

Auditory demonstrations on compact disk for large N

William Morris Hartmann

Department of Physics, Michigan State University, East Lansing, Michigan 48824

(Received 1 April 1992; accepted for publication 23 September 1992)

The popular compact disk entitled *Auditory Demonstrations*, sponsored by the Acoustical Society of America, includes a number of demonstrations that lead to quantitative results. Those demonstrations are evaluated here in the context of a sizeable class in a lecture room. Demonstrations concern masking, loudness, and pitch; specifically they are numbers 2, 3, 6, 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 20, and 25. The evaluations find that most of the demonstrations successfully illustrate psychoacoustical principles in a classroom context; others are less successful or require special circumstances for success. Explanations for success and failure are offered, together with some suggestions for optimizing the chances of success.

PACS numbers: 43.10.Ln, 43.10.Sv, 43.66.Lj, 01.50.My

INTRODUCTION

The compact disc entitled *Auditory Demonstrations*, created by Houtsma *et al.* (1987) at the Institute for Perception Research (IPO) and sponsored by the Acoustical Society of America, is widely used throughout the world in teaching and learning the principles of psychoacoustics. It has been reviewed briefly by Chivers (1989), Eargle (1989), Hartmann (1989), Strong (1989), and Carterette and Kendall (1990). The disc is comprised of 39 demonstrations, illustrating diverse topics organized into seven sections: frequency analysis and critical bands, sound pressure and loudness, masking, pitch, timbre, nonlinearities, and binaural effects. Some of the demonstrations are musical; most of them are not, although the effects illustrated have musical implications.

The demonstrations are of two kinds: There are those in which the object of the demonstration is simply the listening experience. For example, the presentation of noise bands with different bandwidths and different center frequencies has little purpose beyond the elemental sonic experience. Other demonstrations are more quantitative. In these demonstrations the point is made by the numerical values of the data, which come from responses by the listeners. As will be seen below, there are some demonstrations in which the pedagogical message is not apparent at all until the data are analyzed.¹

The purpose of the present article is to test the quantitative demonstrations of the compact disc, those which lead to data. The data are analyzed with the primary goal of determining whether each demonstration succeeds or fails in making the point about auditory perception that it intends to make. Upon occasion, the analysis of the data seems to lead to some new insights about perception itself.

I. CONTEXT

The testing crucible for the demonstrations was the Science of Sound course in the Physics Department at Michigan State University, begun in 1974 and taught each

year since then. In matters of substance, the course owes a debt to pioneers in musical acoustics education: Dayton Miller, John Backus, and Arthur Benade, to name a few. In matters of style, the course draws its inspiration from P. T. Barnum. The course lectures make full use of the wide variety of sonic demonstrations, musical and psychoacoustical, that are the special asset of the discipline of musical acoustics. Students hear sounds through a high quality audio system in the lecture hall; they view the analysis of sounds by electronic instrumentation via a dozen closed-circuit television monitors than hang from the ceiling of the hall.

The course is intended to capture the attention of the nonscience student. In the early years, students who were not enrolled in the course came to class anyway for an afternoon's entertainment. In more recent years, the ratings among the casual observers have dropped due to heavy competition from the television soap operas. The course remains popular enough, however, to enroll a significant number of students from the full spectrum of academic majors. The relatively large population of the course enabled tests of the compact disc demonstrations with a heterogeneous group of young listeners.

Demonstrations from the compact disc were presented to Science of Sound students in the winters of 1990, 1991, and 1992, three disjoint groups. The demonstrations were incorporated into ordinary lectures in order to illustrate psychoacoustical facts or principles.

To clear up ambiguities and to obtain additional data where needed, a special group of 33 volunteers was assembled from the class of 1992 for a one-hour listening session, held on 5 March 1992. These listeners were paid \$5 for their participation. Below, this subgroup of the 1992 class will be called group 92W. The average age of group 92W was 19.9 years (s.d. = 1.8). In response to the question, "For how many years have you regularly practiced a musical instrument or singing," the average was 5.3 years (s.d. = 5.5). It seems likely that these statistics are representative of all the listening groups.

All demonstrations were presented in a fan-shaped lec-

ture hall seating 300. For any given demonstration, the actual audience ranged between 30 and 150. In this room, empty seats have seat cushions and hard backs. The volume of this room is about 1000 cubic meters. The floor is covered with industrial carpet, and there is acoustical tile on the ceiling. Walls are concrete block with 20 square meters of absorbing panel distributed on side and rear walls. The reverberation time, measured at 500–1500 Hz, was 0.4 s.

The demonstrations were reproduced via four large Advent loudspeakers (*circa* 1974) that hang (upside down) in pairs from the ceiling at the left and right front corners of the room. Other details of the presentation are described with individual demonstrations below. The numbering of the demonstrations follows the numbering on the disk.

II. EXPERIMENTS

A. Demonstration 2: Critical bands by masking

For many purposes it is useful to think of the auditory system as a bank of tuned filters. Each filter constitutes a channel called a critical band. Frequency components of a signal that lie in the same critical band interact. Components in different critical bands are processed by different neural circuits and do not interact, though they are often combined by the central auditory system to create a fused entity (Hartmann, 1988). Critical-band filtering is central to a wide variety of psychoacoustical effects, and there are a correspondingly large number of ways to measure the width of a critical band. Demonstration 2 measures the critical bandwidth by a masking experiment using masking noise with different bandwidths.

Masking here simply refers to the fact that it is harder to hear a signal when it is embedded in noise than when it is in quiet. Quantitatively, masking is studied by measuring the signal threshold intensity, i.e., the level of the signal that is barely detectable. The amount of masking is equal to the threshold in noise minus the threshold in quiet, expressed in dB.

1. The disk demonstration

The critical band plays a role in the masking of a sine tone by noise because only the noise energy within a critical bandwidth of the tone contributes to the masking. Demonstration 2 illustrates this fact by masking a 2000-Hz tone with noise bands, centered on 2000 Hz and having four different bandwidths: 10 000, 1000, 250, and 10 Hz. The spectral density of each noise band is the same. Therefore, the four different bands have greatly different power, and greatly different loudness. None-the-less, so long as the noiseband is broader than a critical band the amount of masking of the 2000-Hz tone is expected to be the same. Because the critical bandwidth at 2000 Hz is about 280 Hz (Zwicker, 1961), only the 10-Hz band should give an amount of masking that is appreciably different from the broadband noise.

The disc demonstration measures threshold with a descending staircase method. The 2000-Hz sine tone is presented as a sequence of ten pulses. Each pulse is 5 dB less intense than the previous pulse. The listener's task is to count the number of pulses that he can hear. The smaller the number

of audible pulses, the higher the signal threshold and the more the masking. The masking noise is continuous. For each masker bandwidth the staircase is presented twice.

2. Presentation

The demonstration was presented at a high enough level that all listeners could hear all ten pulses of the staircase when the tone was in quiet, though the tenth and quietest pulse was barely audible for those in the back of the room. There were 99 listeners who heard the demonstration and filled out a ballot form with a single response, based upon their best estimate of the threshold pulse number, given both presentations of the staircase. While each staircase descended, the experimenter indicated the step number by pointing to numbers 1 through 10 written on the blackboard at the front of the room.

3. Results

The average thresholds for the 99 listeners are given in Table I, both in step number and as converted to dB. The data show that the demonstration made its point effectively. Thresholds were the same, well within the standard deviation of 4 dB, for broadband noise and for noise with bandwidths of 1000 and 250 Hz. The average threshold dropped by 15 dB for the bandwidth of 10 Hz.

Physical measurements on the signals from the disc showed that the average thresholds in the table are in reasonable agreement with commonly accepted values. The noise spectrum level was measured to be 48 dB below the level of the most intense signal pulse [Note: noise spectrum level is obtained from noise level by subtracting $10 \log$ (bandwidth in Hz).] According to standard masking tables (Scharf, 1970), the threshold for a 2000-Hz signal should be 20 dB higher than the noise spectrum level, hence it should be -28 dB. The average step from the table, 5.6, corresponds to a level of -23 dB. The discrepancy of 5 dB can be attributed to the suboptimal conditions of the environment for the present experiment.²

The success of this demonstration is due to the fact that masking levels do not depend much upon absolute level. Lis-

TABLE I. Demonstration 2, critical bands by masking: Detection thresholds for a 2000-Hz sine tone masked by noise with different bandwidths. The column "mean step" gives thresholds, averaged over 99 listeners, in units of staircase step numbers. The column "Watkins" is data from a similar experiment (12 listeners) by Anthony Watkins. Column "Threshold" translates step numbers to level in decibels, referenced to the most intense 2000-Hz pulse. Standard deviations are in parentheses. In the quiet condition all 10 steps were heard, with the tenth 45 dB below the first.

Condition	Mean step	Watkins	Threshold (dB)
Quiet	10 (0)		-45 (0)
Noise bandwidth			
10 000	5.8 (0.8)	5.8 (0.09)	-24 (4)
1 000 Hz	5.6 (0.7)	5.3 (0.7)	-23 (4)
250 Hz	5.5 (0.8)	5.5 (0.7)	-22 (4)
10 Hz	8.4 (0.7)	8.0 (1.0)	-37 (4)

teners in the back of the room, where the signal and noise were relatively weaker, found amounts of masking that were almost the same as listeners near the loudspeakers in the front. Further, broadband irregularities in the transfer function of the room affected signal and noise alike. Excellent agreement with classroom data from Watkins (1992) shows that similar results can be obtained in different environments.

4. Impression

Although the demonstration is a quantitative success, the impression that it leaves on the listener is somewhat puzzling. Energy detection theory and the critical band concept provide a unified model, intended to cope equally well with broadband and narrow-band noise maskers. However, broadband and narrow-band situations sound different. When a listener detects a signal in broadband noise, a tone with clear pitch stands out from the noise background. By contrast, when the noise is narrow band, the noise has almost as much pitch as the tonal signal. Detection then seems to depend upon the overall loudness or upon a sense of envelope stability. (Without a signal, the envelope fluctuates more.) In sum, it is possible that energy detection theory is the correct approach to masking by both wide and narrow bands, but to the casual listener it does not sound that way.

