

The Modification of Timbre by Resonances: Perception and Measurement*

FLOYD E. TOOLE AND SEAN E. OLIVE

National Research Council, Division of Physics, Ottawa, Ont. K1A 0R6, Canada

Resonances are fundamental to the production of musical pitch and timbre. They are also the principal source of coloration when they are added in the processes of sound recording and reproduction. The traditional problem in the design and evaluation of audio products has been to find the measurements necessary to recognize the presence of a resonance, the interpretation necessary to characterize its audibility, and the judgment of how much its form must be modified in order for it not to cause objectionable coloration. A review of previous work and new experimental results describe the thresholds of audibility of resonances as a function of frequency, Q , relative amplitude, time delay, program material, listener hearing performance, loudspeaker directivity, and reverberation added during recording or reproduction. The findings are discussed in terms of the measured amplitude and time responses of the systems through which the audio signal is passed. While the emphasis is on reproduced sound, there are some interesting relationships to the perceived timbre of sound in live performances.

0 INTRODUCTION

Resonances are the fundamental sound production mechanisms in musical instruments and the human voice. In musical instruments, high- Q resonances are the pitch-determining elements in vibrating strings, bars, columns of air, cavities, or electronic circuits. Superimposed on these basic sounds are characteristics of other resonances, such as those in the wooden panels of enclosures of various shapes and sizes, coupled with the air resonances within those enclosures. These processes combine to yield the melodic lines and harmonies of music, while giving us the clearly distinctive timbres of the various musical instruments.

In the voice, the resonant vibrations of the vocal cords set the rate at which puffs of air are released from the lungs. These pulses of air drive another resonant system consisting of the pipe extending from the larynx to the mouth, with various modifications due to the sinus cavity and others formed by the tongue, teeth, and lips. Together they generate the set of resonances known as formants, which help give voices their individual characters and permit the modification of those sounds in ritualized ways so that information is communicated.

In recording studios modifications of timbre are cre-

ated by various forms of signal processing. Artificial reverberation, echoes, phasing, flanging, specialized microphones, and equalization all generate colorations due to resonances and interference. Since they are added deliberately, for artistic reasons, they are an integral part of the original performance and must be preserved through the remainder of the record-replay chain.

Any of the numerous devices used in sound recording and reproduction can add resonances or delayed sounds to the original program with a consequent modification of the timbre of the original sound. The effects are frequently heard and usually detract from the naturalness of the sound. Mounting evidence indicates that resonances are, in fact, the major source of coloration in sound recording and reproduction [1], [2].

The sources of such resonant colorations are numerous, and some are difficult, if not impossible, to eliminate completely. It therefore is important to know the limits of perceptibility of these supplemental resonances, so that money and effort expended in their control can be spent efficiently.

1 THE AUDIBILITY OF RESONANCES—A REVIEW

In assessments of the audibility of resonances, one is almost inevitably listening through components or to program material containing unknown amounts of the very defect that is under examination. This means

* Presented at the 83rd Convention of the Audio Engineering Society, New York, 1987 October 16–19; revised 1987 December 30.

that the detection thresholds for resonances are masked thresholds; the ability to hear a resonance contributed by one component may depend on the relative levels of similar resonances from other sources. In addition, the threshold will certainly depend on the signal type, its spectral content, bandwidth, sound level, and the temporal relationship between the resonance and the driving signal.

Few studies have specifically addressed the audibility of resonances or the significance of spectral irregularities in general. Bücklein [3] appears to have been one of the first. He explored the relative audibility of spectral irregularities in samples of speech, music, and white noise presented to listeners through headphones. He concluded that, in general, peaks in the frequency response are more easily heard than the equivalent dips, and that both peaks and dips become more audible as their width increases. He also found that these phenomena were difficult to hear using solo instruments as test sounds, since they were audible only when the frequency of the resonance and the musical tones coincided. Even then several of his listeners had to be

coached in order for the effects to be heard. Sounds with continuous spectra, especially white noise, were found to be most revealing of irregularities in frequency response.

More recently, Fryer [2], [4] reported a detailed determination of the detection thresholds for resonances added to different program material. The results of the study, which examined the effect of both Q and frequency, are summarized in Fig. 1. As shown in Fig. 1(a), the measured parameter is the difference between the spectrum level of the program and the maximum steady-state output from the added resonance. Based on this measure, one concludes that the listeners were most sensitive to resonances of low Q , with the detectability decreasing approximately 3 dB for each doubling of the Q value. It was also found that all resonances were most easily heard with white noise as a test signal, with reduced sensitivity when using classical (symphonic) music, and with much reduced sensitivity when using popular music. These observations are consistent with those of Bücklein and also with those of Stevens [5], who did similar tests in the course

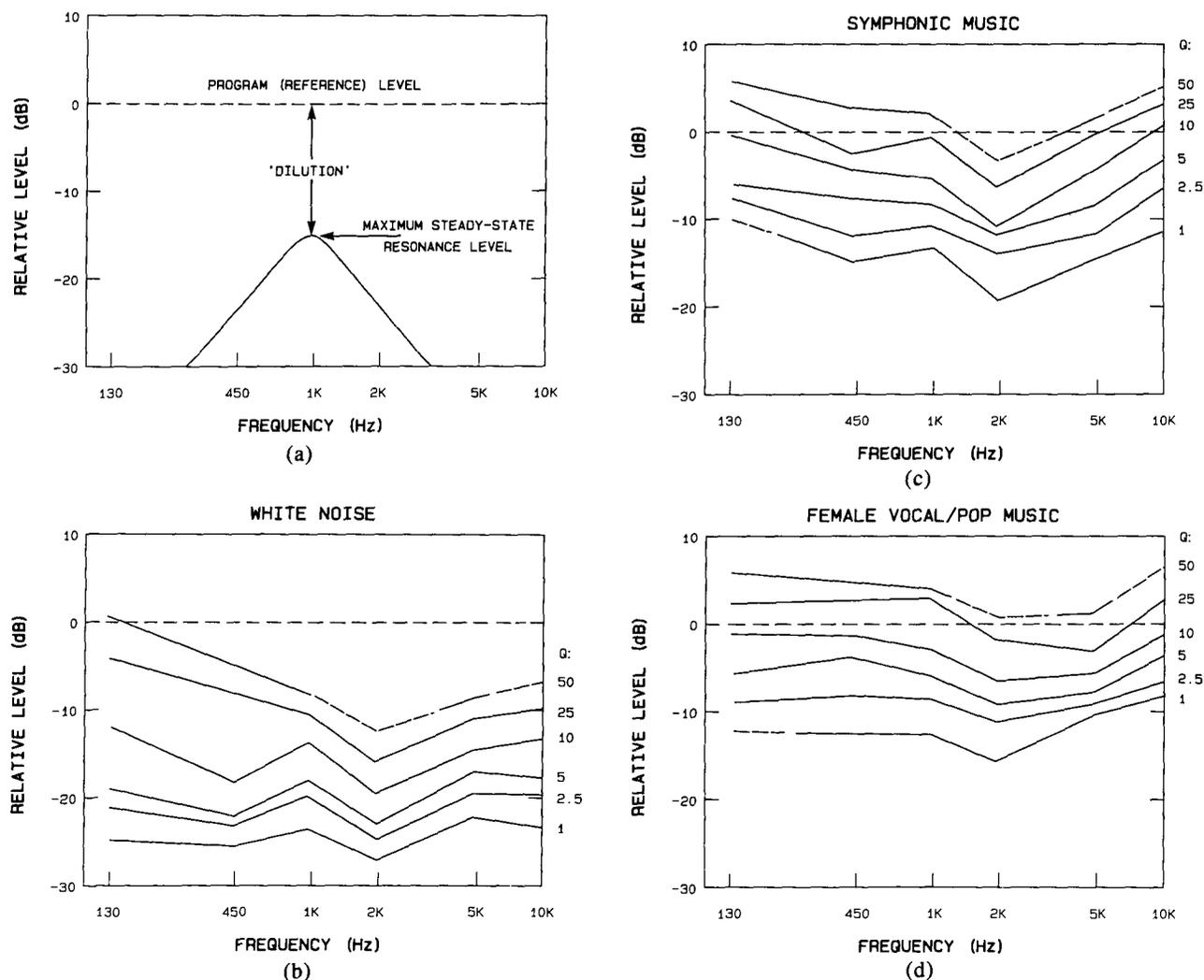


Fig. 1. Detection thresholds for nondelayed resonances of various Q when auditioned using (b) white noise, (c) classical music, and (d) popular music. Curves show the maximum steady-state output from the resonance using the spectrum level of the program material as a reference, as shown in (a). (From Fryer [2], [4].)

of investigating the importance of resonances in loudspeaker enclosures.

Moulana [6] presented a very detailed theoretical analysis of the various forms of resonances that can be encountered in practice, and described the results of extensive subjective tests. The data and discussions are thought provoking and provide some useful insights into the perceptual processes that may be involved. The massive effort is, however, frustratingly short of conclusions that can be applied directly to the design or evaluation of audio devices.

In summary, these findings indicate that irregularities in frequency response and phase response, and the corresponding waveform modifications, can be caused by resonances within an audio system, adding to the total sound output over a narrow range of frequencies. As judged by their maximum steady-state levels, low- Q resonances, producing broad peaks in the response curve, are more easily heard than high- Q resonances that result in quite narrow peaks. A resonance with $Q = 1$, for example, can be heard in noise when its maximum steady-state level is 25 dB below the spectrum level of the program, while one with $Q = 50$ can approach to within 10 dB (even less at low frequencies) of the spectrum level before being heard. With music the detection thresholds are substantially higher. Stevens comments that the detection of a change, as is the case for threshold determinations, need not correspond with a degradation of sound quality. His listeners sometimes felt that low-frequency resonances improved the signal, particularly speech.

Moulana distinguished between driven-state colorations, related to amplitude-response irregularities, transient-state colorations, related to the decaying oscillation after a driving signal is removed, and interstate colorations, related to the amplitude step and frequency shift that can accompany a transition between driven and transient states. He notes that pink noise seems to be most revealing of driven-state colorations, while speech and music are more likely to reveal transient and interstate colorations. Resonances with Q values less than 10 appear to contribute driven-state colorations, while those with Q values greater than 50 tend to be revealed through transient-state colorations. There was some evidence that pitch shift, an interstate coloration, tends to be most audible in the frequency range between 125 and 500 Hz.

As has been discussed previously [1], the finding that the audibility of resonances decreases with increasing Q means that, in the time domain, the duration of ringing is an unreliable indicator of potential coloration. In those loudspeaker evaluations it was found that the presence of resonances could be inferred from irregularities in measurements of either amplitude or phase as functions of frequency. Moulana, in his study, considered the importance of the phase-response irregularities as an independent source of coloration and concluded that "the subjective effect of these local irregularities is negligible if not absent in the first place" [6]. This implies that, while either amplitude or phase

measurements can indicate the presence of resonances, the amplitude response appears to be more directly related to the audible effect.

2 EXPERIMENTS

While some basic facts about the audibility of resonances have been established by the earlier work, other issues suggested the need for additional experiments. First, however, a reference was established by repeating the experiments performed in the earlier work.

All experiments that follow were performed using the equipment shown in Fig. 2. The listener was able at any time to select the unadulterated signal, as a basis for comparison, or the signal with the resonance at maximum level to refresh the memory as to the nature of the relevant coloration, or the signal with the resonance added at the level set by the listener. The task was to adjust the level of the resonance until it was at threshold. Listeners were instructed to start at a level at which the resonance was clearly audible and to reduce the level of the resonance until it just disappeared (that is, there was no difference between the signal without the resonance and the signal with the resonance). At that point the level was to be increased until the coloration just became audible. Between these two settings is the true threshold, but for the purposes of this test, listeners were encouraged to establish a criterion for detection, such as 'just audible' or 'just not audible,' and to use it consistently. The knob used by the listener was attached to a multiturn potentiometer which eliminated knob position as a possible feedback cue. Listeners were given as much time as they needed to make their adjustments, returning at will to review the sound of the pure signal or the resonance. It is believed that the resulting threshold levels are close to the minimum attainable. With only a little practice, the listeners made repeated threshold adjustments with standard deviations

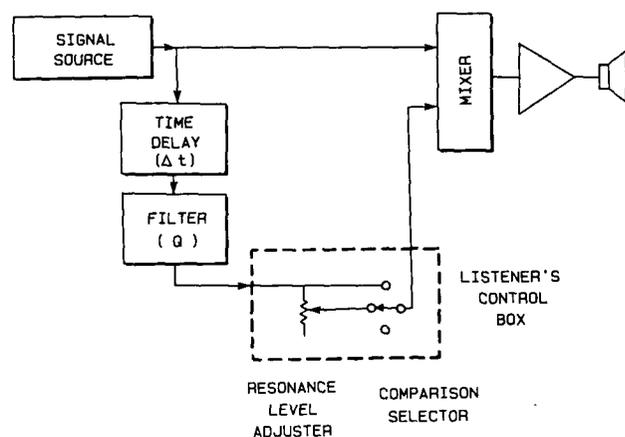


Fig. 2. System used for generating resonances and adding them, with or without a time delay, to a signal. The level of the resonance relative to the pure signal is determined by the listener, who has control of a multiturn adjuster. With the comparison selector the listener can choose at will to listen to the signal with the resonance at its maximum level (top position) or at its adjusted level (middle position), or to the signal alone (bottom position).

that were in the range of 2–4 dB under most signal and listening conditions.

