Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 1*

FLOYD E. TOOLE

National Research Council, Ottawa, Ontario K1A 0R6, Canada

Precision measurements on loudspeakers have been possible for some time now, and over the years, various views of their importance have developed as a result of accumulated experience and scientific investigation. A survey of the literature reveals areas of agreement and disagreement among workers. There is also evidence of geographic concentrations of loudspeaker designers and reviewers who appear to share attitudes toward specific measurements. Part 1 attempts to consolidate published opinion on loudspeaker measurements in preparation for Part 2, which presents the results of some current research on the subject.

0 INTRODUCTION

Evaluating loudspeakers by means of measurements is rather like being a detective looking for clues to the existence and origins of misbehavior. The measurements that we use are not, in themselves, virtues or problems; they are merely indicators. Two ears and a brain do not process sounds as do microphones, measuring instruments, and the eyes. The auditory perceptions of musical sounds are not the same as visual analyses of data from clinical sine waves, impulses, and pink noise. And yet, for now at least, there is no choice but to proceed with the traditional methods bearing in mind *always* that evidence that offends the eye may or may not indicate the presence of a problem that is offensive to the ear.

The relationship between listener perceptions and measured quantitites is a general problem in psychometrics that has many manifestations in audio. The most basic relationships deal with the detection thresholds of sounds of various kinds, with or without the embellishment of added distortions. From these one progresses through the relationships between the measured quantity of a stimulus and the strength of the associated perception. The scientific literature is replete with learned studies of this kind, dealing, as much as possible, with isolated perceptual dimensions that are logically and clearly related to individually measurable parameters.

The situation becomes more complicated when the experiment is expanded to include multiple stimuli, and vastly more complex with the inclusion of stimulus inputs through more than one sensory modality. If the controls are further relaxed, and the repertoire of experimental signals includes real-world sounds propagated in real-world environments, precise experimental control becomes almost impossible and the perceptions are no longer simple. Yet these are the conditions of a typical listening test within which listeners readily tender opinions of preference, sometimes complete with detailed subjective analyses of what they hear.

It is characteristic of commonplace listening experiences that listener opinions will vary. It is argued by some that it is the variety of listener tastes and preferences that has led to the wide diversity of sounds from loudspeakers, all of which pass, loosely, as "high fidelity." Thus has developed the traditional problem of attempting to match the tastes of the consumer with those of the loudspeaker designer; a messy business indeed.

With enough attention to detail, however, it is possible to demonstrate that much, if not most, of the variation in listener opinion stems not from irreconcilable differences among individuals but from the influence of physical and psychological factors, many of them unrelated to the matter in question. There may indeed be clear differences between listeners with widely different hearing ability, but among listeners with similar, particularly similarly good, hearing the differences of opinion can be very small indeed [1].

* Manuscript received 1985 May 30; revised 1986 January 9.

From the results of the strictly controlled listening tests that have been conducted it is possible to clearly identify loudspeakers that are consistently favored by listeners with normal hearing and those that are not. In fact, the fidelity ratings from these subjective measurements can be used as a means to explore the objective domain, seeking to identify those measured variables that are indicators of listener preference.

The ultimate objective of the present investigation could be stated as follows: to define a set of technical measurements, including the manner of their measurement and the form of their presentation, such that interpretations by experienced unbiased observers correspond to the results of controlled listening tests using experienced unbiased listeners. A loudspeaker that has a good technical performance, in certain specific terms, should sound good in the listening room—and vice versa.

1 AN HISTORICAL PERSPECTIVE

Before presenting the results of the present research work it is important to appreciate the historical context into which they fit. In most scientific and engineering endeavors precise measurements are taken for granted as being of fundamental value. In the design and evaluation of loudspeakers, however, the role of measurements has never been clearly defined.

Part of the difficulty stems from the nature of the device itself. A loudspeaker generates a three-dimensional sound field within a complicated acoustical enclosure, the listening room. Without an understanding of the sound propagation characteristics of the room and of the perceptual processes of the human recipient of the sound, it has not been possible to be definitive about the aspects of loudspeaker performance that are most important.

Inherent in the design or evaluation of loudspeakers is the establishment of some rules by which the work will proceed. By tradition, subjective evaluations have remained the final arbiters of sound quality. However, measurements have been assuming an ever-increasing role in loudspeaker work, and lacking definitive scientific data, workers have used their own experiences and beliefs to develop rules for using measurements. In the early years progress was hampered by primitive instrumentation and poor sound sources. The motivations were, nevertheless, very high indeed, and one need look no further than to a paper by Brittain, in 1936–1937 [2], for a remarkably perceptive review of the hearing process and loudspeaker measurements. A "standard living room" for listening tests is described, as are the techniques, such as spatial or frequency-domain averaging, that help in extracting useful data from measurements in normal rooms. Other measurements were made either in an "acoustically absorbing" room or on a tower in genuine free space.

Brittain sums up his opinions in the following list of "common imperfections of electroacoustic systems . . . , roughly in order of importance." (Some of the author's terminology has been updated for the benefit of modern readers.)

1) Frequency response (amplitude response)
2) Harmonic distortion
3) Spurious noises and intermodulation distortion
4) Frequency shift (FM distortion?)
5) Dynamic range compression (power compression)
6) Transient distortion
7) Phase distortion
8) Group delay
9) Electroacoustic efficiency
10) Power-handling capacity
11) Constancy of performance.