B. Demonstration 3: Critical bands by loudness comparison

A second way to measure critical bandwidth is by the loudness of a band of noise. Loudness is a psychological quantity, not a physical quantity. If the bandwidth of a noise increases while the spectral density of the noise decreases proportionately, the total power remains constant. For bandwidths smaller than a critical band, the loudness remains constant as well. However, if the bandwidth is larger than a critical band, the loudness of the noise increases for increasing bandwidth (Zwicker *et al.*, 1957).

The explanation for this growth of loudness is that when the bandwidth is narrow, it is mostly the same tuned neurons that respond as the bandwidth increases slightly. As the bandwidth increases beyond a critical band, the increase enlists different neurons. The addition of new neurons to the chorus appears as increased loudness.

1. The disc demonstration

A reference noise band, centered on 1000 Hz with a 15% bandwidth (145 Hz), is followed by a test band. The test band is initially the same as the reference. Then the width of the test band increases in seven steps, each of 15%, while the spectral density is decreased to keep the power constant. The listener is asked to compare the loudness of the test band with the reference on each of the eight steps.

2. Presentation

Listeners had a paper ballot with blanks numbered 1 through 8. Their task was to estimate the loudness of the test band, given that the reference band was assigned a loudness rating of 100. The demonstration is brief, lasting only 69

seconds. It was presented twice, first as an introduction, then for collecting data.

3. Results

A total of 64 ballots were completed. The average loudness and standard deviation for the eight noise bands is given in Table II. The data show the expected trend. Given that the critical band at 1000 Hz is 160 Hz (Zwicker, 1961), one expects that loudness will not grow much between the first and second step; the data bear out this expectation. Beyond the first step, loudness grows almost linearly with increasing bandwidth.

The standard deviations of the data in Table II are quite large. Examining the original data showed that there were 16 ballots with nonmonotonic dependences, where estimates decreased for increasing bandwidth. Eliminating those ballots led to the data given in Fig. 1 averaged over $N = 48$ listeners. The figure shows a break in slope between the second and third steps, near a bandwidth of 180 Hz. However, this bandwidth is close to the reference bandwidth, and the evidence for a break is weak. The demonstration might be more convincing if it included a few bands with bandwidth less than 145 Hz.

4. Impression

After the second step, the noise becomes monotonically louder as the band becomes wider. However, the loudness increase on the third step does not seem impressive. If listeners are told to expect a sudden increase in loudness at some point in the sequence, they will not locate it at the second step. They will choose some other step in the sequence where the loudness change is more dramatic than on the third. Only the analyzed data, as plotted in Fig. 1, can show the expected break in the loudness function at the critical bandwidth.

C. Demonstration 6: Frequency response of the ear

The human ear does not have a perfect frequency response. It is more sensitive in the range 1000 to 5000 Hz than in lower or higher frequency ranges. The sensitivity is enhanced in the 2000- to 4000-Hz range by outer ear reson-

TABLE II. Demonstration 3, critical bandwidth by loudness comparison: For eight steps of increasing noise bandwidth, the table gives the loudness estimates averaged over 64 listeners. The loudness of a reference noise band was set at 100.

Step	Bandwidth (Hz)	Mean	s.d.
1	145	100.4	1.5
2	167	102.7	6.2
3	192	109.0	11.8
4	221	116.6	18.1
5	254	124.8	23.0
6	292	135.5	28.9
7	335	149.4	34.2
8	386	165.2	43.4

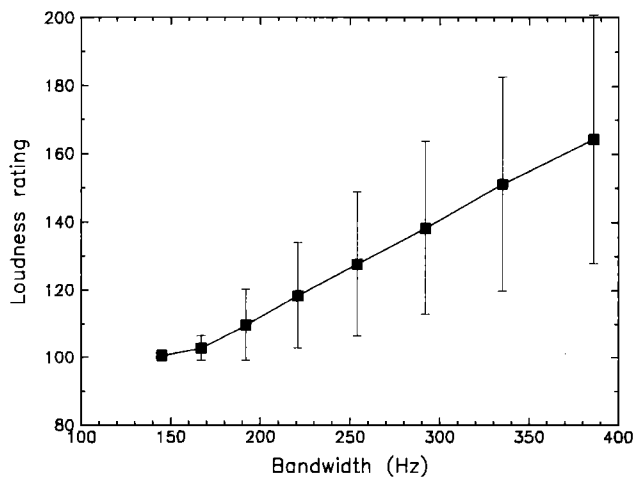


FIG. 1. Demonstration 3, critical bands by loudness comparison: The loudness ratings by listeners for eight noise bands with equal energy but different bandwidths. Ratings were made with respect to the noise with the smallest bandwidth, arbitrarily given a rating of 100. There were $N = 48$ listeners, selected because they gave nondecreasing ratings for increasing bandwidth. The error bars are two standard deviation in overall length.

ances; it is reduced by a middle ear transfer function that is less efficient for high and low frequencies. These elements of the overall transfer function, which are entirely acoustical, are best seen in a plot of the threshold intensity for hearing a sine tone as a function of tone frequency. Measurements of the transfer function by estimations of loudness at higher intensities are affected by the fact that as the level increases, the excitation pattern caused by a low-frequency tone spreads along the basilar membrane, enlisting more and more neurons to sing its tune. Thus, so-called "equal-loudness" contours show a flatter frequency response than the threshold curve.

1. The disc demonstration

The disc demonstration attempts to measure the response of the ear by a threshold measurement. First the gain of the reproduction system is set by a 1000-Hz tone on a "calibration track" on the disc. When the gain is correct, the 1000-Hz tone is barely audible.

The next track measures the threshold of audibility for sine tones of different frequency. First, a 125-Hz tone is tested, its level decreasing in 10 steps of 5 dB each. The listener's task is to count the number of steps that he can hear. Then the experiment continues with similar downward staircases for tones of 250, 500, 1000, 2000, 4000, and 8000 Hz. For each frequency the staircase is presented twice.

2. Presentation

A priori, the idea of measuring auditory threshold in a lecture hall or classroom situation seems only slightly less preposterous than putting a man on the moon with a large slingshot. Background noise, standing waves in the room, and an imperfect sound reproducing system all contribute importantly to error in such a measurement. The demonstration was done anyway.

The calibration track was set to be more than barely audible at the front of the room near the loudspeakers. An advantage of the calibration track is that it gets the audience quiet in preparation for the experiment. Listeners filled out paper ballots with seven blanks, one for each frequency. During each staircase the experimenter pointed to numbers 1–10 written on the blackboard.

3. Results

There were 100 ballots returned. Average step numbers and corresponding levels were given in Fig. 2. There were nine listeners who did not hear any of the steps at 125 Hz.³ Their scores for this frequency were taken to be zero. Also plotted on Fig. 2 are threshold data from Fletcher and Munson (1933), arbitrarily shifted to agree at 1000 Hz. The principal discrepancy due to classroom presentation is that the normal threshold elevation at low frequencies is exaggerated. This is probably caused by low-frequency background noise in the environment.

4. Variability

The error bars in Fig. 2 show that there was considerable variability; the average *sd* over seven frequencies was 5.5 dB. It seemed possible that much of the variability arose because some listeners were near the signal sources while others were distant, a factor that might lead to a constant attenuation for all frequencies, or equivalently an increase in broadband noise background. To test for an attenuation effect, each listener's responses (R_f) were normalized by subtracting his response at 1000 Hz (R_{1000}) to obtain a difference statistic. If the responses at different frequencies f were independently distributed then the variance of the distribution of differences ($R_f - R_{1000}$) should be the sum of the variance for f and the variance for 1000. If a common factor, such as a constant attenuation for all f , correlated the distributions then the variance of the difference distribution should be less than the sum of individual variances.

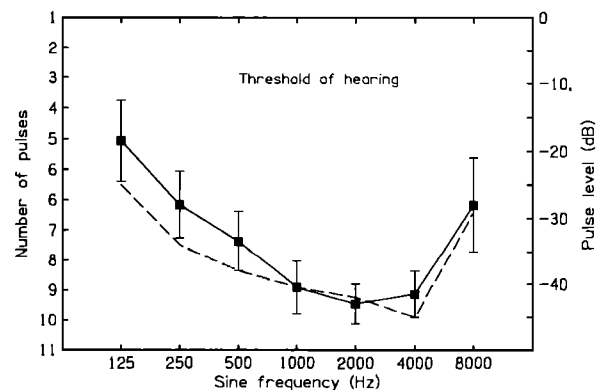


FIG. 2. Demonstration 6, frequency response of the ear: The data points show absolute thresholds determined as the number of audible pulses in demonstration 6. There were $N = 100$ listeners in a room. The decibel units on the right-hand axis show pulse level relative to the first pulse. The error bars are two standard deviations in length. The broken line shows the data of Fletcher and Munson (1933), displaced vertically to fit at 1000 Hz.

The calculation showed that the variance of the distribution of differences, $R_f - R_{1000}$, was indeed less than the variance of R_f plus the variance of R_{1000} , but it was only 22% less. It appears therefore, that constant attenuation, independent of frequency, accounts for much less than half the variance observed in the data.

5. Impression

Whether this demonstration makes its point to the individual listener is a question that can be answered by looking at differences of successive responses for each listener. If the demonstration works, then the differences should be positive as the frequency increases from 125 to 4000 Hz (lower thresholds for increasing frequency) and negative as the frequency increases from 4000 to 8000 Hz.

Successive difference are shown in Fig. 3. The signs agree with expectation, except for the difference between 2000 and 4000 Hz. The error bars show that the *a priori* reservations concerning this demonstration were not justified. The demonstration left the correct impression on more than 84% of the students on four out of six successive frequency comparisons.⁴

D. Demonstration 7: Loudness scaling

The psychophysical technique known as “magnitude estimation” attempts to measure sensation directly by asking subjects to give a number that rates the magnitude of a sensation. It is often combined with “magnitude production,” in which a subject changes the magnitude of a stimulus properly to match a number (Stevens, 1971).