All of the experiments involved two to six subjects, to confirm the repeatability of the effects being examined. Some of the results are shown as group averages. In other cases the results are shown for a single listener so that comparisons can be made across a group of experiments. In no case were the individual differences among the listeners great enough to alter the conclusions drawn from the experimental results.

The loudspeakers and headphones used in the experiments were all high-quality high-fidelity units exhibiting very little evidence of the resonant colorations that were the subject of investigation. All were highly rated in subjective and objective evaluations of sound quality [1].

2.1 Detection Thresholds for Resonances—A Comparison

Using pink noise, thresholds were determined for two subjects who listened, in a quiet anechoic chamber, to a single loudspeaker at 2 m, at ear level, in the forward direction. The results were similar for both listeners, and the averaged results are shown in Fig. 3 along with comparable results taken from the works of Fryer [2], [4], Moulana [6], and Harwood (reported in [6]). Given the opportunities for differences between experiments of this kind, the results show good agreement. Among other differences it should be noted that the present results and those of Moulana were obtained under anechoic listening conditions, while those of Fryer and Harwood were obtained in normal semireverberant rooms. In addition, Fryer used white noise, while the other studies used pink noise.

Sample tests with choral and symphonic music, jazz, anechoically recorded speech, and rock music all yielded elevated threshold levels, similar to those found by Fryer (Fig. 1), although no clear pattern of relative sensitivity was apparent among the different programs.

2.2 Frequency Domain versus Time Domain

The process involved in the perception of resonances raises questions regarding the relative importance of spectral and time-domain cues. Experiments were undertaken in which the matter was addressed directly. The confusion of reflected sounds in a semireverberant room would likely reduce the audibility of phenomena that are strongly related to waveform or to the sequence of events in the time domain. Listening in an anechoic chamber represents the opposite condition. The sustained sounds within noise, speech, or music are likely to mask much of the information in resonant decays. Using narrow impulses (clicks) at a low repetition rate represents the opposite condition.

Four experiments were conducted. In the first pair, two listeners made threshold adjustments with pink noise and with 10- μ s pulses at a repetition rate of 10 pulses per second while listening to a single loudspeaker in an anechoic chamber. The second pair of experiments differed only in that the listening was done in a normal

semireverberant room of domestic proportions (an IEC-recommended listening room [7]: 2.8 m high by 6.7 m long by 4.1 m wide, 77 m³, RT₆₀ = 0.34 ± 0.08 s from 250 Hz to 4 kHz).

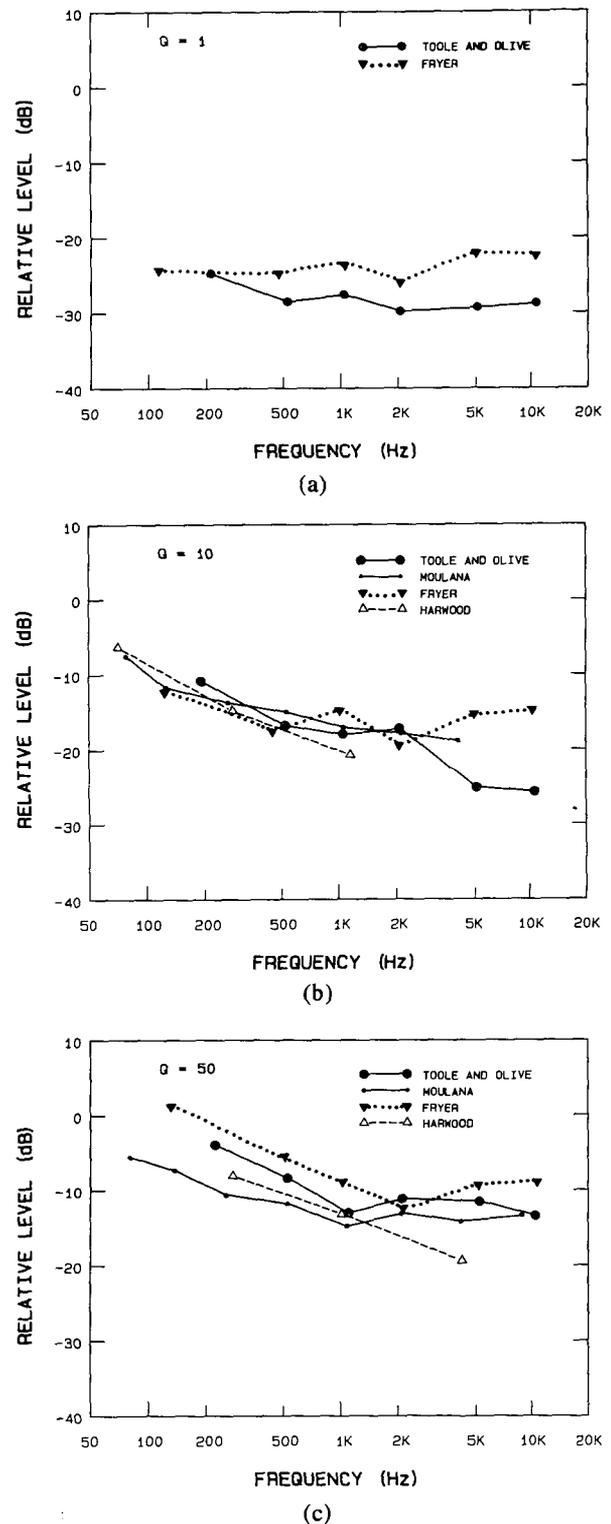


Fig. 3. Detection thresholds for a single nondelayed resonance as determined in the present study and in studies by Moulana [6], Fryer [2], [4], and Harwood (reported in [6]). Not all data were available from all studies. Curves show the maximum steady-state output from the resonance using the spectrum level of the program material as a reference, as shown in Fig. 1(a).

Since the present interest is primarily in the differences or trends in perception as specific experimental parameters are changed, the results are shown as threshold shifts. These indicate the direction and magnitude of threshold changes as a result of making the designated changes in the experimental conditions.

Fig. 4(a) shows the threshold shift as a function of signal type, pulses versus noise, for anechoic listening. It shows that there is relatively little difference for resonances of high Q , but some substantial increases in sensitivity exist for lower Q resonances when the signal is pink noise.

Fig. 4(b) shows the result of the same test performed in the standard listening room. This time there is little difference in the audibility of resonances of any Q at lower frequencies, but at high frequencies the thresholds drop rapidly. Looking for a connection between these first two results, it seems plausible that increased sensitivity with pink noise may be related to listening in a predominantly direct sound field. The direct sound field is dominant at all frequencies in the anechoic chamber and due to tweeter directivity, at high fre-

quencies in this listening room/loudspeaker situation, as illustrated in [1, fig. 18].

In Fig. 5(a) it can be seen that with pink noise the resonances are about equally well revealed either in a free sound field or in a normal room. In contrast, Fig. 5(b) shows that, with pulses, medium- and low- Q resonances are better revealed in a normal room. It is obvious that there are significant effects here, but the underlying mechanism is not immediately evident.

There is a suggestion that the result could be related to the temporal continuity of the signals. On this basis, the most "continuous" resonance signals [pink noise at all Q values and pulses with $Q = 50$ (prolonged ringing)] appear to be about equally audible in either listening environment. The "discontinuous" resonance signals (pulses with Q values of 10 and 1) are more audible in the normal room. A plausible explanation suggests that the reflections and the reverberation in the normal room have the effect of temporally extending the discontinuous events, rendering them more "continuous." Therefore the detection of some resonances is improved when the driving signal itself is continuous

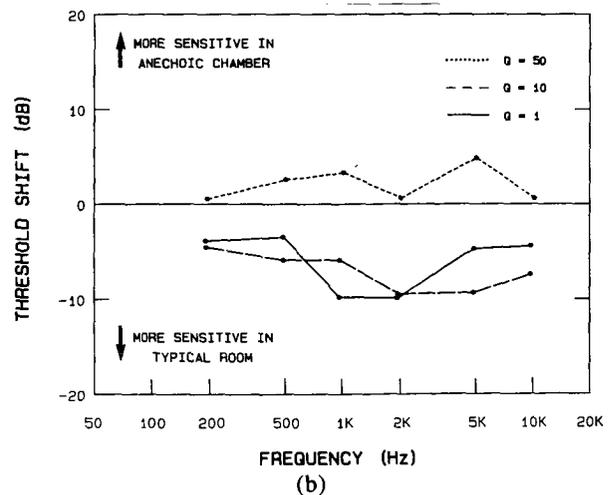
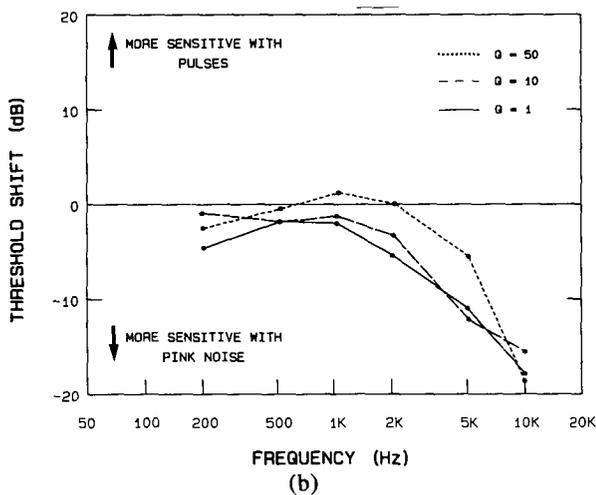
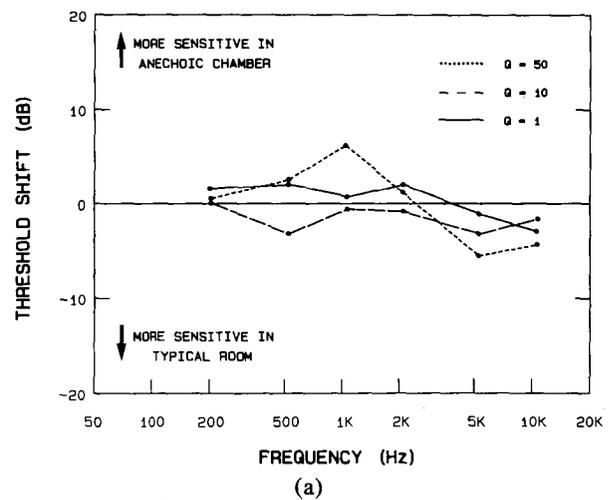
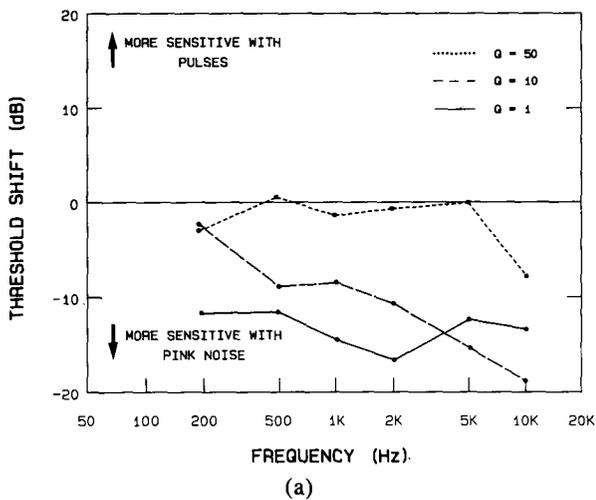


Fig. 4. Experimental results showing the shift in the detection threshold for nondelayed resonances as a function of signal type: pink noise versus pulses, for listening done in (a) an anechoic chamber and (b) a typical listening room. Mean results for two listeners.

Fig. 5. Experimental results showing the shift in the detection threshold for nondelayed resonances as a function of listening environment: anechoic chamber versus typical domestic listening room, while listening to (a) pink noise and (b) pulses. Mean results for two listeners.

or when the listening environment adds repetitions in the form of reflections.