Looking at this list, one is likely to be overcome with a feeling of déjà vu, and rightly so; these conclusions are 50 years old. Yet they hold up well in light of today's knowledge.

The following is a review of more recent opinions expressed in the literature. Unfortunately the writings are fragmented, with useful contributions to be found in consumer audio publications and trade or corporate publications as often as in professional journals. In all there are few spokespersons for such a large and mature industry. Most of what has transpired in the field of loudspeaker system design has occurred behind closed doors by people who range from enthusiastic amateurs with an artist's touch to straightforward engineers. Few of these workers have made their ideas public.

In addition, there are the product reviewers. These people, by their journalistic nature, are more candid about their approaches to technical assessment. Some reviewers eschew measurements entirely, or give them only token respect.

Naturally, influences can bear both ways, from designers to reviewers, and vice versa. There are substantial differences of opinion in both camps, and some interesting similarities. For example, there are suggestions of geographical patterns, with the designers and reviewers of specific areas forming what appear to be mutually compatible closed systems. This may or may not have anything to do with listener taste, since adequately controlled and reported listening tests are almost nonexistent, but it certainly does seem to have something to do with the locally fashionable form of amplitude-response measurement.

Almost all workers acknowledge the value of more than one form of amplitude-response measurement. For different reasons, however, everyone seems to have a favorite, and the following rough classification reflects the predominant view held by each author, at least at the time of publication.

The presentation of these views has been organized around the sequence of sounds arriving at a listener's ears in a normal listening room. First there is the direct sound, normally the sound radiated along the reference axis of the loudspeakers. Next come the sounds that have been reflected once from the adjacent room boundaries—floor, walls, and ceiling—radiated from the loudspeaker at specific and sometimes large angles off axis. Still later there is the reverberant field developed from sounds that have been reflected several times from the room boundaries and objects within the room. These sounds include, to varying degrees, sounds radiated from the loudspeaker in all directions.

1.1 Amplitude Response as a Function of Listening Perspective

The first item on Brittain's list is amplitude response, referring mainly to on-axis measurements in free space. This curve, as well as being the most flattering to most loudspeakers, describes the direct sound at the ears of properly seated listeners. A number of workers such as Hentsch [3] in Switzerland, and Harwood [4], [5] and Mathers [6] at the BBC and Colloms [7], also in the United Kingdom, clearly believe in the preeminence of the direct sound in establishing perceived sound quality. Their emphasis is therefore on amplitude responses measured in the free field on axis and within a listening window sufficient to include the ears of normally positioned listeners, usually up to 30–45° off axis horizontally, and less vertically.

Moving, philosophically, further into the listening room, we embrace the views of those who believe in including some measure(s) of sounds that arrive at a listener's ears after one or more reflections. Shorter [8], writing of his BBC experiences in 1958, some years before Harwood and Mathers, said that "experience . . . suggests that, if a single quantity representing 'effective' response is to be found at all, it will lie somewhere between the axial and the mean spherical response." He goes on to suggest a spherical integration weighted to give the front response more prominence.

Moir appears to be in general agreement with the suggestion of an ideal loudspeaker having uniform sound distribution over "something less than ±90 degrees in front of the loudspeaker" [9] and of measurements that take into account both the direct and the reverberant energies by using temporal or spectral integrations [10]. He reaches no final conclusions but notes simply that at the present stage free-field measurements are "in better agreement with the subjectively judged response than any response curve measured in the listening room" [10].

Remaining in the United Kingdom, Walker [11] and Cooke [12], while acknowledging the importance of an uncolored direct sound, point out that indirect sounds should be similarly free of irregularities to avoid perceptible coloration.

Queen in the United States concluded that for sound reinforcement the direct sound was the dominant factor in spectrum perception [13], but after examining stereo imaging closely, he observed that "loudspeaker designs for home music listening rooms must consider directivity not from the standpoint of audience coverage (direct sound) but from the standpoint of uniformity of the intensity of arriving reflections with respect to frequency." From this he went on to specify loudspeaker design objectives of either a nondirectional horizontal radiation, with restricted vertical dispersion, or "a directional loudspeaker providing high uniformity of directional pattern with frequency" [14].

Heyser, also in the United States, without reference to supporting subjective data, selected a combination of free-field measurements and listening-room measurements time-windowed to include early reflections, clearly expressing a compatible philosophy [15].

Komamura et al. [16], in Japan, performed factor analyses on extensive subjective and objective data, and concluded that, of the measured parameters, frequency responses measured in the listening room and in the anechoic room, at 0, 30, and 60°, have a high correlation with listener preference. This, clearly, is a transitional point of view, since it suggests that certain free-field data and listening-room measurements are compatible.

The next significant shift of perspective is to embrace the listening-room sound field in its entirety. Some workers, noting that listeners are often in a predominantly reverberant sound field, alluded to the "substantial evidence that the total acoustic power output of a speaker system over its frequency range is the most important characteristic determining the frequency response heard by the listener" [17]. Brociner and von Recklinghausen were careful to point out, however, that such measures can be fooled, saying that " . . . a nonsmooth power response measured in the reverberation room indicates a similarly irregular frequency response as measured in an anechoic chamber. While it cannot be stated that a speaker system which shows a smooth response in the reverberation room will necessarily sound good or have a smooth pressure response, the reverse is true."