1. The disc demonstration

Demonstration 7 uses the method of magnitude estimation. It presents 20 bursts of broadband noise at nine different levels, ± 20 , ± 15 , ± 10 , ± 5 , and 0 dB (relative scale). Each burst is preceded by a reference burst at 0 dB,

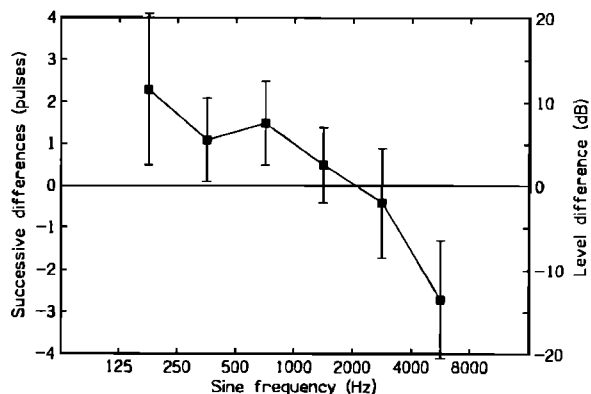


FIG. 3. Demonstration 6, frequency response of the ear: The data show the average and s.d. of the difference in the number of audible steps (pulses) between two successive frequencies in the threshold of hearing experiment. Data points are plotted between the two frequencies involved in the difference. Each step in this demonstration was 5 dB less intense than the previous step. Error bars are two standard deviations in overall length.

and the two bursts constitute one trial. All bursts are 1 s in duration. Reference and test bursts are separated by 250 ms, and trials are separated by 2250 ms. Trials with different test levels are randomized. The experiment with 20 trials is preceded by an orientation track presenting the reference noise (0 dB) and the strongest and weakest noises (± 20 dB).

2. Presentation and results

The listeners were asked to rate the loudness of each test burst on a scale where the reference burst was assigned a rating of 100, and other sounds were rated according to perceived loudness; for example, a noise twice as loud as the standard should be rated 200. No suggestions were made about appropriate ratings for the strongest or weakest sounds on the orientation track. Listeners filled out a paper ballot with blanks marked 1 through 20. As the experiment progressed, numbers 1 through 20 were written on a blackboard.

There were 64 complete ballots returned for this demonstration. Ratings for each of the 20 trials were averaged over the 64 listeners; the averages are given in Table III. Except for the trials at -20 dB, there is good agreement among the average ratings in any column. Indeed, the consistency of the results is an effective advertisement for the “direct” psychophysical method. In the best case (the three ratings at $+15$ dB) the standard deviation ($N - 1 = 2$ weight) is only 2.6% of the mean.

3. Psychophysical laws

The average results of the scaling experiment from the bottom line of Table III were plotted against the noise inten-

TABLE III. Demonstration 7, loudness scaling: For each of 20 trials the rating, averaged over 64 listeners, is given in the columns headed by the noise level for the trial.

Trial	Level (dB)								
	-20	-15	-10	-5	0	5	10	15	20
1									219
2				84					
3	45								
4					98				
5			62						
6									253
7						127			
8							160		
9		44							
10					104				
11			60						
12									213
13									279
14				80					
15							140		
16		40							
17				75					
18	29								
19						123			
20								208	
Average	37	42	61	80	101	125	150	213	265

sity on various coordinates to test alternative psychophysical laws.

Figure 4(a) tests the idea that loudness is proportional to sound pressure. If the idea is correct, the plot should be a straight line. The pronounced curvature of this plot shows that the idea fails. Loudness grows much more slowly than

sound pressure. A plot of the loudness data against sound intensity is curved in the same way, but even more so.

Figure 4(b) tests the logarithmic scale derived by Fechner from Weber's law. According to this scale, loudness should be proportional to sound level as given on the decibel scale. This plot is also curved, but in the opposite way, showing that loudness increases more rapidly than the decibel scale would predict.

Figure 4(c) tests the power law, as suggested by Stevens,

$$\psi = 100(I/I_0)^p,$$

where ψ is the loudness and I is the intensity. Figure 4(c) is in the form

$$\begin{aligned} \log \psi &= \log 100 + (p/10) 10 \log(I/I_0) \\ &= 2 + (p/10)L, \end{aligned}$$

where L is the level in dB. The slope of the line gives the exponent p . Figure 4(c) has a slope of 0.22, to be compared with the textbook value of 0.3. Because of the regression effect, it is expected that the exponent obtained from estimation data should be on the low side. For example, the exponent for magnitude estimation quoted by Stevens and Greenbaum (1966) is 0.27. Class data from A. J. M. Houtsma and F. A. Bilsen (1992) have given exponents of 0.22 and 0.25, in good agreement with the results of the present experiment. These exponents correspond to slopes of 14 and 12 dB per doubling of loudness.⁵

4. Impression

Students are initially skeptical about the technique of magnitude estimation, and simply doing the demonstration does not, by itself, change anyone's opinion. What is impressive about the demonstration is the data, as shown in Fig. 4. The instructor who wants to persuade students of the utility of the method needs to do the calculations and to show the results to the students in the next session. Really impressive would be a setup where students entered their numerical estimates via distributed parallel data entry devices, with the results analyzed and displayed immediately.

E. Demonstration 8: Temporal integration

A sound of short duration is less loud than an equally intense sound of longer duration. Apparently the sensation of loudness involves an integration of intensity, or power, over time, i.e., it depends upon an energetic quantity. Similarly, signal detection, either in noise or in quiet, is generally found to depend upon the dimensionless ratio of signal energy to noise power spectral intensity. Intensity is, of course, not integrated indefinitely. As the duration of a tone or noise is increased, there comes a point at which the auditory system gains no advantage from further increases. This duration is the "integration time," and most experiments find that it is about 200 ms.

1. The disk demonstration

Demonstration 8 studies the detectability of bursts of broadband noise having different durations. There are indi-

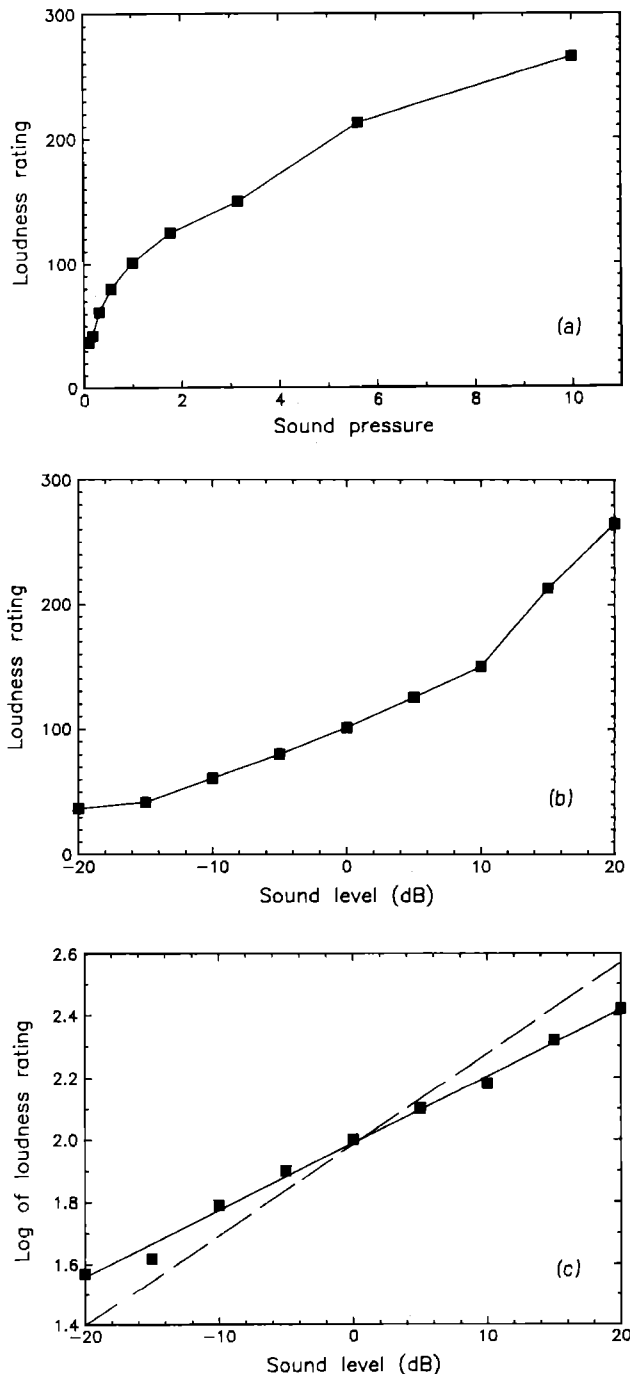


FIG. 4. Demonstration 7, loudness scaling: The loudness ratings, averaged over 64 listeners, for broadband noises of nine different levels, are plotted (a) against the rms sound pressures of the noises, and (b) against the sound levels of the noises in dB. In (c) the log of the loudness ratings are plotted against the levels in dB. The slope of the solid line in (c) establishes the exponent in the psychophysical power law, 0.22. For comparison, the dashed line has a slope of 0.3. In part (a) the sound pressure p is obtained by inverting $L = 20 \log(p)$, where L is the sound level.

vidual test sequences for the durations of 1000, 300, 100, 30, 10, 3, and 1 ms. Each sequence consists of pulses at eight decreasing levels (0, -16, -20, -24, -28, -32, -36, and -40 dB) in the presence of masking noise. Each test sequence is presented twice. The listener's task is to find detection threshold by counting the number of audible steps.

2. Presentation and results

The demonstration was reproduced at a relatively low level, where the most intense noise was a comfortable sound. A total of 111 students heard the demonstration. Eight of them did not understand the instructions or for some other reason left entries blank on the response form, leaving $N = 103$. The average over two presentations and over 103 listeners, as well as the standard deviation over listeners is given in Fig. 5.

The plot of the data shows that temporal integration is taking place. The data for the smallest durations agree approximately with the dashed reference line having a slope of 10 dB per decade of duration. That slope corresponds to simple integration of power to get energy, as expected from energy detection theory.