The apparently anomalous behavior of the $Q = 50$ resonance in Fig. 4(a) may be explained in the observation that the output from such a resonance, driven by the 10/s pulses, is already sufficiently "continuous." Driving it with random noise, therefore, would not make a significant improvement. The shape of the $Q = 10$ curve supports such a hypothesis in that it seems to represent a transitional situation. At high frequencies, the moderate ringing from this resonance dies away long before the next pulse (a discontinuous signal). Moving lower in frequency, the ringing is sustained for a longer time interval, making the resonance progressively more "continuous" relative to the repetition period of the pulses. This result suggests that, under certain circumstances, the frequency with which a resonance is driven by impulsive sounds will influence its audibility. Increasing the repetition frequency should yield lower thresholds.

2.3 Impulsive Sounds—Effects of Repetition and Reverberation

Having established that there are some effects related to the listening environment and possibly to the repetition frequency of the signal, the following experiments addressed the matters directly.

The listening conditions were chosen as follows:

- 1) *Headphones*: Direct coupling to the ears, no reflections or reverberation
- 2) *Anechoic chamber* (loudspeaker at 2 m): Predominantly direct sound, only local reflections from listener's body, no reverberation
- 3) *Small hall—large semireverberant room* (900 m³, RT₆₀ ≈ 2 s, loudspeaker at 8 m): Predominantly reverberant sound.

Resonances with Q values of 1 and 50 were used, to explore both extremes of that variable, at a center frequency of 1 kHz. The electric signal consisted of 40- μ s rectangular pulses at repetition frequencies ranging from 1 to 20 pulses per second. Four listeners with normal hearing participated.

The results showed some important similarities and differences. All listeners exhibited similar trends in threshold as functions of the two variables, listening condition and repetition frequency. They differed in absolute sensitivity. Three of the listeners yielded thresholds that were about 10 dB lower than the fourth listener. The fact that this listener had the least prior experience may have been a factor.

Fig. 6 shows the group-averaged thresholds. For $Q = 1$ the results show a clear increase in sensitivity as the sound at the listeners' ears is modified by increasing reflections and reverberation. The clear, albeit small, difference between headphones and anechoic listening is an indicator of the strength of this variable. The effect of adding large-room reverberation is substantial, increasing the listeners' sensitivity to low- Q resonances by as much as 10 dB at low repetition rates. As a function of increasing repetition frequency two

effects can be seen. First, the threshold levels drop gradually, supporting the idea that signal repetitions are a factor. Second, the curves tend to converge at high repetition rates, supporting the idea that signal repetitions at the source or repetitions caused by reflections in the listening environment have a similar effect.

Fig. 7 shows comparable data for a resonance with $Q = 50$. The result shows a relative independence of listening condition and a reduced sensitivity to repetition frequency, compared to Fig. 6. This supports the observations from Figs. 4 and 5, even to the detail that anechoic listening seems to be very slightly better at revealing the presence of high- Q resonances than other listening conditions [see Fig. 5(b)].

If it is the repetitions in the listening room reverberation that are responsible for improving the sensitivity to small changes in signal spectrum, then there should be a corresponding effect when the reverberation is added to the signal. Two experiments were conducted to test this notion.

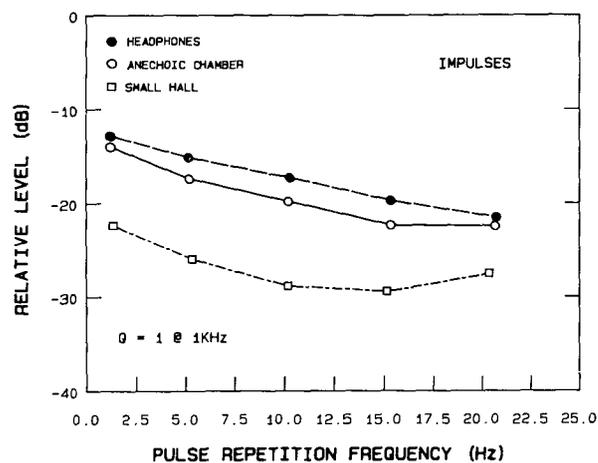


Fig. 6. Detection thresholds for resonances with $Q = 1$ at 1 kHz added without time delay to impulses presented at various repetition frequencies. The combined sounds were auditioned through headphones, and through a loudspeaker in an anechoic chamber and a small hall. Mean results for four listeners.

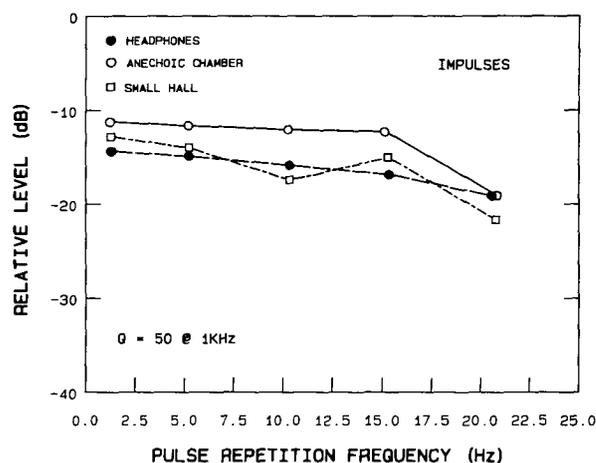


Fig. 7. As Fig. 6, but using $Q = 50$ at 1 kHz.

In the first experiment, impulses at a rate of one pulse per second were modified by resonances in the apparatus shown in Fig. 2. The modified signal was then fed through a reverberation synthesizer (Yamaha DSP-1) before being auditioned through headphones. The parameters of the synthesizer were set as follows: chamber, high = 0.7, initial delay = 5 ms, HPF = thru, LPF = 10 kHz, Rev. Lvl. = 100%. The output was taken from the front-channel mixture.

The reverberation time was varied in steps from 0.3 to 6.0 s and detection thresholds were determined for the same resonance used in the previous experiments ($Q = 1$ at 1 kHz). The result was that, compared to the signal with no reverberation, the detection threshold dropped by about 9 dB for $RT_{60} = 0.3$ s, to plateau at a level about 14 dB lower for RT_{60} in excess of about 1 s. This situation simulates the case in a recording studio in which a signal is modified by a resonance in, for example, a microphone or equalizer, and is then fed through a reverberation device. Clearly, adding reverberation at that stage increases the audibility of the resonance.

The logical variation of this situation is one that simulates the addition of resonant coloration after the signal has been modified by reverberation. In this experiment the reverberation simulator was placed between the signal source and the apparatus for adding the resonance. All other experimental variables were the same as in the previous experiment. The result also was very similar, with the threshold dropping rapidly to a plateau about 14 dB below that for the signal with no reverberation. A reverberation time of less than 1 s was required for the maximum effect to set in.

In summary, for impulsive signals, modifications in the sound spectrum due to the addition of nondelayed low- Q resonances, or the corresponding frequency-response equalization, are more audible if reverberation is added at any stage in the processes of recording or reproduction.

2.4 Effect of Time Delays

In the literature the term "delayed resonances" has been used to describe resonances of the kind already discussed that, in fact, are not delayed. Presumably this misnomer reflects a confusion between true delay and the resonance overhang, or ringing. Introducing a genuine time delay between the driving signal and the output from the resonance is quite different. In reality such delays exist because of propagation time between the driving device and the resonating element. A loudspeaker system, for example, might have a resonant panel that is excited after the small time delay required for the sound to travel through the enclosure (ignoring mechanical coupling). That resonance output might be further delayed by the additional length of the propagation path from the resonating enclosure panel, reflecting off the room boundary and reaching the ear of a listener several milliseconds after the direct sound. Resonant elements in the listening room, separate from the loudspeaker, also present many opportunities for

delayed resonances.

The summation of signals separated by a time delay produces interference patterns, or "comb filtering." If the delayed sound occupied only a narrow frequency band (high Q), there would be little or no evidence of the familiar comb pattern, for broadband low- Q resonances there could be. It is of separate interest to see if there is a gradation from narrow band to broadband in the perceptual effects of these delayed sounds.

The strong differences between the perceptual effects of continuous and discontinuous signals directed the choice of signals for these experiments: continuous pink noise and impulses at a rate of 10 per second. The delays, which varied from zero to 60 ms, were provided by a digital delay unit (Klark-Teknik DN700) with performance adequate to render it audibly transparent except for the delay itself. Three listening conditions were used:

- 1) Headphones
- 2) A typical domestic room (see Sec. 2.2 for details)
- 3) A small hall [see Sec. 2.3, item 3), for details].

These conditions presented the listener with signals that were contaminated by reflections and reverberation in progressively increasing quantity and duration. A Q of 1 was used throughout, since low- Q resonances are the ones most affected by these variables.

The results reveal strongly different perceptions for the two signals. Fig. 8 shows the result for pink noise in headphone listening. The thresholds were lowest for resonances coincident with the signal itself (zero added time delay), and they were elevated progressively as a function of added time delay. Fig. 9 shows the comparable result for impulses. The thresholds were highest at zero time delay and decreased with added time delay—precisely the opposite trend from that seen with pink noise. Readers are cautioned to note carefully the time scales on these and subsequent graphs, as they have been modified to most clearly display the effects.

These differences between pink noise and impulses also extend to frequency dependence. With pink noise,

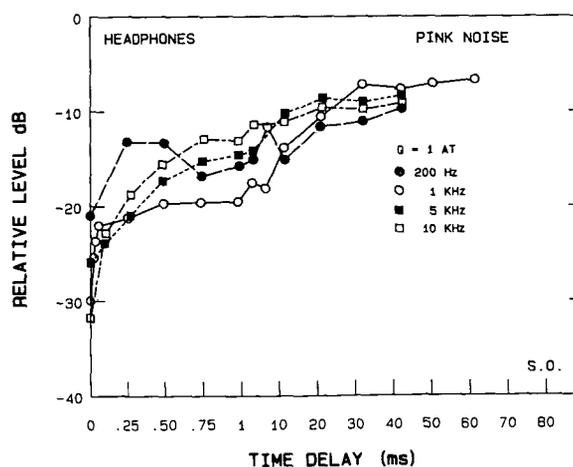


Fig. 8. Detection thresholds for delayed resonances auditioned through headphones using pink noise. Resonances with $Q = 1$ and various center frequencies were added at delays from zero to 60 ms.

the frequency of the resonance was relatively unimportant. In contrast, with impulses a strong frequency dependence was revealed, with a higher resonance frequency resulting in a more rapid drop in threshold as a function of time delay.

Adding the listening condition as a variable, Fig. 10 shows that, for pink noise, there was only a slight effect. At time delays less than about 10 ms, headphone listening was a little better at revealing resonances, but at larger delays, all three listening conditions yielded similar results.

Fig. 11 shows the comparable result for impulses, illustrating a significant environmental effect. At time delays less than about 1 ms the result was similar to that seen in Fig. 5(b), with the resonance detection substantially improved by the addition of reverberation. With resonances delayed by more than about 1 ms the situation was reversed, with the more complicated sound fields elevating the detection thresholds. The added reflections of the simple impulse would appear to have

had the effect of bringing the discontinuous impulses and continuous pink noise perceptually closer. Fig. 12 illustrates the close similarity of the thresholds for these two very different signals when they were auditioned in the small hall. The results were also very similar for this listener and for the listener whose comparable results appear in Figs. 10 and 11.

2.5 Effects with Music and Speech

Pink noise and impulses are convenient signals to generate and manipulate, but while they are representative of important classes of sounds that appear in nature, they are not "real." In the following tests three natural sounds were used:

- 1) Speech recorded in an anechoic chamber
- 2) Speech recorded in a normal semireverberant room
- 3) Chamber music recorded in a concert hall.

The anechoic speech was clearly discontinuous, the "echoic" speech was less so, and the chamber music had a more or less continuous background of rever-

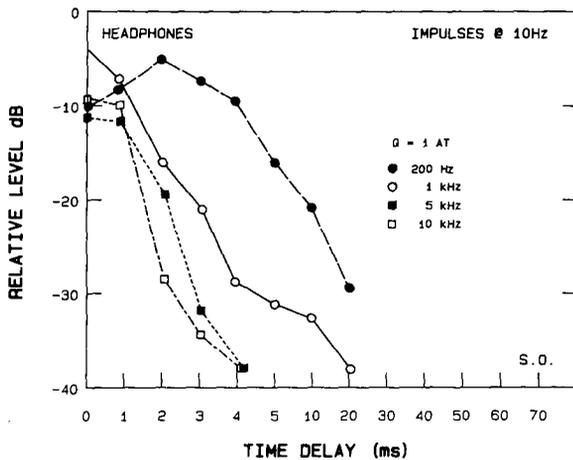


Fig. 9. Detection thresholds for delayed resonances auditioned through headphones using impulses with a 10-Hz repetition frequency. Resonances with $Q = 1$ and various center frequencies were added at delays from zero to 40 ms.