It was at the same time, 1968, that Bose developed his distinctive loudspeaker and, along with it, the view that "when the loudspeaker is properly placed with respect to the rear reflecting wall, the frequency response measured with respect to the total radiated acoustical energy should be flat" [18]. This position was reinforced in a later publication where he concluded that "loudspeakers should be designed to a flat power criterion rather than the conventional flat-frequency-response-on-axis criterion" [19]. Bose, by implication, was referring to his own designs which (then) were either multidirectional or at least widely dispersing, generating a clearly dominant reverberant sound field in the listening room.

Others applied the sound power criterion to conventional loudspeakers, among them several who, perhaps by chance, worked in the same New England (U.S.) area. McShane [20], Allison and Berkovitz [21], Consumers Union [22], Torick [23], and Hirsch [24] all attributed a dominant role to sound power in determining the spectral character of sound in a listening room. Rosenberg [25] and Staffeldt [26] in Scandinavia prevent this from being an exclusively U.S. point of view.

It therefore appears that in the United States and in the United Kingdom there were collections of designers and reviewers sharing perspectives on measurements that would seem to place them at opposite ends of a controversy.

There is probably no better way to illustrate the trans-

Atlantic polarization than to quote from Shorter [8] who says, "The only firm conclusion which can safely be drawn is that with wide-range loudspeakers of conventional directional characteristics a flat axial response may be acceptable but a flat mean-spherical (i.e., power) response is intolerable."

On the face of it, it would seem that both points of view cannot be correct. Consequently there is a tendency for workers, including some of those mentioned above, to perform both kinds of measurements, just to be comprehensive.

Even among the advocates of sound power, the measurements were not seen to be faultless. In addition to the caution stated by Brociner and von Recklinghausen, Allison and Berkovitz commented on a discrepancy in the measurements noting that "the average room does not give the low-frequency support that is commonly assumed." These were attributed to source–boundary interactions, well known in physical acoustics, and identified earlier by Sioles [27], discussed later by Allison [28].

If the purpose of a measurement is to predict loudspeaker performance in a listening room, why not simply do it there? Sioles, in 1963, "recognized that measurements on loudspeakers must be taken in their normal operating environment in order to assess the overall performance unless a large amount of free-field data is accumulated and averaged" [27].

Indeed, the simplicity of the method is appealing, since no special measuring environments are necessary and, using one-third-octave bands of noise, even the instrumentation is simple. Benade, however, cautioned that such filter widths may fail to reveal important irregularities [29].

Nevertheless, one-third-octave measurements are widely used for assessments by workers who have found it to be a good indicator of listener preference. Observing that the optimum listening-room curve was not flat, various ideal functions have been proposed, for example, Møller [30]. Russell [31] embellished the method with some spatial averaging and suggested a range of "ideal" functions. Holl [32] also employed spatial averaging over the listening area, but instead of using a generalized nonflat target curve, he derived a room calibration based on reverberation-time data and corrected the actual measured curve so that the ideal response was at least approximately flat. It is equivalent to measuring sound power in a sound field that is not perfectly diffuse. Holl alluded to controlled listening tests leading him to conclude that "the room measurement is the best indication of what a system will actually sound like." Long [33] reported using a similar method (developed for product reviews by E. Foster, with the cooperation of T. Holl and A. DeKoster).

Schulein [34] tested the technique and concluded that the nonflat room curves in domestic systems were a part of a self-consistent system, including similarly nonflat control-room monitoring. He speculated that improvements in the program material might increase the desirability of an acoustically flat playback system.

The fact that the "ideal" room curve is not flat is a problem for which there appears to be no unambiguous solution. Consequently one sees various approaches to rationalizing the problem; either select a "target curve" and attempt to imitate it, or derive a room "correction" and subtract it from the measurement. Either way important assumptions are involved.

The generalized "target curve" approach used by Møller and Russell assumes that all rooms are similar in their sound diffusion and frequency-dependent absorption. Adjacent-boundary interactions are not included and neither are any considerations of loudspeaker directivity and its effect on the direct-to-reverberant ratio of the sound field. One might expect some uncertainty about the shape of the "ideal" response curve and, in Fig. 1, this is apparent.

The approach used by Holl and Long has rather better prospects since by using standard microphone and loudspeaker placement, the boundary interactions can be maintained as constant factors, or with some calibration data, removed entirely. The room calibrations at middle and high frequencies rely on reverberation-time measurements and thus assume a diffusion in the sound field that may or may not exist. Loudspeakers of different directivities would therefore yield data requiring individual interpretation; the method may be useful within the limited scope of fundamentally similar products, but there are likely to be errors in the evaluation of loudspeakers in general. Nevertheless, the "frequency response" of the room used in Holl's example [32, Fig. 2], shown in Fig. 1, is close to Møller's optimum curve; apparently both workers used acoustically similar rooms. The alternative target curves suggested by Russell have very generous tolerances, although the −3-dB-per-decade slope is in the same spirit as the Møller and Holl curves.

