES 36th International Conference: Automotive Audio* (2009 Jun.), conference paper 7.

[52] D. Clark, "Progress in Perceptual Transfer Function Measurement—Tonal Balance," presented at the *110th Convention of the Audio Engineering Society* (2001 May), convention paper 5407.

[53] M. Binelli and A. Farina, "Digital Equalization of Automotive Sound Systems Employing Spectral Smoothed FIR Filters," presented at the *125th Convention of the Audio Engineering Society* (2008 Oct.), convention paper 7575.

[54] L. A. Gedemer, "Predicting the In-Room Response of Cinemas from Anechoic Loudspeaker Data," presented at the *AES 57*th *International Conference: The Future of Audio Entertainment Technology* (2015 Mar.), conference paper 1-1.

[55] ANSI/CEA-2034, "Standard Method of Measurement for In-Home Loudspeakers," Consumer Electronics Association, Technology & Standards Department, www.CE.org (2013).

[56] https://www.klippel.de/our-products/rd-system/modules/nfs-near-field-scanner.html

[57] D. Logan, "A New Approach to Loudspeaker Measurements," *Voice Coil*, vol. 28, no. 7, pp. 8–12 (2015).

[58] A. Devantier, "Characterizing the Amplitude Response of Loudspeaker Systems," presented at the *113*th *Convention of the Audio Engineering Society* (2002 Oct.), convention paper 5638.

[59] S. E. Olive, P. L. Schuck, S. Sally, and M. Bonneville, "The Variability of Loudspeaker Sound Quality among Four Domestic Sized Rooms," presented at the *99*th *Convention of the Audio Engineering Society* (1995 Oct.), convention paper 4092.

[60] A. S. Bregman, *Auditory Scene Analysis, the Perceptual Organization of Sound* (MIT Press, 1990).

[61] S. E. Olive, S. M. Hess, and A. Devantier, "Comparison of Loudspeaker-Room Equalization Preference for Multichannel, Stereo and Mono Reproductions: Are Listeners More Discriminating in Mono?" presented at the *124th Convention of the Audio Engineering Society* (2008 May), convention paper 7492.

THE AUTHOR

Floyd E. Toole

Floyd E. Toole studied electrical engineering at the University of New Brunswick and at the Imperial College of Science and Technology, University of London where he received a Ph.D. In 1965 he joined the National Research Council of Canada where he reached 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 worked with all Harman International companies and directed the Harman Research and Development Group, a central resource for technology development and subjective measurements, retiring in 2007. He is currently a consultant to Harman.

Dr. Toole's research has focused on the acoustics and psychoacoustics of sound reproduction, directed to improving engineering measurements, objectives for loudspeaker design and evaluation, and techniques for reducing variability at the loudspeaker/room/listener interface. For papers on these subjects he has received two AES Publications Awards and, for service to the society, the Board of Governors Award. For his achievements he has been recognized with both the AES Silver Medal (1996) and Gold Medal (2013) Awards. He is a Fellow and Past President of the AES, a Fellow of the Acoustical Society of America, and a Fellow of CEDIA (Custom Electronic Design and Installation Association). He has been awarded Lifetime Achievement awards by CEDIA and ALMA (Association of Loudspeaker Manufacturing & Acoustics International). In 2015 he was inducted into the CE Hall of Fame. He has authored the book, *Sound Reproduction: The Acoustics and Psychoacoustics of Loudspeakers and Rooms* (Focal Press, 2008).