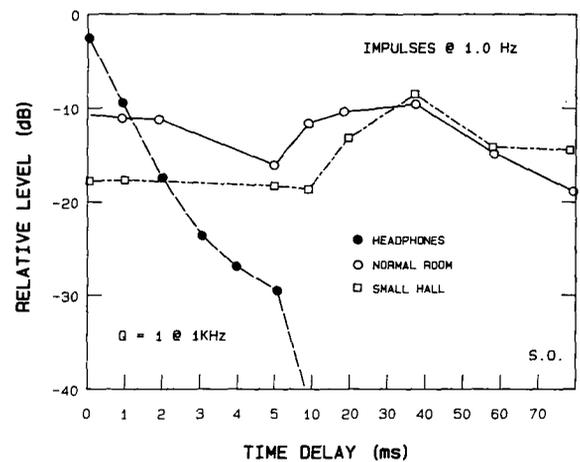


Fig. 11. As Fig. 10, but using impulses at a repetition frequency of 1.0 Hz.

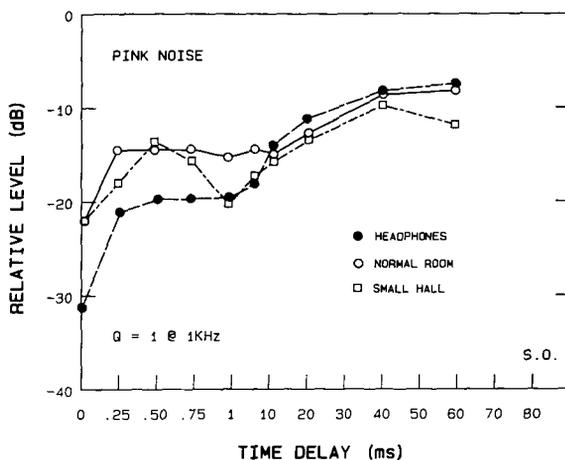


Fig. 10. Detection thresholds for delayed resonances with $Q = 1$ at 1 kHz auditioned using continuous pink noise presented through headphones and through a loudspeaker in a normal domestic room and in a small hall.

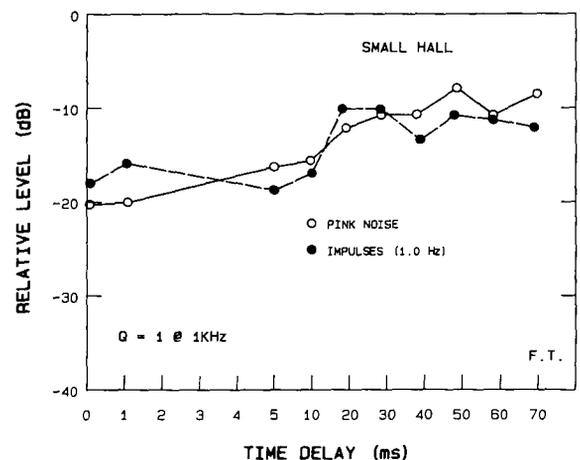


Fig. 12. Detection thresholds for delayed resonances with $Q = 1$ at 1 kHz auditioned through a loudspeaker in a small hall. Results are shown for tests using continuous pink noise and impulses at a repetition frequency of 1.0 Hz.

beration.

If the patterns of listener response continue, there should be some similarity between the experimental results achieved with impulses and the staccato anechoic speech, and between pink noise and the more continuous semireverberant speech and music. The results shown in Figs. 13–15 confirm such similarities. Taking into consideration the different time scales of the plots, the results of Fig. 11 (impulses) and those of Fig. 13 (anechoic speech) exhibit similar trends, while the results of Fig. 10 (pink noise) show the consistency and direction of those seen in Figs. 14 (echoic speech) and 15 (music).

In detecting the presence of small differences in timbre due to resonances it seems that, for resonances excited with delays less than about 1 ms, there might be a slight advantage to listening through headphones or in environments that have little reverberation. For resonances excited with delays longer than about 1 ms,

there is a clear advantage, but only with impulsive sounds or speech recorded without reverberation. Recordings containing normal amounts of reverberation, and listening in typical semireverberant environments, substantially reduce the audibility of delayed low- Q resonances, especially those delayed by more than about 1 ms.

The latter statement was reaffirmed by experiments in which the same sound, speech, was used with and without a reverberant background in the recording and with varying amounts of reverberation in the listening situation (Figs. 16–18).

2.6 Effect of Sound Spectrum

The data shown in Fig. 9 were of special interest in that there was clearly an effect attributable to frequency. In the general body of psychoacoustic data there are several examples of perceptions being dictated, under different circumstances, by signal bandwidth, or by

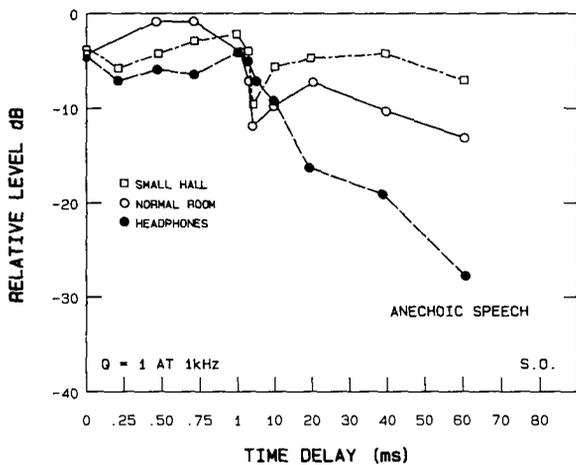


Fig. 13. First of a sequence of three figures illustrating the influence of signal type and reverberation on the audibility of delayed resonances presented through headphones and through a loudspeaker in a normal domestic room and in a small hall. This figure shows the threshold of audibility for delayed resonances with $Q = 1$ at 1 kHz when the signal was speech recorded in an anechoic chamber.

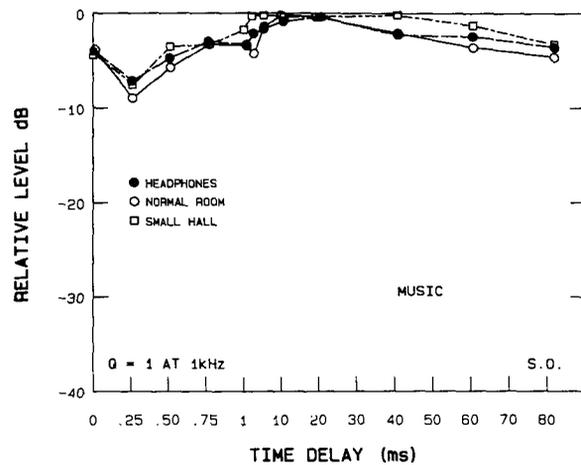


Fig. 15. As Fig. 13, but using music: a recorded excerpt of a chamber orchestra performing in a concert hall.

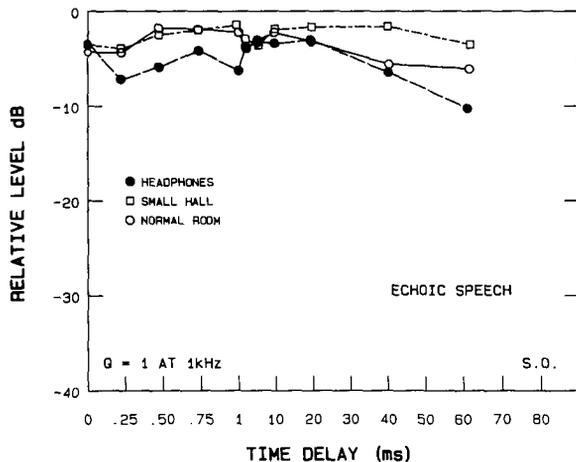


Fig. 14. As Fig. 13, but using echoic speech: a recording of speech including some natural room reverberation.

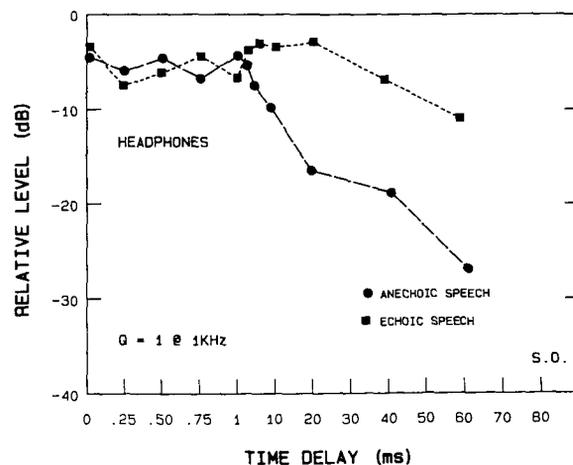


Fig. 16. First of a sequence of three figures illustrating the influence of listening conditions on the audibility of delayed resonances auditioned using speech recorded with and without natural reverberation. This figure shows the threshold of audibility for delayed resonances with $Q = 1$ at 1 kHz when the signal was presented through headphones.

upper or lower cutoff frequencies of band-limited signals. In the present example there is the additional ingredient of the comb filtering caused by the summation of signals delayed by various amounts. It is important to know more about this result.

In the first experiment the signal presented to the listener (including the delayed portion) was low-pass filtered at 24 dB per octave at a number of frequencies including those corresponding to the center frequencies to which the resonances had been tuned in Fig. 9.

The results, shown in Fig. 19, exhibit a pattern that is clearly of the same form as that of Fig. 9. The high-frequency limit of the low-pass filtered signal appears to correspond in some fashion to the center frequency of the low-*Q* resonances. Delayed impulsive sounds containing high frequencies were detected earlier and at much lower sound levels than delayed impulsive sounds restricted to lower frequencies. For example, a reflection delayed by 4 ms containing frequencies up to 20 kHz could be heard when its level was about 30 dB lower than a reflection containing frequencies below 200–500 Hz. It appears that only the portion of the

frequency spectrum near the upper limit is responsible for this pattern of response, and that neither bandwidth nor comb filter effects are major factors.

As a final test of this important observation, an experiment was conducted in which the high-frequency limit was held constant, at 20 kHz, and the signal bandwidth was reduced by progressive high-pass filtering. The result, shown in Fig. 20, indicates that there was no effect. The portion of the signal at or near the high-frequency limit is the dominant factor determining the relative audibility of these delayed sounds.

2.7 Audibility of Delayed Broadband Sounds

The effects of reflected sounds on auditory perception is an involved study because the direction from which the reflection arrives is a factor not only in terms of the audibility of the subsidiary sound as it affects timbre but also in the manner it affects stereo imaging. While this study did not address these issues, some of the experiments incorporated delayed broadband sounds, and although the data are quite limited, the results do fit in with the present discussion.

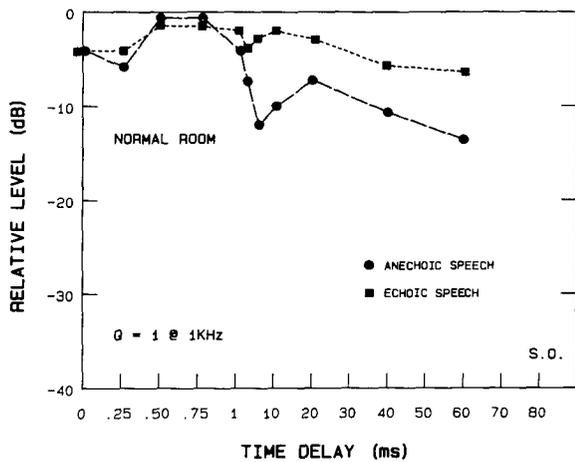


Fig. 17. As Fig. 16, but with signal presented through a loudspeaker at a distance of 2m (6.6 ft) in a normal domestic room.

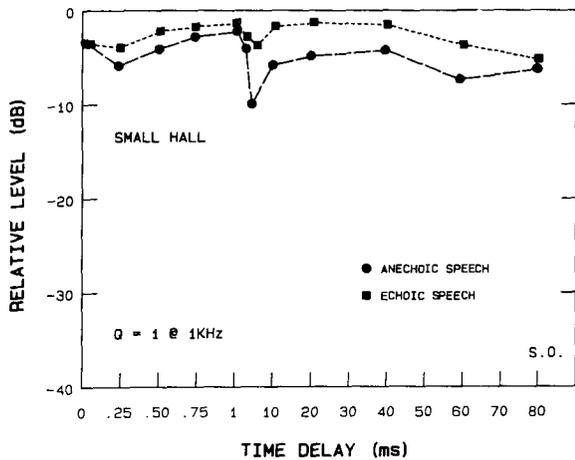


Fig. 18. As Fig. 16, but with signal presented through a loudspeaker at a distance of 8m (26 ft) in a small hall.