1.2 Amplitude Response: A Synopsis

Without recourse to directly comparable measurements to illustrate the point, it is evident that there is serious disagreement among knowledgeable and ex-

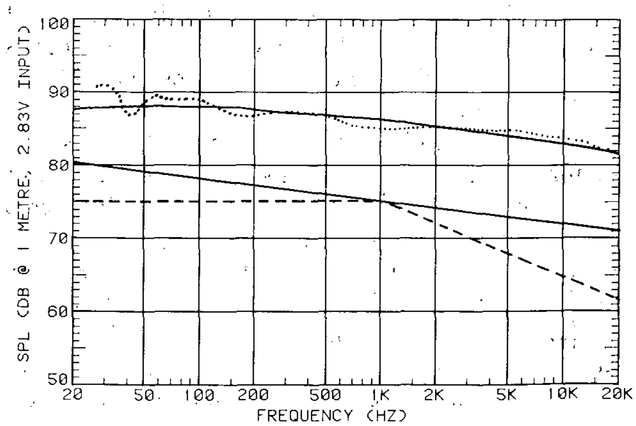

Fig. 1. "Ideal" listening-room amplitude-response curves proposed by Møller (top solid) and Russell (bottom solid and dashed). For comparison a room correction curve used by Holl is shown (dotted).

perienced workers about what constitutes the most useful measure of loudspeaker amplitude response. The magnitude of the disagreement is such that, with the popular forward-facing configuration of loudspeakers, meeting one of the amplitude-response criteria, such as a flat on-axis response, would automatically result in failure of another, the flat sound-power response, because of the directional characteristics of the device. Measurements within the listening room are sensitive to a variety of factors related to both the loudspeaker and the room. If performance in the room is the ultimate criterion, it is unlikely that measurements on the loudspeaker alone would be very reliable, since they convey only part of the information.

To the extent that the general form of the amplitude response is a determinant of listener preference, it would seem that the industry has not yet decided on a design objective.

1.3 Phase Response

In engineering terms it is so logical that phase response should matter that, over the years, several workers have repeated the investigations, only to conclude that in the real world it is, at worst, only a minor problem. With selected test signals and listening with headphones or in an anechoic room, listeners can hear differences attributable to phase shift. Using music, and listening to loudspeakers in normal rooms, even carefully selected listeners appear to have difficulty detecting the presence of quite large phase shifts, much less are they able to establish a preference. While these observations seem to be fairly general for smooth changes in phase as a function of frequency, including substantial group delays, the rapid local phase changes indicative of resonances may be another matter. Always, and especially in loudspeakers with multiple drivers and overlapping crossover regions, there is the uncertainty of whether one is hearing the effects of phase, or of an associated amplitude aberration.

This very point was made by Brittain in 1937, who noted the insensitivity of the ear to phase effects "unless accompanied by some other phenomenon" [2]. Hentsch, in 1951, observed that "phase variations" were unimportant, and large group delays barely detectable, even by experienced listeners [3].

Such results from those early years could be disputed nowadays on the basis of the rather poor microphones and other equipment involved in the tests. They are therefore all the more impressive in view of the fact that much more recent investigations by Preis [35], Bauer [36], Harwood [37], Moir [38], Lipshitz et al. [39], Saponas et al. [40], and Clark [41] among others, have come to basically similar conclusions.

The recent work has been more thoroughly investigative and puts some limits on the thresholds of audibility of various phase and group-delay effects. The limits on overall trends in phase response are very generous, though, and appear not to require special consideration in the design of conventional domestic loudspeakers. However, Saponas et al. are careful to point out that localized sharp discontinuities in amplitude response tend to be associated with high rates of change of phase shift and the combination causes "frequency distortion."

Others, though, have disagreed with these views, among them Bowers and Roe [42] and Gerzon [43]. In [42], the authors compare two loudspeakers, one adjusted to yield a linear phase response and a similar nonlinear phase version. There was "some departure in measured amplitude response in the nonlinear phase system, but as identical drive was applied to identical units this was not counted a relevant variable." In other words, the electrical input to the two loudspeaker systems was the same, but the acoustical output differed in both amplitude and phase response. One is left to wonder which parameter caused the "clearly audible" differences reported by the listeners. In these experiments it is difficult to manipulate one parameter in isolation, particularly when altering electroacoustic devices, a point well made by Barlow [44] who investigated the same effect and concluded that "any audible effect due to the acoustic centres being in the same plane is very small and is masked by reflection" (from the front panel setbacks).

1.4 Audibility of Amplitude-Response and Time-Domain Anomalies

When Bücklein, in 1962, investigated the audibility of various forms and magnitudes of peaks and dips as a function of frequency, he addressed a fundamental problem in amplitude-response assessment [45]. How flat or smooth does a curve really need to be?

He observed, among other things, that an amplitude-response irregularity appearing as a peak may be highly objectionable when the equivalent dip may be barely audible.

Bücklein also found that the audibility of irregularities increased with their width and their amplitude. The latter is to be expected but the former is interesting, as it implies that low-Q resonances are more easily detected than high-Q resonances. This finding was later confirmed by Stevens [46], looking at enclosure resonances, and by Fryer [47] who explored the audibility of resonances as a function of both Q and frequency. Barlow, discussing Fryer's work, summarized the important points as follows: "The ear evidently detects mainly the energy or area under the peak. In some cases, at high Q, the peak can be well above the general level before being detectable, but in other cases a low-Q resonance, well below the general level, is still detectable. Damping a resonance may not give as great an improvement as hoped for, especially at low frequencies" [44].