Deviation from the theoretical ideal might be expected because the reverberation in the room extends all durations beyond their nominal values, an effect that has greater consequences for small durations than for large. Thus the experimental slope ought to be less steep than the theoretical. In fact, however, the experimental slope appears to be *steeper* than the theoretical.

For most of the durations, listeners heard all eight pulses. Performance ran into this ceiling for a duration between 30 and 100 ms. It seems likely that if the demonstration had included pulses with still lower levels than temporal integration would have been observed for durations longer

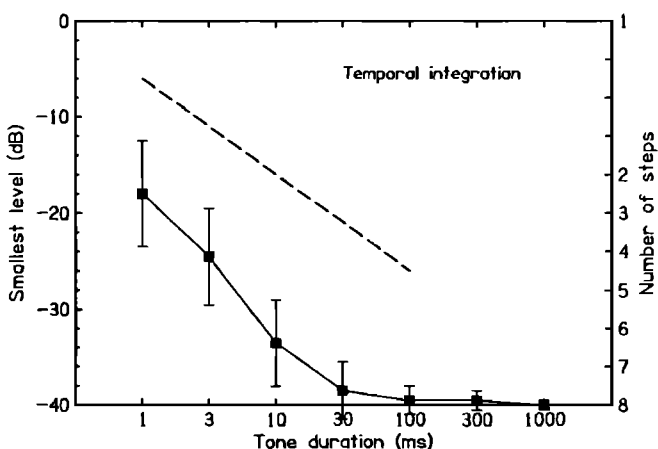


FIG. 5. Demonstration 8, temporal integration: The number of audible steps of broadband noise pulses averaged over 103 listeners as a function of the duration of the pulses. Pulse level, relative to the first pulse, is given on the left. Error bars are two standard deviations in length. The broken line is a reference with a slope of -10 dB per decade of duration. The actual position of the broken line is arbitrary.

than 100 ms, in better agreement with careful psychoacoustical experiments.

3. Impression

The fact that most of the listeners heard all eight of the noise pulses for more than half of the durations made the demonstration seem to be of doubtful value at the time that it was done. When the data were analyzed and the plot was drawn for thresholds at the shortest durations, the value of the demonstration became apparent.

4. Auditory induction

It has been noted that this demonstration suffers from an artifact whereby auditory induction (Warren *et al.*, 1972) causes listeners to hear more noise pulses than are in fact presented (Hartmann, 1989). The effect was strong with the listeners in this experiment. Of the 103 listeners, 31 reported on their answer forms that they heard more than eight pulses, nine or occasionally ten. It is unlikely that these responses were response errors; the experimenter stood at the blackboard pointing to the numbers 1 through 8 in sequence for each staircase. The possibility of auditory induction is a price that one pays for the (otherwise effective) regular descending staircase in a noise background. When a listener responded with a number greater than 8, the answer was taken to be 8 in calculating the statistics above.

F. Demonstration 9: Asymmetry of masking by pulsed tones

It is generally found that low-frequency tones mask high-frequency tones more effectively than high-frequency tones mask low-frequency tones, at least when the level of the masking tone is high. This psychophysical asymmetry of masking is paralleled by a comparable asymmetry in physiological tuning curves observed in electrical recordings from eighth-nerve fibers, again when levels are high. Both effects are almost certainly associated with an asymmetrical spread of excitation on the basilar membrane with increasing level, the excitation in the high-frequency tail of the pattern growing more rapidly than the excitation in the low-frequency tail.

1. The disc demonstration

Demonstration 9 attempts to demonstrate the asymmetry of masking with a two-part experiment. In the first part, a 1200-Hz pure tone is the masker and a 2000-Hz pure tone is the target. In the second part, the roles of the 1200- and 2000-Hz tones are reversed. As in other disc demonstrations, the threshold for the detection of the target is determined by decreasing the target level in a staircase fashion. The listener's task is to count the number of steps for which the target can be heard.

For each step of the staircase, i.e., each level of the target, there is an eight-pulse stream, as shown in Fig. 6. The masker occurs on each pulse; the target occurs only on the even-numbered pulses; therefore, it occurs four times. This repetitive stream helps the listener distinguish between masker and target tones.

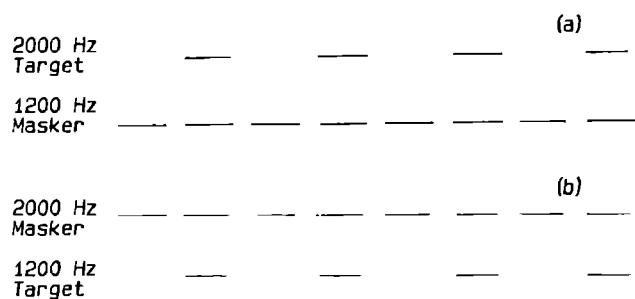


FIG. 6. Demonstration 9, asymmetry of masking: Lines show the sequence of tones in the two parts of demonstration 9. After each set of eighth pulses of the masker (and four of the target), the target level is reduced by 5 dB.

The reduction in the level of the target tone between first and second streams is 15 dB. Thereafter, the reduction is 5 dB per stream. There are 10 streams, and therefore an overall reduction of 55 dB from start to finish.

2. Presentation

It takes some coaching to get listeners to understand that they are supposed to count streams and not pulses. Therefore, the entire set of 10 streams with the 1200-Hz masker was presented while the experimenter pointed to numbers 1 through 10 written on the blackboard in synchrony with the experiment. Then the demonstration was restarted, and listeners counted the number of audible streams for both 1200- and 2000-Hz maskers.

The experiment was done with Science of Sound classes in 1990, 1991, and also with group 92W. The playback levels were different in different years, with results described below.

3. Results and impression

In 1990, the demonstration was presented at a comfortable level to 61 listeners. The average number of audible steps was 6.8 for both the 1200-Hz masker and the 2000-Hz masker. The minimum number of steps heard was 3 (one listener), the maximum was 10 (one listener). This represents a considerable range of levels (35 dB). However, the difference between conditions is more interesting than the absolute values. The upward spread of masking (asymmetry) was defined as the number of steps heard for the 2000-Hz masker minus the number of steps heard for the 1200-Hz masker. The distribution of the upward spread of masking observed in 1990 had a mean of precisely zero and a standard deviation of 1.74 steps, equivalent to 8.7 dB.

According to these data, the demonstration failed to show any asymmetry of masking, contrary to expectation. It was conjectured that the failure was the result of standing waves in the room. Given that standing waves typically lead to a room transfer function with valleys as deep as -15 dB, one might expect the demonstration to fail in any lecture room situation. This demonstration, where a tone of one frequency is masked by a tone of a different frequency, was contrasted with masking of a tone by noise where it can be

expected that the room transfer function affects both the signal and the noise similarly (Hartmann, 1990).

The above argument is, however, an argument for large variability in the measured upward spread of masking; it is not an argument for zero mean value. Therefore, the experiment was repeated in 1991 with 31 listeners at an unpleasantly high level. The average number of audible steps was 4.9 for the 1200-Hz masker and 8.6 for the 2000-Hz masker. The upward spread of masking was 3.77 steps (s.d. = 2.28), equivalent to 18.8 dB (s.d. = 11.4 dB). It is evident that on this occasion, the demonstration revealed considerable asymmetry of masking, as expected. In fact, because almost half the listeners heard all 10 steps for the 2000-Hz masker the upward spread might actually have been more than 18.8 dB.

Because of the discrepant results obtained in different years, demonstration 9 was repeated again with group 92W, once at the beginning of the session and once at the end. In the beginning, it was presented at a level of 75 dB SPL, read on a sound level meter in the middle of the audience, and at the end it was presented at 60 dB. At 75 dB, the upward spread of masking was 17.7 dB (s.d. = 6.4 dB), similar to the result from 1991. At 60 dB the upward spread was 5 dB (s.d. = 7.4 dB), similar to the result from 1990. A histogram showing the upward spread of masking, measured in steps numbers, for the two different intensity levels, as heard by the 92W group, is given in Fig. 7.

The conclusion of the experiment is that asymmetry of masking is easily observed, despite the vagaries of standing waves in a room, when the level is high. However, it is barely observed, or not observed at all, when the level is low. The obvious message for the instructor who wants to convince an

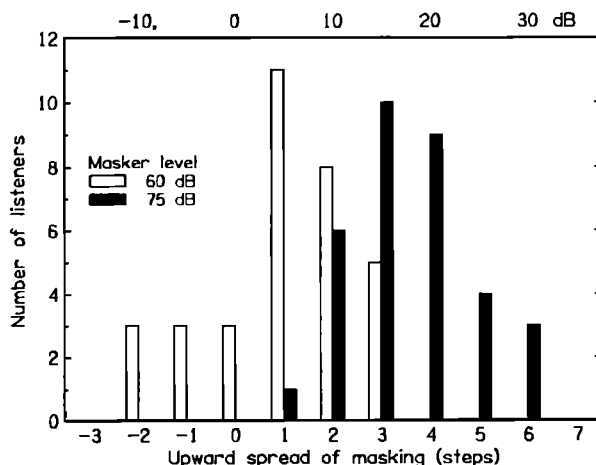


FIG. 7. Demonstration 9, asymmetry of masking: The histogram bars show the number of listeners who heard n more pulse streams with the 2000-Hz masker (1200-Hz target) than they heard with the 1200-Hz masker (2000-Hz target). The value of n is defined as the upward spread of masking and appears on the lower horizontal axis. On the upper horizontal axis is the corresponding difference in target levels, equivalent to 5 dB per step. The open bars are for 60-dB masker level; the filled bars are for 75-dB masker level. The total N is 33.

audience that masking is asymmetrical is that the demonstration should be presented at a high level. At a high level listeners appreciate the point of the demonstration right away.

G. Demonstration 10: Backward and forward masking

Backward and forward masking, also known as “nonsimultaneous” masking, probe temporal aspects of the auditory system. In backward masking a brief tone pulse is made less detectable by a noise burst that follows the tone. Forward masking is the same except that the masking noise burst precedes the tone. Nonsimultaneous masking can be contrasted with “direct” masking, as in demonstrations 2 and 9. Besides the interest in temporal effects in masking, nonsimultaneous masking has the advantage that it avoids nonlinear interaction between the masker and the signal.