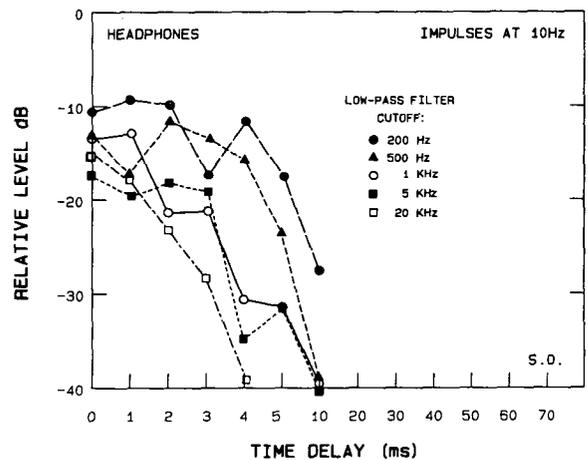


Fig. 19. Threshold of audibility for delayed low-pass filtered impulses auditioned through headphones.

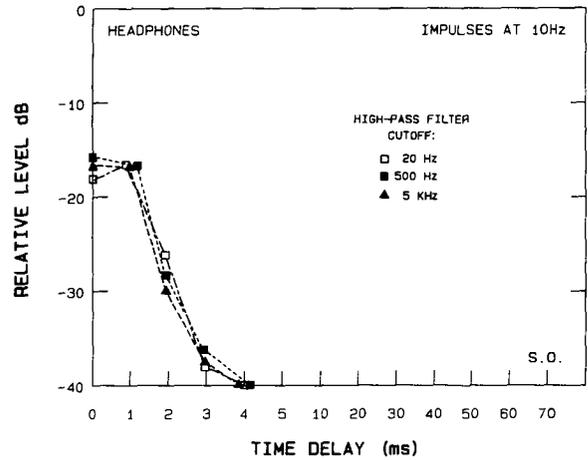


Fig. 20. Threshold of audibility for delayed high-pass filtered impulses auditioned through headphones.

Fig. 21 shows the audibility of various sounds used in the preceding experiments, when they were presented through a single loudspeaker in a typical domestic listening room. In these data it is possible to recognize patterns that have been displayed in the context of detecting delayed resonances. For example, the curve for pink noise resembles those in Figs. 8 and 10. The curves for the other signals have forms that can be recognized in Figs. 13–15. In general the data confirm the observations in the other context, that the sensitivity to a delayed sound depends on the nature of that sound and the extent to which the sound is delayed. For small delays, ranging from zero to about 1 ms, the most revealing sound is pink noise. At larger delays, the delayed sounds are best revealed by signals that are impulsive or at least contain impulsive components, like speech, recorded without reverberation.

2.8 Direct-to-Reverberant Ratio—Stereo versus Mono

In the course of investigations of loudspeaker sound quality, it was found that loudspeakers that rated poorly in monophonic listening tests frequently received higher ratings in stereophonic tests [8]. The question arises whether the difference in ratings was caused by a reduced ability to detect colorations in the stereophonic mode of listening, or whether the spatial enhancements of stereo were, in effect, of perceptually higher priority than coloration.

A related question pertains to the importance of the ratio of the direct to the reverberant sound fields in the vicinity of the listener. In the course of normal events this ratio is affected either by the dispersion of the loudspeaker or by the distance between the loudspeaker and the listener. Loudspeakers with wide dispersion, including those that are deliberately multidirectional, generate a more energetic reverberant sound field in a typical room than more directional devices. Similarly, as the listener moves away from any loudspeaker in a normal room, the reverberant sound becomes a progressively greater proportion of the total sound.

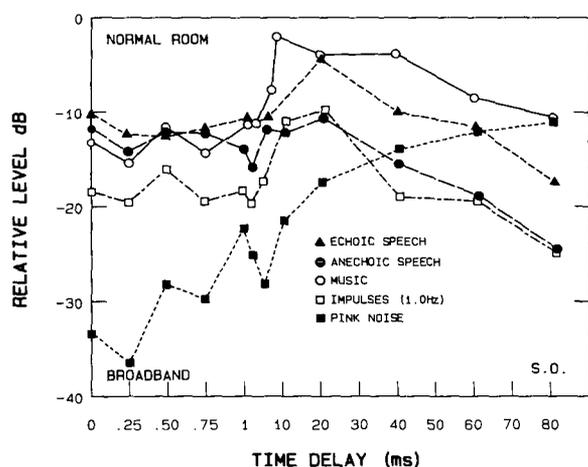


Fig. 21. Threshold of audibility for delayed broadband sounds of various types, auditioned through a single loudspeaker in a normal domestic room.

Both of these issues were addressed in experiments conducted using a pair of composite loudspeakers, as shown in Fig. 22. The forward-facing loudspeakers were fitted with sound-absorbing “blinkers” to isolate a strong direct sound component at the listener’s location. The remaining three loudspeakers were driven in parallel from a separate amplifier, and provided the reverberant component of the total sound heard by the listener. Adjusting the relative drives to the front and rear loudspeakers provided a convenient means of varying the direct-to-reverberant ratio. Using sound pressure levels measured at the listener’s head location as the basis, sound fields with direct-to-reverberant ratios of +6 dB, 0 dB, -6 dB, and -12 dB were randomly presented to the listeners. (+6 dB means that the direct sound is 6 dB higher than the reverberant sound; -12 dB means that it is 12 dB lower than the reverberant sound.)

In addition to varying proportions of direct and reverberant sound, listeners also made threshold adjustments in both monophonic (left or right loudspeaker alone) and stereophonic (both loudspeakers driven simultaneously) listening modes.

The results for stereo listening, using the four ratios of direct and reverberant sound, covering a range of 18 dB, are shown in Fig. 23; again the signal was pink noise. The curves can be seen to weave around each other in an apparently haphazard manner within a range of 5–10 dB. No general trends are apparent in these data.

Fig. 24 shows the threshold shift in a comparison of measurements made in stereo and in mono at the different direct-to-reverberant ratios. Again the variations do not reveal any large or consistent trends.

When listening to pink noise in a typical domestic room, there is nothing in these data to suggest that listening in stereo alters the ability to detect low- Q resonances. Neither, it seems, does substantially chang-

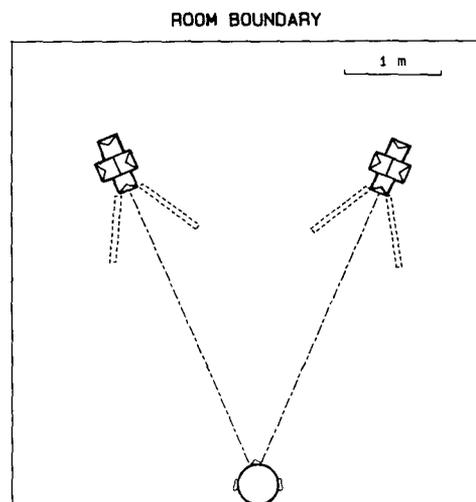


Fig. 22. Arrangement of two composite loudspeakers, each consisting of four small carefully matched high-fidelity loudspeakers. The direct sound from the front pair was isolated somewhat by acoustically translucent baffles.

ing the proportions of direct and reverberant energy in the sound at the listener's location.

It is important to keep in mind that the listening room provided a normally reverberant sound field as a background to this experiment. This condition was appropriate to the specific questions being addressed, as outlined in the introduction to this section. It remains for further study to reveal the manner in which timbral changes are influenced by reflections and reverberation superimposed on a listening environment otherwise free from reflections.

2.9 Effect of Hearing Performance

Six listeners ranging in age from 25 to 62 years were tested. All had essentially normal hearing for their ages. In other words, any measured threshold deviations were accounted for by presbycusis, the seemingly unavoidable deterioration of hearing with age.

These tests used pink noise and were of the basic form described in Sec. 2.1. The results were sufficiently

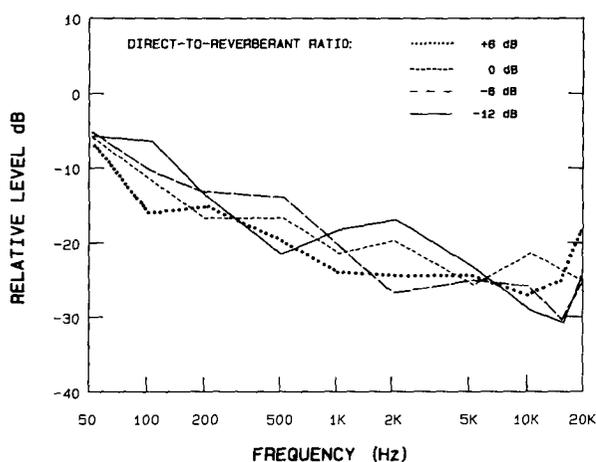


Fig. 23. Detection thresholds for nondelayed resonances with $Q = 1$, as determined using pink noise in a typical domestic room. The results are shown for one listener who auditioned a stereophonic center image. The direct-to-reverberant ratio at the listener's position ranged from +6 dB to -12 dB.

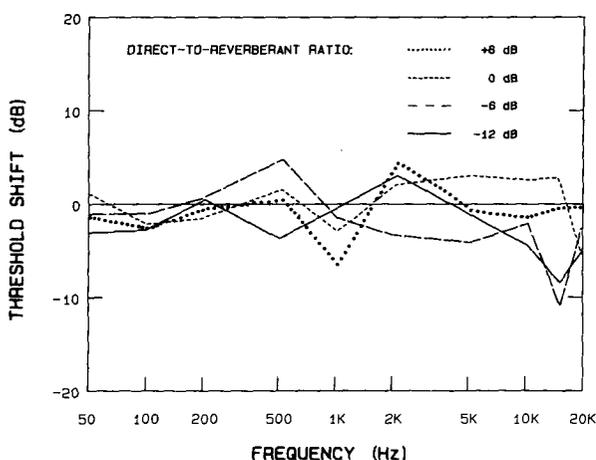


Fig. 24. Difference between detection thresholds for nondelayed resonances measured in stereophonic listening, as shown in Fig. 23, and those measured in monophonic listening (one loudspeaker operating).

similar so that no differences in detection thresholds for resonances could be attributed to either the hearing performance or the age except, possibly, at high frequencies with high- Q resonances. Listeners with noticeably reduced upper frequency limits of hearing when tested with pure tones, were able to detect low- Q resonances at center frequencies up to 20 kHz due, perhaps, to the spectral spread of these phenomena.

To date, no measurements have been made using listeners whose hearing has deteriorated abnormally.

3 COMPARISONS WITH EARLIER WORK

The literature of psychoacoustics contains accounts of numerous studies of simultaneous forward and backward masking, which are the perceptual mechanisms that most likely underlie the effects being examined here. Most of these have been rather academic in their approach, using contrived signals in simple listening situations, and the results are difficult to relate to the present work.

More directly relevant are studies done in the context of concert-hall acoustics, and here there are several instances of directly comparable data. For example, the data in Fig. 8, showing the detection threshold for pink noise as a function of time delay, are similar to those shown in Atal and Schroeder [9, fig. 2], Zurek [10, fig. 2, open circles], and Salmi and Weckström [11, table 1]. The upward shift in thresholds with increasing sound field complexity, observed in this work (e.g., Fig. 13), is paralleled in Zurek in experiments using reversed-polarity signals in headphones [10, fig. 5].

Our examples of the audibility of broadband sounds of different structure, shown in Fig. 21, are consistent with Somerville et al. [12, table 1] data for sounds with 10-ms delay. These workers found that pink noise yielded the lowest threshold, with anechoic speech, choral music, piano, string quartet, and orchestral music showing progressively higher thresholds. The 10-ms data in Fig. 21 agree closely with this observation, but since the measurements extended over a wide range of time delays, it is possible to observe trends not revealed in the earlier study.

The work of Schubert [13] was much more extensive, and the results closely confirm the data of Fig. 21, even including the reduction in thresholds at longer delays for impulsive sounds [13, fig. 7]. Schubert used pizzicato violin as his example of this class of sound. Although the form and trends of the curves were similar, Schubert's thresholds were approximately 10 dB lower than those in this study. The explanation seems to lie in the different angle of incidence for the delayed portion of the combined sound. In this study the direct and delayed sounds arrived along the same axis; in Schubert's study the delayed sound arrived from 30° off the axis of the direct sound. In a separate investigation Schubert found that this angular shift was sufficient to drop the thresholds by approximately the amount of the observed difference. This reference may be difficult

to find, and it is written in German, but there is a brief summary of the results in Kuttruff [14, chap. 7] in the context of a good general discussion of the subjective effects of combined sound fields. Barron, in his study of concert-hall acoustics, confirmed some of Schubert's results in a very similar test [15, fig. 2].

4 PHYSICAL MEASUREMENTS

In practical situations it is important to know the relationship between the perceptual thresholds and physical measurements on the systems. For perspective, Fig. 25 shows the amplitude responses of the resonances used in this study. When added to the signal path, these shapes modify the total system frequency response in a manner depending on the amplitude and time delay of the resonance output compared to the main spectrum.