In both Stevens's and Fryer's tests, listeners detected low-Q ($Q \leq 10$) resonances at levels 10 to 20 dB below average program level with as much ease as they found high-Q ($Q \geq 20$) resonances with output equal to or above the average level. In terms of measured amplitude response this means that some audible colorations may be present even though the curve may appear to be quite flat. On the other hand, some narrow spikes or sharp dips that are offensive to the eyes may be virtually ignored by the ears.

Translated into the time domain, there should be a similar sensory contradiction wherein the high-Q resonances with their prolonged decays that are so offensive to the eyes when viewed on an oscilloscope or in "waterfall" diagrams of signal energy in time and frequency [48], [49] may be less audible than rapidly damped lower amplitude overhangs. Hentsch, in fact, concluded just that when he found that viewing transient phenomena on an oscilloscope "is not a certain means of determining the subjective effects of the deformations" [3].

Shorter [50], in 1946, and Corrington [51], in 1955, developed gating systems for looking at the energy in the decaying tails of tone bursts. Both workers were convinced that this view of events in the time domain offered better indications of audible colorations than conventional amplitude-response measurements that included the visual clutter of nonresonant interference effects. Hentsch, in 1951, saw a way around this by using spatial averaging of anechoic swept-tone amplitude responses to more clearly reveal the spectral irregularities caused by resonances. He concluded that "the study of amplitude response alone is sufficient to determine the quality of reproduction in the transient regime" [3]. In 1961 Larson and Adducci [52] arrived at the same conclusion and observed further that "little correlation exists between the transient performance of a loudspeaker and musical listening tests."

In 1971 Saponas, Matson, and Ashley (40) examined a variety of transient and steady-state test signals and concluded that most of the commonly used transient signals were inappropriate for evaluating loudspeakers, and for even a simple impulse, it would be "nearly impossible to look at the output from a microphone and say if this is a good or bad result." They concluded that "the sine wave is still very much the king of test signals."

Driscoll, writing in 1974, basically agreed with this result saying that while he thought that impulse tests can be "a useful pointer to at least some qualities of a loudspeaker's performance," the test by itself cannot be used to rank products "in any order of performance quality" [53]. Unwin expressed almost the directly opposite point of view, placing great importance in the visual form of the impulse response and arguing that "testing a loudspeaker with a steady sine wave tells us almost nothing about the way it will handle transients which are the most important constituents of musical performance" [54]. Unfortunately no proofs were offered.

In distinct contrast to all of these endeavors to identify and quantify the sources of resonance and transient colorations there is the conclusion of Bose who "learned that it is possible to produce music without audible coloration from distortion, resonances or transient response irregularities by the use of a multiplicity of full range loudspeakers" [19]. These tests were done with loudspeakers that generated a predominantly reverberant sound field so that, in addition to the small differences among the various drivers, there was the confusion of the mostly reflected sound field to disguise the audible effects of these technical imperfections. Briggs, writing earlier, in 1958, noted that directing the axis of a loudspeaker toward a wall or ceiling "camouflages peaks in response to a remarkable extent because the room gets to work on the sound waves before they reach the ear of the listener" [55]. These findings might be related to that of Harwood [5] who obtained the "remarkable result" that listeners could not hear a series of low-Q resonances at spacings of less than one octave over the entire frequency range.

Although there are some clear differences of opinion, there are nevertheless some recognizable patterns in this collection of findings and they can be described as follows: audible colorations are often caused by resonances in a loudspeaker system. These resonances will be revealed, in varying degrees, in both steady-state amplitude responses and time-domain responses to interrupted sounds.

In order for amplitude responses to reveal unambiguous clues to the presence of resonances the measurement must have high resolution in the frequency domain and incorporate some form of spatial averaging to remove the visual clutter caused by acoustical interference. This done, evidence of potentially audible coloration will take the form of upward thrusting peaks deviating from the underlying general shape of the curve. Wide, low-amplitude bumps can be as annoying as much higher, narrow peaks. Dips in the amplitude response appear to be much less important than peaks. The technique is limited, however, in that it will not necessarily reveal the low-Q resonances 10 to 20 dB below the level of the primary signal that have been shown to be sources of audible coloration.

Explorations of time-domain performance can identify most resonance problems, but the interpretation appears to run contrary to common preconceptions. The highly visible long "tails" caused by high-Q resonances are not always the most severe problems. The broader band, better damped tails of low-Q resonances need to be ferretted out of the time-domain data, as these are often associated with the dominant audible coloration.

It is interesting to speculate why this phenomenon occurs. First, from a purely statistical point of view, a broad resonance will be excited more often by sounds in music than a narrow resonance. Second, since musical sounds at middle and high frequencies are either transient in nature or are at least strongly amplitude modulated, a low-Q resonance will reach full amplitude more often and more quickly than a high-Q resonance. And, finally, since these are forced resonances, once the exciting signal is removed, the resonant system reverts to its natural frequency. For a high-Q resonance the resonant overhang will have the same, or close to the same, frequency as the signal that stimulated it, and its presence may be difficult to detect, given that musical instruments are themselves resonant devices. A low-Q resonance, on the other hand, can ring at a frequency significantly different from the one that initiated the response, thus imparting a monotonal coloration to a range of exciting frequencies.

The frequency at which the amplitude aberration occurs also matters. All studies have indicated a reduction in audibility at both frequency extremes. Bücklein's results show two regions of high sensitivity: between about 200 Hz and 600 Hz and 2 kHz and 6 kHz. Fryer's data indicate a slightly increased sensitivity between 2 kHz and 5 kHz.