1. The disc demonstration

Demonstration 10 has three distinct parts. The first presents the signal alone, with its level decreasing by ten steps in a staircase fashion, -4 dB per step. The second part presents the same staircase in the backward masking condition. First the delay between signal and masker is 100 ms. Next the delay is 20 ms, and finally it is 0 ms. For each delay the staircase is presented twice. The third part of the demonstration presents the forward masking condition. Delay durations and procedures are the same as for backward masking. In all three parts the listener’s task is to find a detection threshold by counting the number of audible signal pulses.

2. Presentation and results

The demonstration was presented to 87 listeners in 1992. The staircase in quiet was first presented to illustrate the method and to make sure that all listeners could hear all ten steps of the staircase in the absence of any masking. Next the demonstrations of backward and forward masking were done. Each demonstration was presented three times, the first pass was for orientation only, the second and third passes were for taking data. Listeners indicated the number of audible steps on a written ballot.

The raw data were subjected to a screening process. Only complete ballots were counted. Further, some ballots were rejected as nonmonotonic whenever a greater number of audible pulses was reported for shorter delays and whenever the reported number of audible pulses did not depend at all on the duration of the delays. It was decided that these listeners were confused about the task and that their data would be unreliable.

There remained 74 valid ballots for backward masking and 58 for forward masking. Data showing the number of audible steps for each duration are given in Table IV. Data in units of steps may be converted to decibel units by multiplying by 4. The table shows that forward masking is more effective than backward masking, especially as the delay becomes small. The overall increase in thresholds as the delays decrease to zero was about 16 dB for both backward and forward masking. This can be contrasted with increases of 40 or 50 dB for precision headphone experiments (Fastl,

TABLE IV. Demonstration 10, backward and forward masking: The top part of the table gives the number of audible signal pulses for three different durations of the gap between the masker and the signal pulse. Successive pulses differed in level by 4 dB. Standard deviations in parentheses are also given in number of pulses. The bottom of the table describes the data selection.

Gap duration (ms)	Backward masking	Forward masking
100	6.9 (1.1)	5.1 (2.2)
20	5.3 (1.3)	2.9 (2.4)
0	2.8 (1.7)	1.3 (1.8)
Number counted	74	58
Number nonmonotonic	8	24
Number incomplete	5	5
Number of listeners	87	87

1976/77). The difference can be attributed to the smearing of temporal details by the reverberant response of the room, and to the fact that naive listeners become confused about the sound of the target pulse.

3. Impression

For the novice listener demonstration 10 may be the most confusing of all the demonstrations on the disc. Table IV shows that many ballots were invalid, as many as one third of them for forward masking condition. Given a brief (10-ms) sine signal in a narrow-band masker listeners are confused about what the signal is, particularly as the delay shrinks to zero. The standard deviations in the top of Table IV look large, especially in view of the original screening done on the data.

When this demonstration is presented by headphones, a listener hears more signal pulses for delays of 20 and 100 ms than when presented by loudspeakers in a room. When heard with headphones, the forward masking condition nearly covers the full 40-dB range, from 10 signal pulses heard for a delay of 100 ms to one or zero pulses heard for a delay of 0 ms.

H. Demonstration 11: Pulsation threshold

The continuity effect (Warren *et al.*, 1972; Warren, 1984) is an auditory illusion wherein listeners hear a pulsed signal as continuous when the silent gaps between the signal pulses are filled with noise. In order for the illusion to work, the noise in the gaps has to be intense enough that it would mask the signal if the signal were actually present during the gaps.

The continuity effect has been turned into a psychoacoustical method called “pulsation threshold.” The greatest signal level for which the signal is heard as continuous is defined as the pulsation threshold (Houtgast, 1972). The method has been regarded as an alternative to forward masking; suppression is evident in both, whereas direct masking does not show suppression.

1. The disc demonstration

In demonstration 11, 125-ms bursts of a 2000-Hz sine tone (*S*) alternate with 125-ms bursts of noise (*N*) (1875–

2125 Hz). The noise level remains constant and the tone level remains constant for four alternations, i.e., for the sequence $N, S, N, S, N, S, N, S, N$. Then the tone level is decreased by -1 dB for the next sequence of four alternations. This pattern continues for 15 steps. Therefore, the maximum reduction in tone level is 14 dB.

2. Presentation

Students were introduced to the pulsation threshold by a demonstration of the continuity effect using music. A recorded orchestral performance of Tchaikovsky's Sleeping Beauty Waltz was pulsed, 300 ms on and 300 ms off. Then masking noise was added in the silent gaps. The noise level was gradually increased until pulsation threshold was reached, at which point listeners could hear the orchestra playing continuously throughout the gaps. The music in the gaps originated entirely within the listener's heads. There was a general consensus that this was a dramatic and surprising effect.

After the musical introduction, the disc demonstration was presented at a comfortable listening level for the masking noise as heard in the front of the lecture room. The demonstration, as it appears on the disc, goes through the 15 steps only once. However, the demonstration is short, and it was played twice, giving the listeners the chance to make two judgments of pulsation threshold.

3. Results

A total of 87 students heard the demonstration. Three did not make estimations on both passes, leaving a population of $N = 84$. There were 7 listeners who claimed that the tone never sounded continuous, though some of these suspected that with another step or two on the staircase they might have been willing to agree that the tone was continuous. This left a population of $N = 77$. For these, the average step number from the first pass was 11.1 (s.d. = 2.5), and from the second pass, a step number of 10.5 (s.d. = 2.6). The two step numbers from each student were averaged and the results were plotted in the histogram of Fig. 8.

The average step number over both passes is 10.8. Including the 7 listeners for whom the tone did not reach pulsation threshold, the median step number is also 10.8. The demonstration can be counted as a success; more than 90% of the listeners heard the continuity effect.

4. Impression

Pulsation threshold is an entertaining effect with immediate appeal. It needs to be demonstrated in an acoustically dry environment. Excessive reverberation will kill the effect, as the author once learned to his chagrin in an Austrian baroque hall with a wood floor and plaster walls and ceiling. The problem is that the masker reverberates throughout the interval that is supposed to be for signal only making the signal hard to hear at all times.

I. Demonstration 12: Dependence of pitch on intensity

The pitch of a sine tone is a malleable percept. It can be affected by a wide variety of changes in the listening condi-

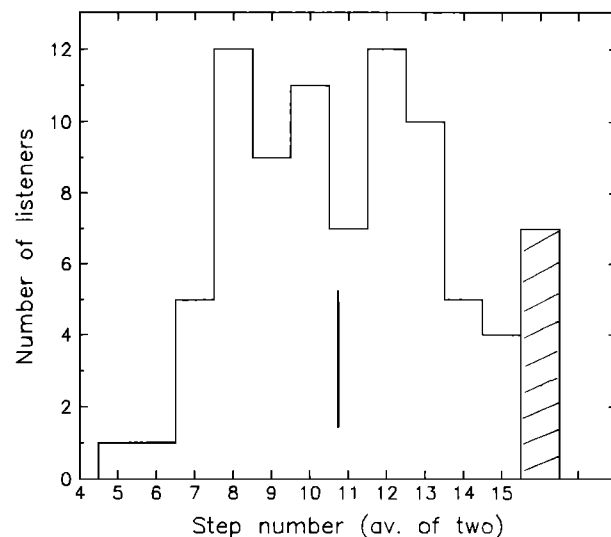


FIG. 8. Demonstration 11, pulsation threshold: The ordinate gives the number of listeners who heard the 2000-Hz tone as continuous on the n th step where n is given on the abscissa. Values of n were averaged over two presentations. On successive steps the tone level was reduced by 1 dB. The hatched bar at the right is for seven listeners who never heard a continuous 2000-Hz tone. The total N is 84. The vertical line in the center gives the average.

tions. For example, the pitch of a sine changes with increasing intensity, as was discovered in experiments by Stevens (1935). Stevens actually did the experiments very badly and obtained pitch changes far in excess of those found by subsequent experimenters using better methods. However, he did get the *signs* of the changes right. The rule known as Stevens' rule is about the signs. It says that with increasing intensity, the pitch of a sine goes down if the frequency is less than 1000 Hz and goes up if the frequency is greater than 1000 Hz. Further work has shown that Stevens' rule applies only on the average. For any given listener, the pattern of pitch increases and decreases is idiosyncratic and not at all monotonic with frequency. Nor is pitch always monotonic with intensity at a particular frequency (Verschuere and van Meeteran, 1975; Burns, 1982).

1. The disc demonstration

Demonstration 12 tests the dependence of pitch on intensity for sine tones of six different frequencies: 200, 500, 1000, 2000, 3000, and 4000 Hz, in that order. Prior to the test there is a calibration track with a feeble 200-Hz tone. The playback level should be set so that this tone is barely audible. If the level is set for comfortable listening to the narration and not for an audible calibration tone, then it is likely that the listeners will not hear the first tone of the test (200 Hz) at all.

The demonstration continues with tests at the six frequencies. Each tone is first presented at a low level, then at a level 30 dB higher. The listener's task is to say whether the pitch goes up, or down, or stays the same as the intensity is increased. The six-frequency test takes less than half a minute.

2. Presentation and results

Demonstration 12 was presented to the 33 listeners of group 92W. It was presented four times, one pass for orientation, and then three passes for listeners to record their responses on a ballot form. Allowed responses were: "goes up," "goes down," or "stays the same." The high level tone registered 80 dB on a sound level meter in the middle of the room.

In analyzing the listener responses, the first data pass was ignored and data from the last two passes were summed. Responses were given numerical values: goes up = $+1/2$, goes down = $-1/2$, stays the same = 0. Therefore, for each frequency, the possible scores for two passes were 1, $1/2$, 0, $-1/2$, and -1 . There were two ways to get a score of zero, two responses "no change," or one response "goes up" and one response "goes down."

The predictions of Stevens' rule are that scores should be negative (pitch goes down with increasing intensity) for the first two tones, 200 and 500 Hz and should be positive (pitch goes up with increasing intensity) for the last three tones, 2000, 3000, and 4000 Hz. Figure 9 shows that the average data agree perfectly with this prediction, although the error bars are rather large.