4.1 Measurement of Resonances Added without Delays

Fig. 26 shows both amplitude and phase responses of the system when a single resonance at 1 kHz has been added at the threshold level as determined with the least revealing program material: popular music data from Fryer [2], [4, fig. 1(d)]. In other words, in an electric signal path that is otherwise perfectly flat, these are the measured amplitude and phase aberrations when the resonance has been adjusted to be at the threshold of detectability when listening to program material. It must be remembered that these irregularities are superimposed on the total responses of the record-reproduction chain, including microphones, rooms, loudspeakers, and so on. Nevertheless, the impression is likely to be that they are surprisingly large. Expressed as a tolerance, in the popular manner, the $Q = 1$ response curve is about ± 1.5 dB, the $Q = 10$ curve is ± 3 dB, and the $Q = 50$ curve is ± 5 dB. Described in this fashion the variations may seem more acceptable, which perhaps explains the popularity of the specification. In performance measurements of audio

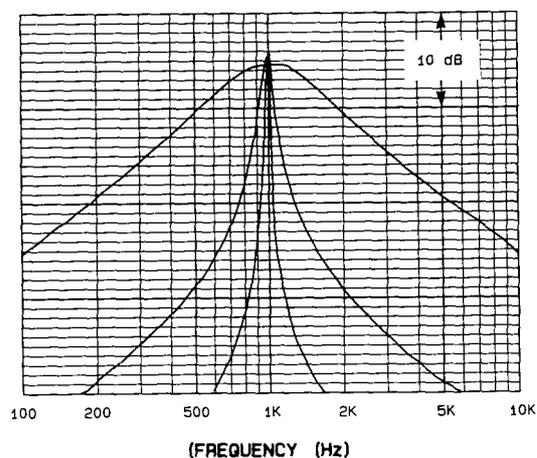


Fig. 25. Frequency responses of three filters used in these experiments. From the widest to the narrowest peaks the Q values are 1, 10, and 50.

components in general such large variations should be easily visible.

With some other sounds, however, resonances are more easily heard and the thresholds are correspondingly lower. Fig. 27 shows amplitude and phase response measurements for an electric signal path with the resonance levels adjusted to correspond to the thresholds determined with nondelayed pink noise (Fig. 3). Measurements are shown for the three values of Q and for three frequencies, 200 Hz, 1 kHz, and 5 kHz.

Fig. 26 shows what might reasonably be interpreted as the maximum measured variations that are permissible under any circumstances, while Fig. 27 shows variations that might be permissible in a stringent high-fidelity design objective. It is evident in Fig. 27 that the removal of visible medium- and high- Q resonances still leaves the designer with the difficult task of identifying the very subtle deviations associated with potentially audible low- Q resonances at any frequency, and medium- Q resonances at high frequencies.

Fig. 28 shows time-domain measurements of selected threshold conditions to see if the resonances are more visible from this measurement perspective. Tone bursts and impulse responses are shown for the conditions described in Fig. 27 at a frequency of 200 Hz—the largest variations in the group. The output from the 32-cycle tone burst is fairly explicit in revealing the high-level high- Q resonance, but the other two resonances are only subtly in evidence. Such stable and prolonged driving signals precisely at the resonance frequency would be fortuitous in music. Impulsive sounds are much more common. Since these would not allow high- Q resonances to build to maximum amplitude, their potential audibility would be much reduced compared to lower Q resonances that are able to respond more rapidly. The stylized impulses used in Fig. 28 illustrate the effect clearly. The amplitude of the ringing

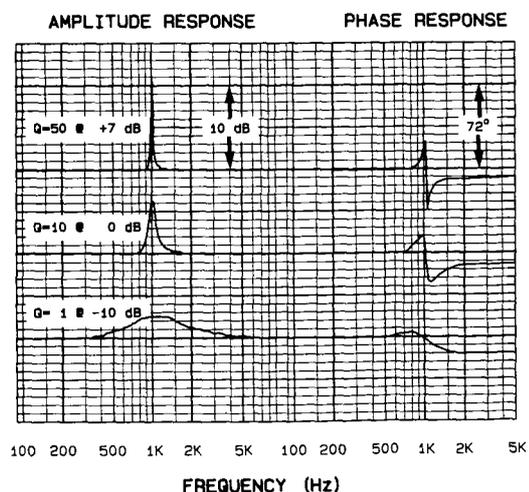


Fig. 26. Amplitude and phase responses of an otherwise linear and flat transmission system to which has been added a single nondelayed resonance at 1 kHz. For each Q the level of the resonance has been set to approximate the threshold conditions when listening to program material that is poor at revealing these phenomena.

at its onset is rather similar in all three impulse responses, while the durations are very different. Since all three of these signals represent conditions at threshold, the implication is that it is the *initial* amplitude, not the duration, of the ringing that is related to the auditory detection process.

4.2 Measurements of Resonances Added with Time Delays

As discussed earlier, the evidence of interference, or comb filtering, in amplitude and phase response measurements will depend on the relative levels of the summed signals and the width of the common frequency band. Small level differences yield large amplitude fluctuations, and vice versa. Increasing the time delay yields closer spacing of the fluctuations—the “teeth” of the comb—in the frequency domain. Fig. 29 shows amplitude response measurements of the system with the parameters set to the threshold conditions at 200

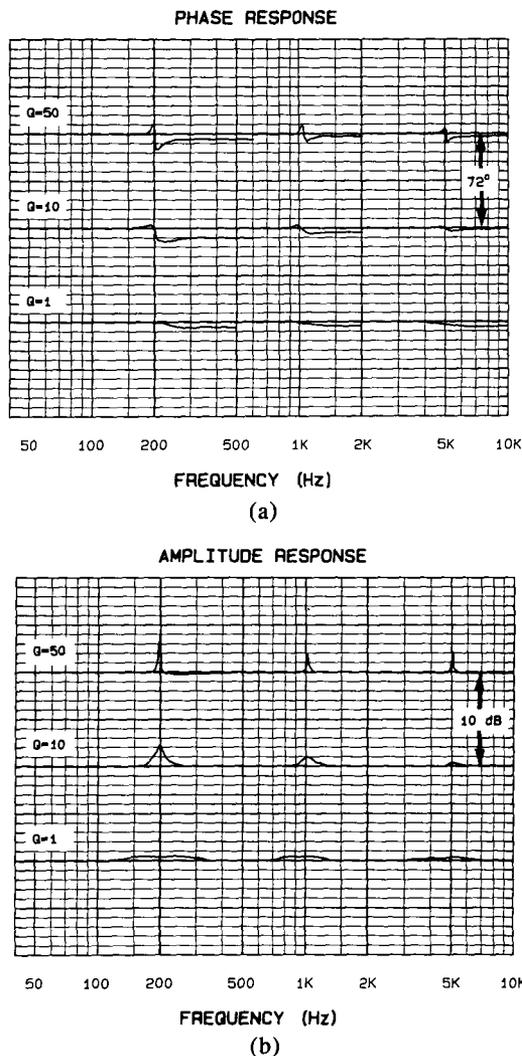


Fig. 27. (a) Amplitude and (b) phase responses of an otherwise linear and flat transmission system to which have been added nondelayed resonances at 200 Hz, 1 kHz, and 5 kHz, having Q values of 50, 10, and 1. The levels of the resonances have been set to approximate the threshold conditions when listening to pink noise—the signal that is best at revealing these phenomena.

Hz in Fig. 9 (impulses, $Q = 1$, headphones). There is nothing in these curves to suggest a simple relationship between any aspect of this measurement and what is heard.

These events were also examined in the time domain through the impulse responses and energy–time curves (the magnitude of the time-domain response) with the system set at a variety of threshold conditions. These data also failed to reveal a simple pattern related to the detection thresholds.

The lack of a straightforward “linear” relationship between measures of system performance with delayed resonances (or delayed signals in general) is not entirely unexpected. Forward and backward masking are strongly nonlinear effects and they are dependent on a number of variables. It is reasonable to presume that these perceptual processes are important, if not the dominant, factors in the audibility of these effects.

5 A HYPOTHESIS

In Sec. 4.1 it was suggested that, in some circumstances, the detection thresholds for resonances may be better characterized by the initial amplitude of ringing than either by the steady-state resonator output or by the duration of the ringing. Let us probe a little deeper into this suggestion.

Fig. 30 shows the data of Fig. 3 in a modified form. The data from all of the studies were averaged, and this combined information was displayed as threshold versus Q for each test frequency. The presentation shows that, when assessed according to the maximum steady-state value of the resonator output, the detection thresholds are dependent on both Q and frequency. While it is possible to work with such data, it requires a nonlinear transformation before the measured parameter can be related to the detectability. A further persistent problem has been that the transformations are not intuitively straightforward (e.g., most people would naturally think that the sustained ringing from high- Q resonances would be the most objectionable).

Following the suggestion in Sec. 4.1, the data of Fig. 30 were altered to show the thresholds in relationship to the peak-to-peak magnitude of the first cycle of oscillation after a 50- μ s impulse, rather than the maximum resonator output when driven by a steady-state signal. This measurement was made at the output of the resonator before it was added to the pulse in the mixer as shown in Fig. 2. (It would have been more difficult to measure accurately after the summation). In Fig. 31 it can be seen that this has had the effect of virtually removing Q as a variable in the relationship. We appear to be one step closer to a plausible mechanism.

Observing that the thresholds for frequencies about a decade apart were separated by approximately 10 dB, it seemed reasonable to incorporate a compensation based on the energy in a fixed number of cycles of oscillation (in this case, by implication, the first cycle). All of the data were normalized to the 200-Hz data

(arbitrarily chosen) by applying a correction factor to each curve calculated, using

$$\text{correction factor (dB)} = 10 \log (\text{frequency}/200)$$

The result, shown in Fig. 32, is that the frequency dependence of the detection threshold is all but removed. Although such a simple demonstration does not prove

cause, there appears to be a relationship between the detection threshold of resonances, as revealed by pink noise, and the energy in the first (or first few) cycles of oscillation of the resonator output. Looking for an explanation, it may be reasonable to argue that, in the continuous pink noise signal, the sustained portions of the ringing are masked. The initial portions are then identified along with the transient features of the signal

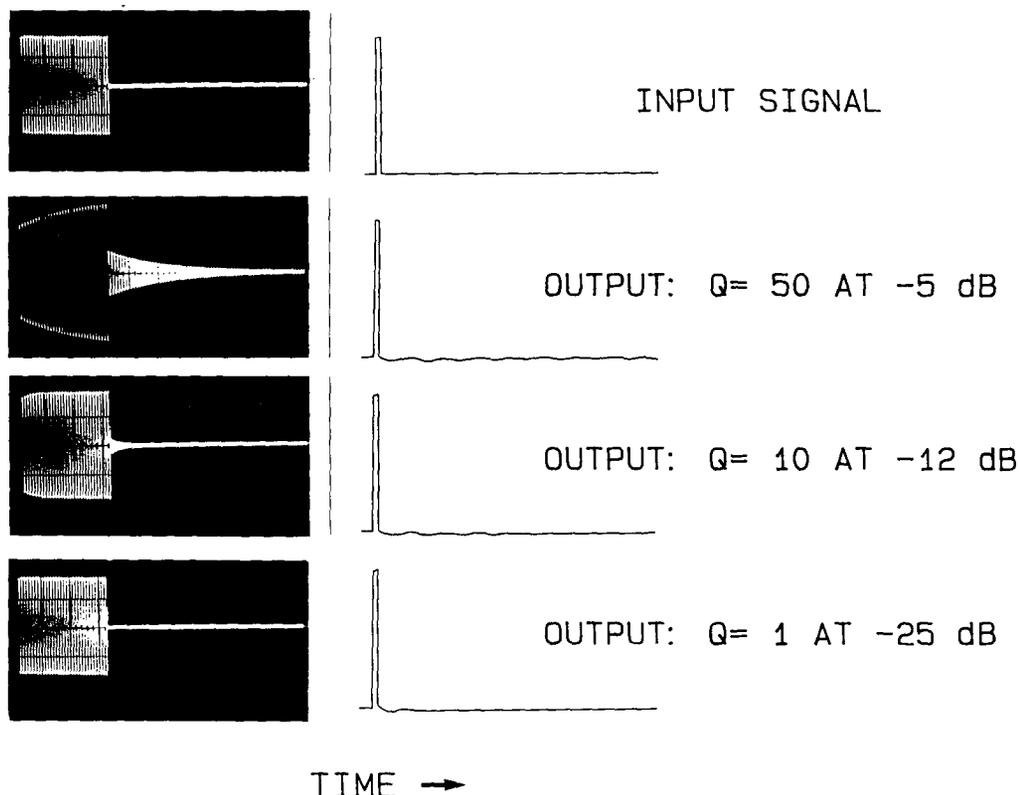


Fig. 28. System response to 32-cycle tone bursts at the resonance frequency and to impulses when conditions are set to threshold at 200 Hz with pink noise as a signal. See Fig. 27 for the corresponding amplitude and phase responses.