Between about 600 Hz and 2 kHz the ears seem to be marginally less sensitive to these problems. It may be more than coincidence that this is the frequency range within which there appears to be a kind of physiological crossover of auditory function. For example, below about 1 kHz the inner ear is capable of neurally encoding accurate information about the acoustic waveform; above this frequency there is an increasing reliance on the amplitude envelope, rather than carrier information. In sound localization this is apparent in the importance of interaural time differences at lower frequencies while, at high frequencies, the sensitivity to timing information seems to apply principally to the amplitude-modulated envelope rather than to the fine structure of the underlying signal [56]. It is fortunate that this frequency range, within which there may be some uncertainty about the dominant perceptual cues, includes the majority of woofer- or midrange-to-tweeter crossovers, where amplitude and phase response problems are routinely found.

1.5 Nonlinear Distortions

Shorter, in 1958, offered the advice that "with some forms of distortion it would probably be easier, as well as more profitable, to remove the cause by appropriate design than to discover rules for assessing the effect" [8]. In electronic components this principle has been applied with considerable success, but in loudspeakers, it is commonplace to measure distortions in whole percentages at modest sound levels. Either listeners are remarkably tolerant of these imperfections or there is still a major problem to be solved.

From accounts of different investigations, however, one can reach a number of conclusions. Bose concluded that with most high-fidelity loudspeakers "audible nonlinear distortion in music or speech is definitely one of the more minor of their shortcomings, notwithstanding the fact that distortion measurements . . . can be quite large" [18]. Kantrowitz arrived at a fairly generous 3% as the level of CCIF (intermodulation) distortion above which it became objectionable in high-frequency drivers [57]. Here, though, is a case where one might suspect that the source material may have had an influence on the results since the 1962-vintage disk recordings, and playback devices had similar problems of their own. The more recent work of Fryer [47] indicated similarly large values for the detection of intermodulation distortion. Noting that typical loudspeakers at typical listening levels produce less than 1% intermodulation distortion, he concluded that "it is not a particularly serious issue for designers."

However, there are several forms of nonlinear distortion and a number of important variables influencing their audibility. Cabot, in a recent review of previous work, discusses the subject in some detail [58]. While acknowledging some of the high detection thresholds that have been reported, he comments on the need for more experiments with better controls and using the superior program material available today. Selecting what he regards as possibly "the most reliable work to date," Cabot arrives at a performance objective of 0.05% total harmonic distortion, with the provision that high-order components not be dominant. If this is the case, then there is considerable room for improvement in loudspeakers, and Shorter's advice might prove to be difficult and rather expensive to implement.

Even with this small sampling of the literature on nonlinear distortion it is evident that there is no unanimity of opinion. Clearly, there is a need for more work on this subject.

2 DISCUSSION

Given the age and size of the industry it may appear to be somewhat surprising that there is so little agreement on the desirable measured performance objectives of loudspeakers. That this is true for something as basic as the amplitude response is particularly disconcerting.

Errors in the physical measurements are not likely to be at fault here, the differences at issue are simply too large. There are some suggestions that the manner of measurement and data presentation may be important, but, there would appear to be secondary factors. There is even the suggestion that geography might be a consideration.

Throughout the published accounts of loudspeaker design and evaluation methods there are references to listening tests and subjective preferences lending support to one or another technical evaluation scheme. Rarely are the subjective measurement methods described in any detail. Knowing now the attention to detail that is necessary to obtain reliable subjective data, it is reasonable to speculate that another important variable underlying these uncertainties is listener opinion.

Recent work on subjective measurements [1] has produced a large body of reliable subjective data on a number of loudspeakers. In Part 2 of this paper these data provide the basis for an examination of certain aspects of measured loudspeaker performance.

3 REFERENCES

[1] F. E. Toole, "Subjective Measurements of Loudspeaker Sound Quality and Listener Performance," *J. Audio Eng. Soc.*, vol. 33, pp. 2–32 (1985 Jan./Feb.)

[2] F. H. Brittain, "The Appraisement of Loudspeakers," *GEC J.*, pt. 1, vol. 7, pp. 266–276 (1936

Nov.); pt. 2, vol. 8, pp. 121–130 (1937 May).

[3] J.-C. Hentsch, "La Fidelité des Haut-Parleurs dans la Reproduction des Phénomènes Transitoires," *Tech. Mitt.-PTT* (Bern), no. 6 (1951 June).

[4] H. D. Harwood, "The Hows, Wheres and Whys of Testing High Quality Loudspeakers," *Audio*, vol. 55, pt. 1, pp. 16–20 (1971 Aug.); pt. 2, pp. 18–21 (1971 Sept.).

[5] H. D. Harwood, "Some Factors in Loudspeaker Quality," *Wireless World*, vol. 82, pp. 45–54 (1976 May).

[6] C. D. Mathers, "Design of the High-Level Studio Monitoring Loudspeaker Type LS5/8," British Broadcasting Corp., Rep. BBC-RD-1979/22 (1979 Nov.).

[7] M. Colloms, *High Performance Loudspeakers*, 3rd ed. (J. Wiley, New York, 1985).