As expected, not all listeners perceived the pitch dependence on intensity in accordance with Stevens' rule, but most of them followed the rule, at least in part. The breakdown for the 33 listeners was as follows: There were 9 listeners who said that pitch increased with increasing intensity preferentially for high-frequency tones (the high-frequency part of Stevens' rule only). There were 6 listeners who said that pitch decreased with increasing intensity preferentially for low-frequency tones (the low-frequency part of Stevens'

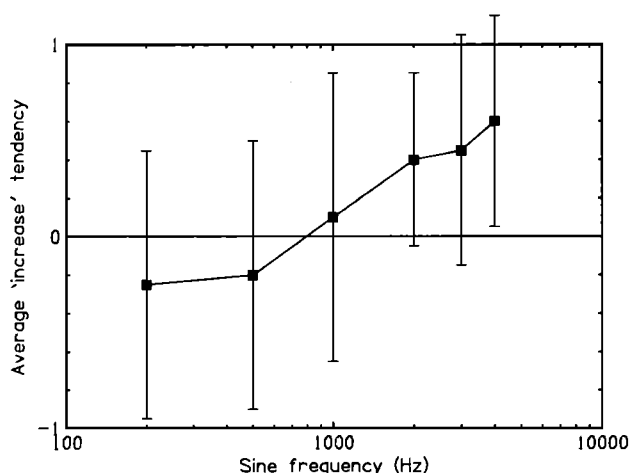


FIG. 9. Demonstration 12, dependence of pitch on intensity: The tendency for the pitch of a sine tone to increase (positive values) or decrease (negative values) with increasing intensity is shown for sine tones of six different frequencies. A score of $+1$ means that pitch increased 100% of the time, a score of -1 means that the pitch decreased 100% of the time. Scores were averaged over two passes by each of 33 listeners. Errors are two standard deviations in length.

rule only). There were 9 listeners who agreed with both the high-frequency and the low-frequency parts of Stevens' rule. Finally, there were 9 listeners for whom no part of Stevens' rule applied.

J. Demonstration 14: Influence of masking noise on pitch

The shift of the pitch of a sine tone as noise is added is a second example of the malleability of pitch. The strongest of these noise-induced effects appears to be an upward shift in the pitch of a tone when noise is added in a frequency range below the frequency of the tone. (Terhardt and Fastl, 1971). The effect is considered to be strong enough and reliable enough that it is used as a tool in pitch research (e.g., Houtsma, 1981).

1. The disc demonstration

Demonstration 14 is straightforward. A 1000-Hz sine tone alternates with a 1000-Hz sine tone that is partially masked by noise, low-passed below 900 Hz. In all, there are eight pairs for comparison. The listener's task is simply to compare the pitches of masked and unmasked sine tones.

2. Presentation and results

The demonstration was presented to the listeners of group 92W at two different sound levels, 60 and 75 dB, as measured with a C-weighted sound level meter in the middle of the audience. The 33 listeners filled out a ballot in response to the question, "Which pitch is higher: with noise, without noise, no change." At 60 dB the responses were as follows: with noise 8, without noise 7, no change 18. At 75 dB the responses were as follows: with noise 13, without noise 10, no change 10.

The results of the demonstration were somewhat surprising. Only at the higher level did a plurality of the population agree that the pitch is higher in the partially masked condition. The conclusion, at least from this group of listeners, is that for most people the demonstration does not make the intended point. It is relatively more successful at high level than at low.

3. Impression

Those people who hear the pitch change with added noise believe the effect to be strong, and they cannot understand why everyone does not hear that effect.

K. Demonstration 15: Octave matching

When asked to choose the frequency of a tone that seems to be an octave higher than a standard tone, most listeners choose a frequency that is more than a factor of two higher than the standard (Ward, 1954). While most experiments measuring this "octave enlargement" effect have required listeners to adjust a tone to make an octave (Sundberg and Lindqvist, 1973; Ohgushi, 1983), the effect is also found in a forced-choice experiment (Dobbins and Cuddy, 1982).

1. The disc demonstration

Demonstration 15 illustrates the octave enlargement effect with an eleven-interval forced-choice experiment. Each interval consists of a pair of tones: a low tone with frequency of 500 Hz, followed by a high tone near 1000 Hz. The listener's task is to choose the pair that makes the best octave. The alternative pairs are presented one after the other, with the high-tone frequency f_2 increasing in sequence as a staircase: 985, 990, 995, 1000, 1005, 1010, 1015, 1020, 1030, 1035 Hz. The entire staircase of eleven tone-pair steps takes 44 s; it is presented twice.

2. Presentation

Listeners were given a ballot on which they answered the question, "For how many years have you regularly practiced a musical instrument or singing?" Listeners filled in two additional blanks on the ballot with the step number for the pair that made the best octave on each of the two passes of the staircase. The entire demonstration was presented once as a trial run; it was presented a second time for listeners to record their responses. During the staircase, the experimenter pointed to numbers 1 through 11 written on a blackboard to indicate each step as it occurred.

3. Results

There were 91 ballots returned complete with an answer to the question on "practice." The initial data set consisted of a response to each pass of the staircase. There were 13 ballots with only one response. It was assumed that a single response was given by listeners who did not change their minds on the second pass. Each single response was therefore counted twice. In the end there were $91 \times 2 = 182$ judgments in the data set. The agreement between passes tended to be good. On only six ballots did the two responses differ as many as three step numbers (i.e., a spread greater than 10 Hz, or 1%).

The average step number, including all listeners, was 5.82 corresponding to a high-tone frequency of about 1009 Hz and an octave enlargement of 0.9%. The standard deviation was 2.00 steps or 10.0Hz. The average of 0.9% is in reasonable agreement with data from the literature for this frequency range.

A histogram of the distribution of choices showed two peaks, one near step 4, the other at step 8. Because the ballots were originally sorted by years of musical practice, it was impossible not to notice that the two peaks came from different populations of listeners. Listeners with more musical experience tended to choose smaller octave enlargements. It was decided to make separate histograms for musically experienced and less experienced listeners.

The 91 listeners had between 1 and 20 years of practice, with a median value of 5.5 years. The histogram in Fig. 10 shows the choices of two groups, with approximately equal N . The hatched bars are for listeners with more than 5.5 years of practice ($N = 42$) the bars are for listeners with less than 5.5 years ($N = 49$). Figure 10 shows the considerable difference between these two groups. The modal response for the experienced listeners appears at the step 4, corre-

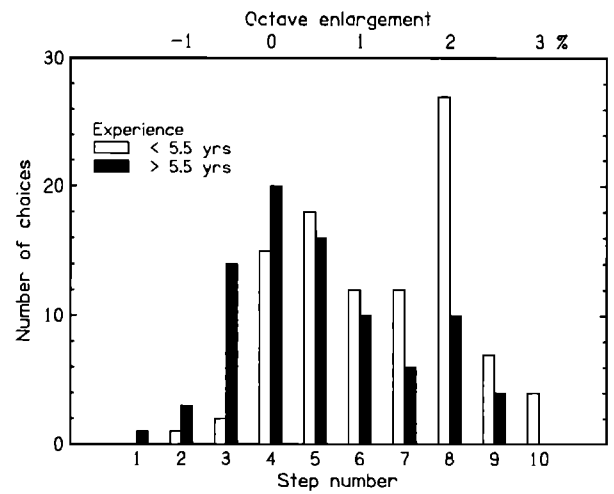


FIG. 10. Demonstration 15, octave matching: The ordinate gives the number of preference judgments in favor of the tone on step n as the best octave above 500-Hz reference tone. The value of n is on the abscissa. The top axis shows the enlargement percentage for step n . There were 182 preference judgments, two from each listener. The filled bars are from 42 listeners having practiced music for more than 5.5 years. The open bars are from 49 listeners having practiced music for less than 5.5 years.

sponding to a precise physical octave. However, the skew of the distribution leads to a significant mean shift [1.08 steps (s.d. = 1.91)] so that an octave enlargement appears in the mean response for the experienced group.

A literal interpretation of the data in Fig. 10 suggests that there may actually be two octaves, for both experienced and inexperienced listeners. One octave corresponds to a precise factor of 2; the other is enlarged by about 2%. However, the data from this demonstration should not be pushed too far. The sequence presented on the cd, monotonically increasing intervals with the expected value in the middle of the sequence, is good for a demonstration but is too biased to make a good experiment. What the data do say is that one should not rely entirely on the mean shift observed in an octave enlargement study but ought to look at the distribution of responses.

L. Demonstration 16: Stretched and compressed scales

Demonstration 16 consists of Ernst Terhardt's three delightful renditions of the famous drinking song extolling the glory of the Hofbrauhaus. But some renditions are more delightful than others. All of them are in two-part harmony, with the low bass part in the key of C. The melody is variously in the key of C, or C sharp, or B, i.e., it is in tune with the bass or it is mistuned half a step up or down.

Terhardt found that the stretched tuning (bass in C and melody in C sharp) is preferred by almost half the population. The explanation for the effect is that listeners have become accustomed to hearing harmonic relationships that are stretched by mutual interaction of the excitation patterns in a tonotopic neural representation. The memory trace established by harmonic sounds is then reflected in expectations for melodic intervals.

1. Presentation and results

Demonstration 16 requires few instructions from the experimenter; unlike most of the demonstrations on the compact disc, this one virtually runs itself. The narrator (Ira Hirsh) asks which of the three renditions sounds best in tune. The listener has to make the decision. There is no feedback.

The demonstration was presented to the 33 listeners of group 92W. It was presented two times, first for orientation, then for decisions; listeners indicated their preferences on a written ballot.

The voting showed that 17 listeners preferred the melody in the key of C; 15 preferred the melody in the key of C sharp, and 1 voted for the melody in the key of B. Thus, 45% of the listeners preferred the stretched tuning, a percentage that agrees with Terhardt's finding of about 40%.