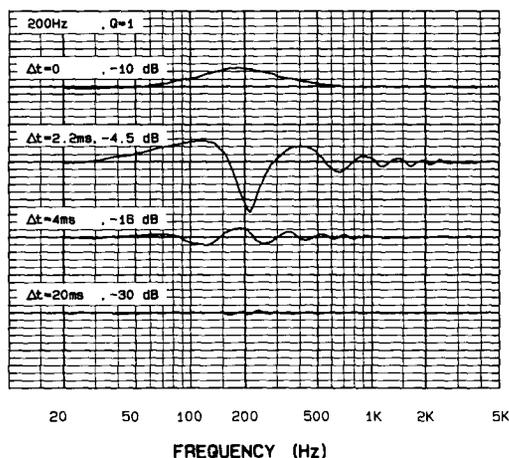


Fig. 29. Amplitude responses of an otherwise linear and flat transmission system to which have been added delayed resonances with $Q = 1$ at 200 Hz. The threshold conditions have been selected from the data of Fig. 9: impulses auditioned through headphones—the most sensitive condition for this signal.

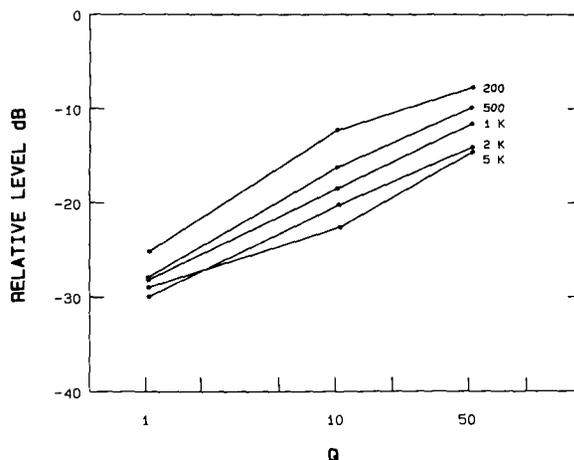


Fig. 30. Detection thresholds for single resonances determined using noise. These data are rearranged from Fig. 3 and represent a mean of all studies. The vertical axis is the relative level of the steady-state resonance output compared to the spectrum level of the signal, as shown in Fig. 1(a). In this presentation the threshold varies with both Q and frequency.

itself in the perceptual selection provided by the precedence effect. Although it is a continuous signal, pink noise can be viewed as a sequence of impulses of random amplitude, occurring randomly in time. Applause is a similar class of sound. Empirical evidence that the precedence effect is effective with such a sound is the ease with which a source of pink noise can be localized in a semireverberant room.

It may be relevant that the most distinctive timbral cues in the sounds of many musical instruments have been found to be in the onsets of transients, not in the harmonic structure or vibrato of sustained portions [16], [17]; “. . . the manner in which the various partials of the tone build up to their final amplitudes . . . is quite important in identifying the instrument; tones recorded without the initial transient are much harder to identify” [18, p. 102]. There may well be a relationship between the perception of timbral changes caused by resonances added to an unstructured continuous sound like pink noise, and the perception of timbre of musical instruments in the unstructured more-or-less continuous reverberation of a concert hall. The experiments in [16] and [17] used recordings of instruments in anechoic chambers, but the authors were not specific about the reproduction circumstances, indicating only that they used loudspeakers.

Nondelayed medium- and low- Q resonances in impulsive or transient sounds (as are commonly found in the sounds of speech and many musical instruments) can be detected in sound fields with numerous reflections and reverberation when they are at levels as much as 10 dB lower than when they are auditioned in anechoic surroundings or through headphones (Figs. 5 and 6). This fact is a clear justification for the attention given to the delicate balance of reflections and reverberation in the design of performing spaces. Too little reflected sound energy and the sound appears dull and lackluster; too much and the musical articulation suffers. It has

long been recognized that the right amount of reflected and reverberated sound adds a pleasant tonal richness to certain musical instrument sounds.

Benade has been particularly emphatic about the importance of early reflections in aspects other than their contribution to the sense of spaciousness [19]. He cites as their advantages 1) the accumulation of loudness (Haas effect), 2) the accumulation of information about onset, spectrum, pitch, space, and time location, and 3) the fact that a “spatially enriched” sound field also tends to communicate to the listener a sound that is more accurately representative of the total sound output from the instruments. Within the early reflected and reverberant sound fields there needs to be a spatial integration of the sounds radiated in different directions from musical instruments [19]. He summarizes this as a “generalized precedence effect” [20, sec. 6.1], in which “there is an *accumulation of information* from the various members of the sequence” (of early reflections). All of this is consistent with observations within the present work.

Vaughan [21] viewed the same process from the opposite perspective. He considered the perceptual consequences of the interaction of a multidirectional sound field with the complex directional properties of the ears and arrived at rather similar conclusions. According to his findings, good rooms for music should present listeners with:

- 1) *Richness*: Strong multiple reflections from all angles shortly after the original sound.
- 2) *Density*: A large number of reflections in a short period of time, growing progressively denser.

His list goes on, with the consistent theme that the right temporal and spatial complexity in sound fields can be beneficial to the appreciation of musical sound. To the extent that the detection of timbral subtleties is a factor in this, the present work has added a quantitative argument to this same end.

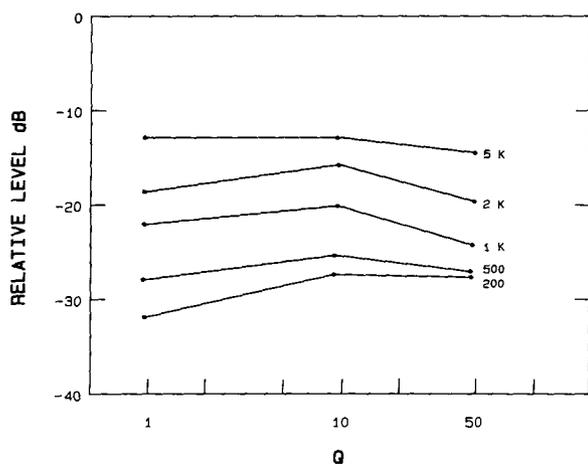


Fig. 31. As Fig. 30, but with the vertical axis modified to show the relative level of the peak-to-peak oscillation of resonator output in response to a 50- μ s impulse. In this presentation the threshold varies mainly with frequency, which is in reverse vertical order from Fig. 30.

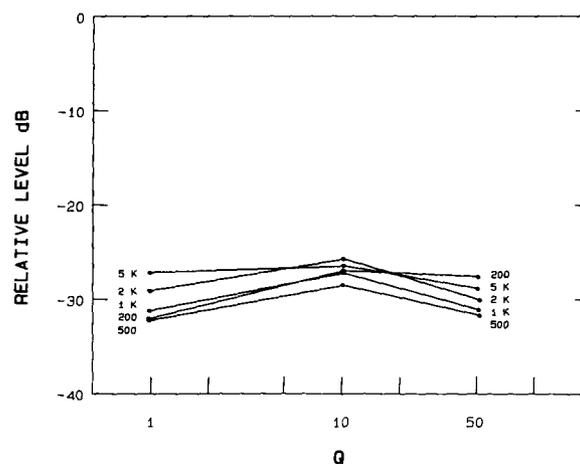


Fig. 32. As Fig. 30, but with the vertical axis modified to show the sound energy in the first n cycles of oscillation of the output of the resonator in response to a 50- μ s impulse. In this presentation the threshold is approximately independent of both Q and frequency.

6 SUMMARY AND DISCUSSION

The audibility of resonances is a major factor in subjective assessments of instrumental timbre, both in live performance and in reproduction through loudspeakers or headphones. The variables in these situations are numerous, especially in the case of sound reproduction, where the entire sound recording-reproduction chain is being scrutinized.

The audibility of timbral modifications depends on the Q , the frequency, and the amplitude of the resonance. This much is well known. However the relationship to the measured evidence of the resonance is not always intuitively straightforward. Changes in perceived sound quality due to added resonances also depend on other factors, such as the kind of program material used in the audition, reverberation in the reproduction environment, and reverberation (real or synthesized) in the recording. It is therefore to be expected that, other factors being equal, listening tests can sometimes yield different opinions about devices or recordings being evaluated.

With many common sounds, under common listening circumstances, it is surprising just how much the measured performance of a signal path can be modified without significantly altering perceived timbre. On the other hand, given the right signal and listening circumstances, it is possible to hear resonances that would be very difficult to detect by *any* technical measurements.

It is important to note that this study has concentrated on defining the detection thresholds of resonances, the level at which any change in sound quality is just, or just not, noticeable. While a purist criterion for sound reproduction would require that all resonances (and delayed sounds) should be below these threshold levels, it should not automatically be assumed that audible resonances or delayed sounds are detrimental. Certain modifications of timbre, loudness, and spatial impression can be quite clearly annoying, while others may not. Some may even be preferred. The results of this study define the levels at which certain changes in system performance can just be perceived, and these preferential options set in.

The measurements most commonly used to reveal the presence of resonances are those that are based on the maximum steady-state level as revealed in the amplitude response or phase response of the system. Measurements in the time domain fall into two classes, those that assess the response of the system to tone bursts, allowing the system to stabilize at full steady-state output before termination of the burst, and the impulse response itself or its magnitude (energy-time curve). The well-known waterfall diagrams, or cumulative spectra, can be viewed as a composite of tone-burst responses at many frequencies simultaneously decaying from the steady-state level (though they are usually measured using transient signals in FFT [22] or time-delay spectrometry techniques). In the following summary, the measured parameter is taken to be the maximum steady-state level.

Resonances are added to the signal with essentially no time delay when the resonating elements are tightly coupled to the source of energy. In sound production these would be the multitude of resonances that comprise the timbre of vocal and musical instrument sounds. In sound recording and reproduction there are several opportunities for such resonances. For example, they can occur in the diaphragm or suspension system of a loudspeaker, in air resonances in the enclosure, as well as in mechanically driven panel resonances. Mechanical or electrical resonances in microphones and recording-reproduction devices (tonearms, cartridges, analog and digital tape or disk devices) would also generally be of this kind. Changes in sound spectra imposed by equalizers also fall into this category, most obviously in the case of parametric equalizers where the signal manipulations are described explicitly in terms of the frequency and Q of filters. These, clearly, are timbral changes of great importance in the fields of music production and reproduction.

The audibility of resonances without time delay, based on steady-state measurements, can be summarized as follows:

- 1) Low- Q resonances, producing broad peaks in the measurements, are more easily heard than high- Q resonances producing narrow peaks of similar amplitude.
- 2) The detectability of resonances decreases approximately 3 dB for each doubling of the Q value.
- 3) In general, pink or white noise are the most sensitive indicators of these resonances, with speech and music progressively less sensitive. Continuous signals with dense broadband spectra seem to be advantageous.
- 4) With discontinuous, impulsive, or transient sounds (which occur as components of some speech and musical instrument sounds) the addition of signal repetitions in the form of reflections and reverberation during recording or reproduction can increase the audibility of medium- and low- Q resonances (the improvement can be as much as 10–14 dB), but they will have little effect on resonances of high Q ($Q \gg 10$). This means that some timbral subtleties (both virtues and faults) will be better revealed when listening in typical semi-reverberant rooms than when listening through headphones or in acoustically dead environments.

It should be noted that, while reverberation assists the perception of timbre in transient sounds, it can be detrimental to the timbral identification of sustained sounds, such as vowels and organ tones [23], [24].

Both findings imply that, in the making of recordings, the monitoring environment should have some important acoustical similarities to the intended listening environment. There is the further implication that the necessary similarities, in this respect at least, might be satisfactorily achieved electronically, in the form of synthesized reverberation.

- 5) The duration of ringing is itself an unreliable indicator of the audibility of these resonances.

Resonances added to a signal after a time delay may be perceived in different ways, depending on a number of factors, probably because of the complexities of

temporal (forward and backward) masking. Delayed resonances occur when there is a propagation delay between the direct sound and the resonating element. They also occur if a simultaneously resonating component, radiated in a different direction, reaches the listener through a reflecting pathway. Musical instruments and loudspeakers both can radiate dissimilar sounds in many directions. The interactions of these off-axis sounds with room boundaries and other objects in rooms create many opportunities for delayed resonances in sound production and reproduction.