[8] D. E. L. Shorter, "A Survey of Performance Criteria and Design Considerations for High-Quality Monitoring Loudspeakers," *Proc. IEE*, vol. 105, pt. B, pp. 607–621 (1958 Nov.); reprinted in *J. Audio Eng. Soc.*, vol. 7, pp. 13–28 (1959 Jan.).

[9] J. Moir, "Speaker Directivity and Sound Quality," *Wireless World*, vol. 85, pp. 61–63, 96 (1979 Oct.).

[10] J. Moir, "Loudspeakers in Rooms, Which Response," *Hi-Fi News Rec. Rev.*, vol. 27, pp. 22–25 (1982 June).

[11] P. J. Walker, "The Loudspeaker in the Home," *J. Brit. Inst. Radio Eng.*, vol. 13, pp. 377–380 (1953 July).

[12] R. E. Cooke, "Misleading Measurements," *Hi-Fi News Rec. Rev.*, vol. 21, (1976 Oct.); reprinted in *KEFTOPICS*, vol. 1, no. 3, KEF Electronics Ltd., Maidstone, England.

[13] D. Queen, "Relative Importance of the Direct and Reverberant Fields to Spectrum Perception," *J. Audio Eng. Soc. (Project Notes/Engineering Briefs)*, vol. 21, pp. 119–121 (1973 March).

[14] D. Queen, "The Effect of Loudspeaker Radiation Patterns on Stereo Imaging and Clarity," *J. Audio Eng. Soc.*, vol 27, pp. 368–379 (1979 May).

[15] R. C. Heyser, "Breakthrough in Speaker Testing," *Audio*, vol. 57, pp. 20–30 (1973 Nov.).

[16] M. Komamura, K. Tsuruta, and M. Yoshida, "Correlation between Subjective and Objective Data for Loudspeakers," *J. Acoust. Soc. Japan*, vol. 33, pp. 103–115 (1977 Mar.).

[17] V. Brociner and D. R. von Recklinghausen, "Speaker System Design Using a Reverberation Chamber," presented at the 34th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 16, p. 342 (1968 July), preprint 579.

[18] A. G. Bose, "On the Design, Measurement and Evaluation of Loudspeakers," presented at the 35th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 17, p. 86 (1969 Jan.), preprint 622.

[19] A. G. Bose, "Sound Recording and Reproduction, Part 1: Devices, Measurements, and Perception," *Technol. Rev. (MIT)*, pp. 19–25 (1973 June); "Part 2: Spatial and Temporal Dimensions," *ibid.*, pp. 25–33 (1973 July/Aug.).

[20] C. L. McShane, "The Meaning of Quantitative Loudspeaker Measurements," presented at the 36th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 17, p. 344 (1969 June), preprint 651.

[21] R. F. Allison and R. Berkovitz, "The Sound Field in Home Listening Rooms," *J. Audio Eng. Soc.*, vol. 20, pp. 459–469 (1972 July/Aug.).

[22] "Loudspeaker Accuracy: CU's Tests," *Consumer Rep.*, vol. 38, pp. 456–457 (1973 July).

[23] E. Torick, "In the Loudspeaker Testing Lab," *High Fidelity*, vol. 27, pp. 69–73 (1977 Oct.).

[24] J. D Hirsch, "Testing Speakers," *Stereo Rev.*, vol. 47, pp. 24–25 (1982 Aug.).

[25] U. Rosenberg, "Loudspeaker Measurement and Consumer Information," Statens Provningsanstalt, 11486 Stockholm 5, Sweden, Rep. 244 (1973).

[26] H. Staffeldt, "Correlation between Subjective and Objective Data for Quality Loudspeakers," *J. Audio Eng. Soc.*, vol. 22, pp. 402–415 (1974 July/Aug.).

[27] G. W. Sioles, "Loudspeaker Measurements in Live Rooms," *J. Audio Eng. Soc.*, vol. 11, pp. 203–206 (1963 July).

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

[29] A. H. Benade, *Fundamentals of Musical Acoustics* (Oxford University Press, New York, 1976), chap. 12.

[30] H. Møller, "Relevant Loudspeaker Tests," Brüel and Kjaer Application Note 15-067; also "Relevant HiFi Tests at Home," presented at the 47th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 22, p. 272 (1974 May).

[31] R. H. Russell, "Speaker Evaluation: Ear or Machine?," *The Audio Amateur*, pt. 1, pp. 10–15, issue 1; pt. 2, pp. 20–25, issue 2 (1975).

[32] T. Holl, in conversation with J. Atkinson, "Speakers, Rooms and Response," *Hi-Fi News and Rec. Rev.*, vol. 28, pp. 40–43 (1983 May).

[33] R. Long, "Loudspeaker Testing and the Listening World," *High Fidelity*, vol. 31, pp. 18–19 (1981 June).

[34] R. B. Schulein, "In Situ Measurement and Equalization of Sound Reproduction Systems," *J. Audio Eng. Soc.*, vol. 23, pp. 178–186 (1975 Apr.).

[35] D. Preis, "Linear Distortion, Measurement Methods and Audible Effects: A Survey of Existing Knowledge," presented at the 2nd International Conference of the Audio Engineering Society, Anaheim, California, 1984 May 11–14, preprint C1005.

[36] B. B. Bauer, "Audibility of Phase Distortion," *Wireless World*, vol. 80, pp. 27–28 (1974 Mar.).

[37] H. D. Harwood, "Audibility of Phase Effects in Loudspeakers," *Wireless World*, vol. 82, pp. 30–33 (1976 Jan.).