Demonstration 16 on stretched tuning offers an opportunity for an interesting comparison with Demonstration 15 on the enlarged octave. Ward's (1954) original work on octave enlargement did not regard the effect as an isolated phenomenon. Instead, the melodic stretch implied by the enlarged octave was proposed as a basis for calculating individual musical scales, at least for sine tones.

It is therefore interesting to see whether there is any correlation between octave judgments in demonstration 15 and tuning judgments in demonstration 16. One might expect that those listeners with sizeable octave enlargement would prefer the stretched tuning in the two-part harmony.

To look for this correlation, the data from Demonstration 15 were reanalyzed, dividing the population into two groups, determined by preference in demonstration 16: the stretched-tuning group ($N = 15$) and the unstretched-tuning group ($N = 17$). Average step numbers for the perceptual octave were calculated for each group.

The calculation found that the average octave step number was 5.83 for the stretched-tuning group, and 5.80 for the unstretched-tuning group. These two averages are essentially identical. The comparison indicates, contrary to expectation, that octave enlargement and stretched tuning are mutually independent. This experimental result is awkward enough that it ought to be checked elsewhere with a different group of listeners.

2. Impression

Listeners are amazed when they hear this demonstration because they never expected to find themselves in any doubt about mistuning by a semitone. When it's over they very much want to know the right answer.

M. Demonstration 17: Frequency difference limen or jnd

The difference limen or just-noticeable difference (jnd) is the smallest change that is detectable by an observer. The value of a difference limen is important because it sets an upper limit on the information capacity of the corresponding communications channel. The difference limen for frequency sets the world's record for precision in a sensory system. Over the frequency range of speech, 100–8000 Hz, approximately 1500 different frequencies can be discriminated.

1. The disc demonstration

Demonstration 17 measures the frequency difference limen by requiring listeners to distinguish ever smaller frequency differences. There are ten groups of trials with four trials in each group. Each trial consists of a pair of tones, one with frequency of 1000 Hz, the other with frequency of $1000 + \delta f$ Hz. The higher tone is either the first tone of the pair or the second, and the listener's task is to say which. In the first group of pairs, the increment δf is 10 Hz. In the second group, the increment is 9 Hz, and so on, until, for the tenth group, the increment is only 1 Hz. In all, the demonstration requires the listener to judge 40 pairs of tones.

2. Presentation and results

Listeners heard the first two groups of pairs to become familiar with the task. Then the demonstration was restarted, and all ten groups were presented. Listeners indicated whether the pitch went up (U) or down (D) on a ballot with appropriate blanks.

A total of 64 listeners returned ballots. Of these, five listeners did not understand the forced-choice nature of the task and left blank spaces when they were uncertain. For eight listeners the number of correct responses was less than 20, i.e., less than chance. Data from these listeners were removed from the data set.

The remaining 51 ballots were analyzed according to the average number of correct responses for each frequency difference, i.e., for each group of four tone pairs. The results are shown in Fig. 11. The plot of the data has a shallow slope, making it difficult to estimate a jnd. Formally, the data cross the 75%-correct line at a frequency difference between 3 and 4 Hz. This can be compared with frequency difference limens at 1000 Hz found by Wier *et al.* (1977): 1.9 Hz at 40 dB SPL

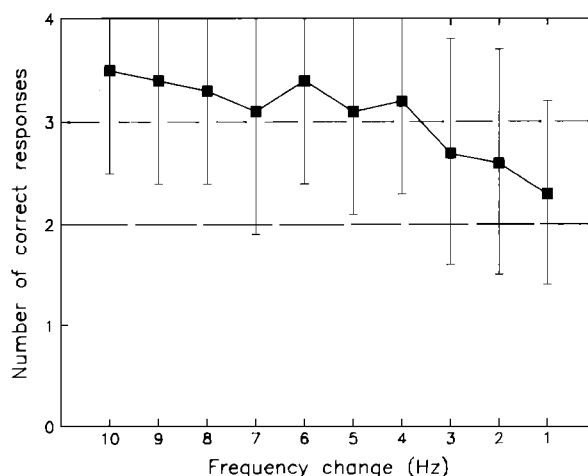


FIG. 11. Demonstration 17, frequency difference limen or jnd: For each frequency change, 10 through 1 Hz, there were four trials. The number of correct responses, averaged over 51 listeners, is shown on the ordinate. Broken lines at 2 (50%) and 3 (75%) give the levels for chance performance and the jnd, respectively.

and 1.3 Hz at 80 dB SPL. The sound levels heard by the listeners in the present experiment were between 40 and 80 dB. The discrepancy in measured jnds between the two experiments is probably due to an initial subject screening and improvement in performance given training on the task in the study of Wier *et al.*⁶

N. Demonstration 20: Virtual pitch

A periodic complex tone has harmonic partials and a definite pitch. The pitch corresponds, at least approximately, to the frequency of the first harmonic or fundamental. This is true whether or not the fundamental is actually present in the spectrum. The pitch that is synthesized by the brain from a collection of harmonics is known as virtual pitch (Terhardt, 1974). There is evidence (Walliser, 1969; Terhardt, 1972 a,b) that the virtual pitch of a complex tone is slightly lower than the pitch of a sine tone with the same period. The effect is magnified if the complex tone is missing lower partials.

1. The disc demonstration

Demonstration 20 presents five complex tones with fundamental frequency of 200 Hz. The first tone is comprised of the first ten harmonics, all of equal amplitude. The second tone is the same except that the fundamental (first harmonic) is removed. On the third tone both the first and second harmonics are removed, etc., so that the fifth tone consists of only harmonics 5 through 10. The tones are all 1 s long and there is a gap of 1 s between tones. Therefore, the entire demonstration lasts 9 s. It is repeated once.

The creators of the disc and accompanying booklet have (wisely) made the instructions for this demonstration vague. The narrator asks if the pitch of the complex tone changes as lower harmonics are removed.

2. Presentation and results

The demonstration was presented to the 33 listeners of group 92W. Instructions were left vague. The paper ballot included four lines, together with the instruction to describe pitch changes or other changes heard as the tone went through its four steps of evolution. The demonstration was played about eight times (about 16 presentations of the steps). Most of the listeners commented on the change in tone color, becoming "thinner, buzzier, brighter, more hollow, or more tinny," as the sequence progressed. Of greater interest are their comments on pitch. A synopsis of these comments is given in Table V.

It is evident from the table that listener responses varied. Comparing responses with years of musical experience revealed no particular correlation. The 4 listeners who responded "no change" evidently perceived the virtual pitch as expected. There is some evidence of the effect noted by Walliser and Terhardt from the 7 listeners who heard the pitch as increasingly flat.

Most of the listeners, however, heard the pitch go up at some point as the sequence progressed. It seems almost certain that this does not represent a small virtual pitch shift. Instead it is very likely that these listeners responded to spec-

TABLE V. Demonstration 20, virtual pitch: A synopsis of listener comments and the numbers of listeners who gave them. There is some double counting in categories marked *.

Comments on pitch	Number of listeners
Increasingly flat	5
Increasingly flat until last tone*	2
Increasingly flat after tone 3	1
No change	4
Increasingly higher	3
Increasingly higher after tone 2	6
Increasingly higher after tone 3	3
Higher only after tone 4*	5
Tones are mixtures of different pitches	4
No comment on pitch	2

tral pitches of the harmonics in the complex tones. Hearing out a harmonic of the complex tone was actually encouraged by an artifact in the demonstration, dropping harmonics in a regular sequence. When the lowest harmonic is dropped, the lowest harmonic remaining in the tone is highlighted. This effect was evident in the responses of the listeners who heard mixtures of tones, and those who indicated the actual musical intervals of the lowest harmonics as they increased.

3. Impression

Whether demonstration 20 succeeds or not depends upon one's intention. If the goal is to show that the virtual pitch (pitch of the missing fundamental) is maintained for different spectra, then it fails. A different demonstration, without the sequence promoting analytic listening, is needed for that.⁷ If the goal is to show how spectral pitch and virtual pitch can be confounded, it succeeds admirably.

O. Demonstration 25: Analytic versus synthetic pitch

A complex periodic tone with a missing fundamental and only a few harmonics has an ambiguous pitch. It is possible for a listener to analyze the tone according to its several harmonics. Alternatively, the listener may synthesize a virtual pitch corresponding to the missing fundamental. It is extremely rare for a listener to be able to do both, i.e., to hear both analytically and synthetically.

1. The disc demonstration

Demonstration 25 presents a sequence of two complex tones. The first has components at 800 and 1000 Hz; the second has components at 750 and 1000 Hz. The listener's task is to say whether the pitch goes up or down. An analytic listener hears out the low-frequency component descending from 800 to 750 Hz and says that the pitch goes down (by 112 cents, i.e., by about a semitone). A synthetic listener hears the first complex as the fourth and fifth harmonics of 200 Hz and the second complex as the third and fourth harmonics of 250 Hz. This listener says that the pitch goes up (by the interval of a major third, ratio = 5/4).

2. Presentation and results

The demonstration was presented to a 1992 class of 137 listeners. According to informal vote (show-of-hands) there were 114 who heard the pitch go down, 14 who heard the pitch go up, and 9 who could not decide. This ratio, about eight-to-one in favor of analytic listening, was typical of the lopsided results from informal experiments in 1990 and 1989.

The demonstration was presented to the 33 listeners of group 92W. According to written ballot, there were 27 who heard the pitch go down, 5 who heard the pitch go up, and 1 who said that he could hear it in both ways.

An unusual result was obtained in an informal vote by 42 listeners in 1991, with 29 hearing the pitch go down and 13 hearing the pitch go up. It is more common to find only 10% of a population voting for synthetic listening (Houtsma and Fleuren, 1991) than to find 30%. The paper by Houtsma and Fleuren also showed that if the harmonic order of the partials is increased, the demonstration is made more ambiguous in the sense that the percentages of listeners showing analytic or synthetic listening approaches 50–50. In fact, the original version of this experiment by Smoorenburg (1970) used harmonics 9 and 10 of 200 Hz and harmonics 7 and 8 of 250 Hz.

3. Impression

The one listener who claimed to be able to hear the pitch change as both upwards and downwards seems to be a unique individual. The great majority of listeners strongly hear only one direction and they are surprised to learn that anyone could hear it differently.