6) The audibility of single delayed resonances is summarized in the following. The detection of delayed sounds appears to be dominated by two quite distinctive patterns of perception that depend on the temporal and spectral structure of the sound.

a) With continuous broadband sounds, such as pink or white noise, resonances are most readily heard at zero or very small time delays. The threshold rises rapidly for delays up to about 1 ms, and it continues to rise slowly for longer delays. In headphone listening the effect is substantial; in delays ranging from zero to 60 ms the threshold can be elevated by as much as 25 dB.

b) With transient sounds, including nonreverberant speech, resonances are least audible at delays less than about 1 ms, and the threshold drops at longer delays. The rate of drop increases with frequency. At a given delay in excess of about 1 ms, sounds containing the highest frequencies will be the most audible. In headphone listening the effect is substantial; in delays ranging from zero to 20 ms the threshold can be lowered by as much as 40 dB.

7) Changing the temporal characteristics of the sound modifies the patterns of threshold versus delay that were described above. This occurs when, for example, reverberation is added to the signal during the recording process, or when reflections and reverberation are added to the signal by the listening environment. In both cases the *differences* between the patterns noted in item 6) are reduced. All of the thresholds tend to be elevated toward the maximum (least sensitive) levels.

The effect is quite strong. The substantial differences noted in headphone listening [see item 6)] were much reduced by the effects of listening in the reflections and reverberation in a typical domestic listening room. Listening in a small hall removed virtually all of the differences; the effect of time delay was substantially diminished.

Adding reverberation to the recorded signal had a similar effect. Speech recorded with and without reverberation yielded very different detection patterns when auditioned through headphones (and, by inference, in an acoustically damped room). The delayed resonances were more audible in the anechoic speech. In contrast, the speech recorded with reverberation yielded thresholds that were high and very similar to those found when reverberation was increased in the listening environment. The point at which reverberation is added to a sound seems not to be a factor; the result is con-

sistently to reduce the audibility of delayed resonances and to diminish the effect of time delay as a variable.

8) In item 6b) it was noted that the most easily detected delayed transient sounds were those containing the highest frequencies. Further investigations indicated that this effect was not strongly related to the bandwidth or to interference (comb filtering). Instead, it appears that the sound energy near the high-frequency limit of the signal spectrum has special perceptual significance. It may therefore be inferred that the audibility of reflections can be reduced by attenuating the high frequencies either at the source or by sound absorption at the reflecting surface. In this respect, it could be fortunate that most loudspeakers become more directional at high frequencies, effectively low-pass filtering the off-axis sounds reflected by the room boundaries.

The audibility of broadband delayed sounds was not examined in depth in this study, but there were clear indications of the same patterns discussed in the preceding summary. At time delays of less than 10–20 ms, delayed broadband sounds were most readily revealed by continuous pink noise; at longer time delays impulsive signals were most revealing. At any time delay, speech and music with reverberation were the least revealing sounds.

Viewed in total, the indications are that reflections and reverberation improve the ability to hear small timbral features in impulsive and transient sounds due to nondelayed medium- and low- Q resonances. In music production this might help explain why listeners clearly prefer the sound of instruments played in rooms having a certain amount of reverberation. It may also help to explain why listeners prefer loudspeakers with the smoothest and flattest frequency responses, those with the fewest resonances [1]. The fact that loudspeakers with similarly good performance both on and off axis are preferred to those exhibiting irregular response away from the principal axis is especially important. These off-axis sounds account for much of the reflected sound field in the listening room. By providing the listener with a temporal sequence of near replicas of the direct sound, they appear to lead to a further perceptual enhancement of timbral subtleties, whether they are in the program material (good) or added by the loudspeaker (bad).

Timbral changes also arise from spectral manipulations using equalizers in recording or mastering in control rooms, or in the listening room. The audible effects will depend on the amount of reverberation in the program material and in the listening environment. An amount of equalization appropriate for one listening condition may not be equally appreciated in another. It is one more source of variability in the trouble-prone record-reproduction cycle.

When a single delayed sound is very different in character from the direct sound, there is a situation corresponding to a delayed resonance. In this case, reverberation reduces the audibility of the delayed sound, a fortunate coincidence since these sounds would likely be detrimental to the overall effect.

While it is tempting to speculate further about the relationship between these findings and the situation of loudspeakers in rooms, it is probably advisable to wait until the completion of further work on the subject.

The high sensitivity to resonances under certain circumstances creates a serious problem for those attempting to identify all potential resonant colorations in measurements. The simple fact is that there are situations in which it may not be possible to do so. Fortunately these situations of highest sensitivity tend not to coincide with either the typical musical sounds on recordings or the typical listening circumstances. Consequently, for most practical purposes, it is possible to detect the presence of audible resonances using conventional measurements.

With electronic signal paths the interpretation of measurements is straightforward, since there are virtually no visual distractions. In the case of loudspeakers, however, a major problem is to separate the fluctuations in measured response that are caused by interference from those caused by resonances. The acoustical interference of sounds originating at different locations on a loudspeaker, such as from different drivers or cabinet diffraction, causes irregularities in measurements that change as a function of microphone position. An energy average of measurements made at a number of microphone positions (spatial averaging) can be used to reduce the visibility of interference effects while retaining the frequency resolution necessary to reveal the presence of resonances. This process has been described, with numerous examples, in [1]. It is rewarding to note that the residual frequency-response fluctuations in the most preferred loudspeakers in that reference [1, fig. 8] were more closely comparable in scale with those seen here in Fig. 27, rather than those in Fig. 26. Apparently, integrated over the duration and musical variety of a series of those listening tests (typically four to six 30-min sessions), listeners were able to detect and to establish clear preferences based on very small changes in timbre caused by resonances.

With any measurement it is clearly important that there be adequate frequency resolution to reveal the presence of high- Q resonances. Poor resolution in the frequency domain will have the effect of broadening and lowering bumps associated with high- Q resonances, an effect that in itself is tolerable only when the modified evidence is sufficient to attract the appropriate attention. In some of the situations depicted in Fig. 27 there are few enough data to work with even with adequately high resolution.

The popular one-third-octave measurements are useful only to reveal gross features in the frequency domain, especially when the filter center frequencies are fixed rather than swept. Less obvious, but much more common at the present time, are the frequency resolution limitations imposed by any of the simulated free-field measurement methods (such as FFT, TDS, tone burst) wherein there is time-domain gating, and the effective frequency resolution is the inverse of the gating interval. Presented on a logarithmic frequency scale, this has

the effect of presenting the viewer with more resolution than may be necessary at high frequencies and not enough at low frequencies. In many practical measuring situations evidence of the $Q = 50$ resonances in Fig. 27 would all but disappear at frequencies below 1–2 kHz.

Earlier work on loudspeaker sound quality evaluations revealed that loudspeakers that rated poorly in monophonic listening tests frequently received higher "fidelity" ratings in stereo listening. It now appears that this change is indicative of the high perceptual priority of spatial presentation rather than a reduced ability to detect resonances in a spatially enriched sound field. It is reassuring to note that the loudspeakers with the highest ratings in either mono or stereo listening were those with the least evidence of resonances [8].

On a more familiar level, it is now possible to explain why two loudspeakers never seem to sound *exactly* the same when auditioned using pink or white noise, even though the measurements may be virtually identical. In most cases the measurements simply lack the resolution to reveal all of the audible resonances. Even in the best of cases some resonances that are audible will be difficult to identify in measurements, especially if, in addition, there are the effects of acoustical interference to confuse matters. It is also possible to explain why those audible differences that seem so conspicuous when listening to noise, diminish when listening to music.

Finally, it may be possible to explain why listeners are able to hear relatively small changes in the bass response of loudspeakers through the multitude of very energetic resonances that populate every listening room. The behavior of loudspeakers at low frequencies is described by low- Q filters, while the room resonances, although mountainous in scale, are relatively high- Q phenomena.

While shedding light on a number of issues, the present experiments have also raised some new questions. More experiments are in progress.

7 ACKNOWLEDGMENT

The authors wish to acknowledge the financial assistance of this work by Audio Products International, Paradigm Electronics, Inc., and PSB Speakers, and the donation of equipment by these companies and by KEF Electronics Ltd.

8 REFERENCES

- [1] F. E. Toole, "Loudspeaker Measurements and Their Relationship to Listener Preferences," *J. Audio Eng. Soc.*, vol. 34, pt. 1, pp. 227–235 (1986 Apr.); pt. 2; pp. 323–348 (1986 May).
- [2] P. A. Fryer, "Loudspeaker Distortions, Can We Hear Them?" *Hi-Fi News Rec. Rev.*, vol. 22, pp. 51–56 (1977 June).
- [3] R. Bücklein, "The Audibility of Frequency Response Irregularities" (1962), reprinted in English in

J. Audio Eng. Soc., vol. 29, pp. 126–131 (1981 Mar.).

[4] P. A. Fryer, "Intermodulation Distortion Listening Tests," presented at the 50th Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 23, p. 402 (1975 June).

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

[6] K. Moulana, "Tonal Colouration Caused by a Single Resonance," Ph.D. thesis, University of Surrey, UK (1975 Aug.).

[7] "Sound System Equipment, Part 13, Listening Tests on Loudspeakers," International Electrotechnical Commission, Publ. 268-13 (1986).

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

[9] B. S. Atal and M. R. Schroeder, "Perception of Coloration in Filtered Gaussian Noise—Short-Term Spectral Analysis by the Ear," *Proc. 4th Int. Cong. on Acoustics* (Copenhagen, 1962 Aug., paper H31).

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

[11] J. Salmi and A. Weckström "Listening-Room Influence on Loudspeaker Sound Quality and Ways of Minimizing It," presented at the 71st Convention of the Audio Engineering Society, *J. Audio Eng. Soc. (Abstracts)*, vol. 30, p. 360 (1982 May), preprint 1871.

[12] T. Somerville, C. L. S. Gilford, N. F. Spring, and R. M. D. Negus, "Recent Work on the Effects of Reflectors in Concert Halls and Music Studios," *J. Sound Vib.*, vol. 3, pp. 127–134 (1966).

[13] P. Schubert, "Die Wahrnehmbarkeit von Rückwürfen bei Musik" (The Perceptibility of Reflections in Music), *Z. Hochfrequenztech. Elektroakust.*, vol. 78, pp. 230–245 (1969).

[14] H. Kuttruff, *Room Acoustics*, 2nd ed. (Applied Science Pub, London, 1979).

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

[16] M. Clark, Jr., D. Luce, R. Abrams, H. Schlossberg, and J. Rome, "Preliminary Experiments on the Aural Significance of Parts of Tones of Orchestral Instruments and on Choral Tones," *J. Audio Eng. Soc.*, vol. 11, pp. 45–54 (1963 Jan.).

[17] E. L. Saldanha and J. F. Corso, "Timbral Cues and the Identification of Musical Instruments," *J. Acoust. Soc. Am.*, vol. 36, pp. 2021–2026 (1964).

[18] J. Backus, *The Acoustical Foundations of Music* (Norton, New York, 1969).

[19] A. H. Benade, "Critique of Certain Concert-Hall Design Criteria," presented at *112th Meeting of the Acoustical Society of America* (1986 Dec.), paper A-5.

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

[21] D. Vaughan, "The Perception and Appreciation of Richness and Timbre in Rooms for Music," *Appl. Acoust.*, vol. 15, pp. 289–306 (1982).

[22] 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).

[23] R. Plomp and H. J. M. Steeneken, "Place Dependence of Timbre in Reverberant Sound Fields," *Acustica*, vol. 28, pp. 50–59 (1973).

[24] R. Plomp and G. Langhout, "Effect of Reverberation on the Identification of Stationary Vowels," *J. Acoust. Soc. Am.*, vol. 77, pp. 1261–1262 (1985 Mar.).

THE AUTHORS



F. E. Toole

Floyd E. Toole was born in Moncton, New Brunswick, in 1938. He received a B.Sc. degree in electrical engineering in 1960 from the University of New Brunswick and a Ph.D. and D.I.C. in electrical engineering in 1965 from the Imperial College of Science and Technology, University of London, England. Since



S. Olive

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 multipurpose concert halls and theaters.

Dr. Toole is a member of the Audio Engineering Society, the Acoustical Society of America, and the Canadian Acoustical Association. In 1986 he was awarded a fellowship in the Audio Engineering Society.

Sean Olive was born in Brockville, Ontario, Canada, in 1959. He received a bachelor's degree in music in 1982 from the University of Toronto and continued graduate studies in the Tonmeister program at McGill University in Montreal, receiving a master's degree in music (sound recording) in 1986.

In addition to teaching sound recording at McGill University, Mr. Olive has engineered and produced numerous recordings of live concert performances and studio sessions in Montreal.

Since 1985 May, he has been a guest researcher at the National Research Council, in Ottawa, where he has been conducting research in psychoacoustics related to the perception and measurement of loudspeaker and listening-room performance. He is frequently involved in the measurement and evaluation of audio products, and his efforts in research have been supported through research grants provided by such Canadian loudspeaker companies as Paradigm, PSB, and, currently, Audio Products International. He is a member of the Audio Engineering Society.