[38] J. Moir, "Phase and Sound Quality," *Wireless World*, vol. 82, pp. 80–84 (1976 Mar.).

[39] S. P. Lipshitz, M. Pocock, and J. Vanderkooy, "On the Audibility of Midrange Phase Distortion in Audio Systems," *J. Audio Eng. Soc.*, vol. 30, pp. 580–595 (1982 Sept.).

[40] T. A. Saponas, R. C. Matson and J. R. Ashley, "Plain and Fancy Test Signals for Music Reproduction Systems," *J. Audio Eng. Soc.*, vol. 19, pp. 294–305 (1971 Apr.).

[41] D. Clark, "Measuring Audible Effects of Time Delays in Listening Rooms," presented at the 74th Convention of the Audio Engineering Society, *J. Audio*

Eng. Soc. (Abstracts), vol. 31, p. 972 (1983 Dec.), preprint 2012.

[42] J. Bowers and S. Roe, "Phase and Loudspeakers," *Hi-Fi News Rec. Rev.*, vol. 21, pp. 56–61 (1976 Apr.).

[43] M. A. Gerzon, *Wireless World (Letter to the Editor)*, and reply by Harwood, vol. 82, pp. 60–61 (1976 Mar.).

[44] D. A. Barlow, "Loudspeaker Coloration," *Wireless World*, vol. 84, pp. 34–36 (1978 Mar.).

[45] R. Bücklein, "The Audibility of Frequency Response Irregularities," *J. Audio Eng. Soc.*, vol. 29, pp. 126–131 (1981 Mar.), a translation of a 1962 original paper in German.

[46] W. R. Stevens, "Loudspeakers—Cabinet Effects," *Hi-Fi News Rec. Rev.*, vol. 21, pp. 87–93, 97 (1976 Sept.).

[47] P. Fryer, "Loudspeaker Distortions—Can We Hear Them?," *Hi-Fi News Rec. Rev.*, vol. 22, pp. 51–56 (1977 July).

[48] J. M. Berman and L. R. Fincham, "The Application of Digital Techniques to the Measurement of Loudspeakers," *J. Audio Eng. Soc.*, vol. 25, pp. 370–384 (1977 June).

[49] C. P. Janse and A. J. M. Kaizer, "Time-Frequency Distributions of Loudspeakers: The Application of the Wigner Distribution," *J. Audio Eng. Soc.*, vol. 31, pp. 198–223 (1983 Apr.).

[50] D. E. L. Shorter, "Loudspeaker Transient Response—Its Measurement and Graphical Representation," *BBC Quart.*, vol. 1, pp. 1–9 (1946 Oct.).

[51] M. S. Corrington, "Correlation of Transient Measurements on Loudspeakers with Listening Tests," *J. Audio Eng. Soc.*, vol. 3, pp. 35–39 (1955 Jan.).

[52] R. J. Larson and A. J. Adducci, "Transient Distortion in Loudspeakers," *IRE Trans. Audio*, pp. 79–85 (1961).

[53] R. Driscoll, "New Transient Test for Loudspeakers," *The Gramophone*, pp. 789–790 (1974 Oct.).

[54] P. Unwin, "Loudspeaker Coloration," *Hi-Fi News Rec. Rev.*, vol. 19, pp. 71–75 (1974 Mar.).

[55] G. A. Briggs, *Loudspeakers*, 5th Ed. (Wharfedale Wireless Works, Idle, Bradford, Yorkshire, 1958).

[56] J. Blauert, *Spatial Hearing* (1974); transl. by J. S. Allen (M.I.T. Press, Cambridge, MA., 1983).

[57] P. Kantrowitz, "Distortion of High-Frequency Loudspeakers," *J. Audio Eng. Soc.*, vol. 10, pp. 310–317 (1962 Oct.).

[58] R. Cabot, "Perception of Nonlinear Distortion," presented at the 2nd International Conference of the Audio Engineering Society, Anaheim, California, 1984 May 11–14, preprint C1004.

THE AUTHOR

Floyd E. Toole was born in Moncton, New Brunswick, in 1938. He received the B.Sc. degree in electrical engineering in 1960 from the University of New Brunswick and the Ph.D. and D.I.C. in electrical engineering in 1965 from the Imperial College of Science and Technology, University of London, U.K. Since then he has been with the acoustics section, division of physics, National Research Council, Ottawa.

Dr. Toole's early research was concerned with sound localization and the mechanisms of binaural hearing. After an interval of activity in the measurement and control of noise, including organizational and standards-writing work with the Canadian Standards Association, he returned to audio. In recent years he has been involved with loudspeakers, rooms, and listening tests. A routine program of measurements and listening tests is regularly used by loudspeaker manufacturers, acoustical consultants, and audio publications for purposes that range from product design to product reviewing. A parallel, research-oriented effort is aimed at improving the precision and utility of measurements and listening tests. Part of this energy has been put into working groups of the International Electrotechnical Commission where he is active in standards writing for loudspeaker measurements, listening tests, headphones, and amplifiers. In the field of professional audio, Dr. Toole has designed recording studios, control-room monitor loudspeakers and sound reinforcement systems for large concert halls and theaters.

Dr. Toole is a member of the Audio Engineering Society, the Acoustical Society of America, and the Canadian Acoustical Association.