III. DISCUSSION AND CONCLUSIONS

Demonstrations that illustrate basic psychoacoustical principles pose a peculiar paradox. *A priori* one would think that psychoacoustics, concerned as it is with human perception, would be an ideal source of sonic experiences for classroom demonstrations. This idea is not actually wrong. However, much of what is interesting in psychoacoustics is concerned with small changes, thresholds for detection, just-noticeable differences, and judgments of ambiguous stimuli. Such delicate matters are, of course, the most treacherous kind of material for a classroom demonstration. That the creators of the ASA *Auditory Demonstrations* compact disc managed to make practical demonstrations out of such difficult material is a remarkable accomplishment.

The present article has evaluated 15 compact disc demonstrations of the kind that produce data. Seven of these involve detection thresholds, and three of them involve judgements of ambiguous stimuli. The evaluations were made over the course of three years, by testing the demonstrations in a classroom with an audience of N between 30 and 150. The response data were collected by written ballot.

The data show that most of the demonstrations work well; they make the points that they intend to make, though demonstrations 6, 9, and 14 are level dependent and demonstration 11 requires a short room reverberation time. De-

monstration 10 is so badly affected by reverberation (even of short duration) that it should be done with headphones.

Success or failure with the demonstrations can be measured in two ways, by the data they yield and by the impression that they leave upon the listener. Some of the demonstrations succeed only if data are collected and plotted. This is true of some of the loudness demonstrations, notably numbers 3, 7, and 8, because the demonstrations themselves leave no strong impressions. Collecting data is also necessary for some demonstrations involving pitch, particularly numbers 14, 15, 16, 17, and 25. On these pitch experiments, listeners may have strong impressions but individual differences are important. Only if data are collected does a coherent pattern emerge from the individual results. On demonstrations 15, 16, and 25 the crudest forms of data collection (e.g., show of hands) may suffice. Other pitch demonstrations require data averaging.

Some of the demonstrations reviewed above leave a successful impression on the listener without data collection, though even these demonstrations can benefit from data collection. These are numbers 2, 6, 9, and 11. Among the other demonstrations that leave a successful impression without data collection, and therefore are not reviewed above, are numbers 1, 4, 5, 16, 19, 22, 23, 27, 28, 29, 30, 31, 33, 34, and 35.

Instructors who wish to use the *Auditory Demonstrations* compact disc in class should heed the observation by Carterette and Kendall (1990) that the demonstrations are not self-contained pedagogical vehicles. The problems involved in using the compact disc are partly due to the wide diversity of formats and listener tasks in the demonstrations themselves, and partly due to the fact that psychoacoustics is not an easy subject to learn. Careful preparation by the instructor is required, employing all the art of the professor's trade, in order to prepare students to appreciate the message contained in a demonstration. The instructor who takes the time to explain the demonstrations in detail and to collect data systematically will find that *Auditory Demonstrations* is a valuable resource in psychoacoustics education.

ACKNOWLEDGMENTS

I am grateful to Dr. E. M. Burns, Dr. Laurent Demany, and Dr. A. J. M. Houtsma for useful discussions on many points of the demonstrations. Mr. Jian-Yu Lin helped set up session 92W and contributed to data analysis. This work was supported by a grant from the NIDCD of the National Institutes of Health.

¹Those demonstrations which are mostly experiential and qualitative are numbers 1, 4, 5, 18, 19, 21, 22, 23, 26, 27, 28, 29, 30, 33, 34, 35, and 36. Other demonstrations which are mostly qualitative but give hints of possible quantitative application are numbers 13, 24, 31, 32, and 37.

²The creators of the demonstration 2 intended to duplicate the levels of the corresponding demonstration (number 1) on the Harvard tapes of 1978. There the spectrum level was 45 dB below the level of the most intense signal pulse.

³A. J. Watkins has suggested setting the playback level using the 125 Hz steps.

- ⁴The average value of differences for successive frequencies is a statistic that could have been computed by subtracting means from the threshold data of Fig. 2. The standard deviations, however, required a calculation specialized to individual differences for successive frequencies.
- ⁵Watkins (1992) plotted individual loudness estimations for 33 listeners. Four of them were straighter on a linear dB scale (Fechner) and 29 were straighter on a log scale (Stevens).
- ⁶Watkins (1992) also found a mean between 3 and 4 Hz. His 20 listeners, self-selected as musical or nonmusical, led to group averages of 2.3 Hz ($N = 9$) and 5.0 Hz ($N = 11$), respectively.
- ⁷Demonstration 23 on the compact disc, "Virtual pitch with random harmonics," actually makes this point successfully.
- Burns, E. M. (1982). "Pure-tone pitch anomalies I. Pitch-intensity effects and diplacusis in normal ears," *J. Acoust. Soc. Am.* **72**, 1394–1402.
- Carterette, E. C., and Kendall, R. (1990). "Auditory demonstrations—A review," *Music Percept.* **7**, 435–438.
- Chivers, R. C. (1989). "Auditory demonstrations on compact disc," *Acustica* **69**, 192.
- Dobbins, P. A., and Cuddy, L. L. (1982). "Octave discrimination: An experimental confirmation of the 'stretched' subjective octave," *J. Acoust. Soc. Am.* **72**, 411–415.
- Eargle, J. (1989). "Auditory demonstrations and anechoic orchestral music recording," *Audio* **73**, 98–103 (July).
- Fastl, H. (1976/77). "Temporal masking effects II: Critical-band noise masker," *Acustica* **36**, 317–331.
- Fletcher, H., and Munson, W. A. (1933). "Loudness, its definition, measurement and calculation," *J. Acoust. Soc. Am.* **5**, 82–108.
- Hartmann, W. M. (1988). "Pitch perception and the segregation and integration of auditory entities," in *Auditory Function*, edited by G. M. Edelman, W. E. Gall, and W. M. Cowan (Wiley, New York), pp. 623–645.
- Hartmann, W. M. (1989). "Auditory demonstrations on compact disc—A review," *J. Acoust. Soc. Am.* **85**, 521–522.
- Hartmann, W. M. (1990). "Psychoacoustical demonstrations—Potential and peril," *J. Acoust. Soc. Am. Suppl.* **1** **87**, S31.
- Houtgast, T. (1972). "Psychophysical evidence for lateral inhibition in hearing," *J. Acoust. Soc. Am.* **51**, 1885–1894.
- Houtsma, A. J. M. (1981). "Noise-induced shifts in the pitch of pure and complex tones," *J. Acoust. Soc. Am.* **70**, 1661–1668.
- Houtsma, A. J. M., Rossing, T. D., and Wagenaars, W. M. (1987). "Auditory Demonstrations" compact disc, Acoustical Society of America, Eindhoven, Netherlands.
- Houtsma, A. J. M., and Bilsen, F. A. (1992). Personal communication.
- Houtsma, A. J. M., and Fleuren, J. F. M. (1991). "Analytic and synthetic pitch of two-tone complexes," *J. Acoust. Soc. Am.* **90**, 1674–1676.
- Ohgushi, K. (1983). "The origin of tonality and a possible explanation of the octave enlargement phenomenon," *J. Acoust. Soc. Am.* **73**, 1694–1700.
- Scharf, B. (1970). "Critical bands," in *Foundations of Modern Auditory Theory*, edited by J. Tobias (Academic, New York), Vol. 1.
- Smooenburg, G. F. (1970). "Pitch perception of two-frequency stimuli," *J. Acoust. Soc. Am.* **48**, 924–942.
- Stevens, S. S. (1935). "The relation of pitch to intensity," *J. Acoust. Soc. Am.* **6**, 150–154.
- Stevens, S. S. (1971). "Issues in psychophysical measurement," *Psychol. Rev.* **78**, 426–450.
- Stevens, S. S., and Greenbaum, H. B. (1966). "Regression effect in psychophysical judgement," *Percept. Psychophys.* **1**, 439–446.
- Strong, W. J. (1989). "Review: Auditory demonstrations," *J. Catgut Acoust. Soc.* **1**, 67.
- Sundberg, J. E. F., and Lindqvist, J. (1973). "Musical octaves and pitch," *J. Acoust. Soc. Am.* **54**, 922–929.
- Terhardt, E., and Fastl, H. (1971). "Zum Einfluss von Stortönen und Storgerauschen auf die Tonhöhe von Sinustönen," *Acustica* **25**, 53–61.
- Terhardt, E. (1972a). "Zur Tonhöhenwahrnehmung von Klängen I. Psychoakustische Grundlagen," *Acustica* **26**, 173–186.
- Terhardt, E. (1972b). "Zur Tonhöhenwahrnehmung von Klängen II. Ein Funktionsschema," *Acustica* **26**, 187–199.
- Terhardt, E. (1974). "Pitch, consonance, and harmony," *J. Acoust. Soc. Am.* **55**, 1061–1069.
- Verschuure, J., and van Meeteren, A. A. (1975). "The effect of intensity on pitch," *Acustica* **32**, 33–44.
- Walliser, K. (1969). "Zusammenhänge zwischen dem Schallreiz und der Periodenthöhe," *Acustica* **21**, 319.
- Ward, W. D. (1954). "Subjective musical pitch," *J. Acoust. Soc. Am.* **26**, 369–380.
- Warren, R. M. (1984). "Perceptual restoration of obliterated sounds," *Psychol. Bul.* **96**, 371–383.
- Warren, R. M., Obusek, C. J., and Ackroff, J. M. (1972). "Auditory induction: Illusory continuity of the fainter of two sounds," *Science* **176**, 1149–1151.
- Watkins, A. (1992). Private communication.
- Wier, C. C., Jesteadt, W., and Green, D. M. (1977). "Frequency discrimination as a function of frequency and sensation level," *J. Acoust. Soc. Am.* **61**, 178–184.
- Zwicker, E. (1961). "Subdivision of the audible frequency range into critical bands (Frequenzgruppen)" *J. Acoust. Soc. Am.* **33**, 248.
- Zwicker, E., Flottorp, G., and Stevens, S. S. (1957). "Critical bands in loudness summation," *J. Acoust. Soc. Am.* **29**, 548–557.