Level discrimination of sinusoids as a function of duration and level for fixed-level, roving-level, and across-frequency conditions

Andrew J. Oxenham[a)]
Institute for Hearing, Speech, and Language, and Department of Speech-Language Pathology and Audiology (133FR), Northeastern University, Boston, Massachusetts 02115

Søren Buus
Institute for Hearing, Speech, and Language, and Communication and Digital Signal Processing Center, Department of Electrical and Computer Engineering (442DA), Northeastern University, Boston, Massachusetts 02115

(Received 7 May 1999; revised 17 August 1999; accepted 28 October 1999)

The ability of listeners to detect level differences between two sinusoidal stimuli in a two-interval forced-choice procedure was measured as a function of duration and level in three conditions: (1) the pedestal was fixed in level and the stimuli in the two intervals had the same frequency of either 1 or 2 kHz (fixed-level condition); (2) the pedestal was roved in level over a 20-dB range from trial to trial, but the stimuli still had the same frequency of either 1 or 2 kHz (roving-level condition); and (3) the pedestal was roved in level over a 20-dB range and the two stimuli differed in frequency, such that one was around 1 kHz while the other was around 2 kHz (across-frequency condition). In the fixed-level conditions, difference limens decreased (improved) with both increasing duration and level, as found in previous studies. In the roving-level conditions, difference limens increased and the dependence on duration and level decreased. Difference limens in the across-frequency conditions were generally highest and showed very little dependence on either stimulus duration or level. The results may be understood in terms of different internal noise components with additive variances: In the fixed-level conditions, sensation noise, which is dependent on stimulus attributes such as duration and level, is dominant. In more difficult conditions, where trace-memory and/or across-channel comparisons are required, a more central, stimulus-independent noise dominates. [S0001-4966(00)02502-9]

PACS numbers: 43.66.Fe, 43.66.Ba [RVS]

INTRODUCTION

In a recent study, Buus *et al.* (1997) made a distinction between ''loudness discrimination'' and ''level discrimination.'' Level discrimination (also termed intensity discrimination) was used to describe measurements of difference limens for stimuli that differ only in level, while loudness discrimination was used to describe measurements where the listener is required to judge the loudness of stimuli differing in more than one dimension (e.g., in level and duration, or in level and frequency). It was suggested that the difference between the two measures may reflect different underlying processes. For two otherwise identical sounds, level discrimination may be based on an optimal combination of independent information from different frequency channels (Florentine and Buus, 1981). For two unlike sounds, loudness discrimination may be based on the overall sensation of loudness, rather than a combination of information from independent frequency channels.

Buus *et al.* proposed that this distinction is not merely semantic, and that the difference may be reflected in the dependence of the difference limen (DL) on stimulus duration and stimulus level. Level discrimination is dependent on stimulus duration: generally, the just-noticeable change in level, $\Delta L_{\rm DL}$ [defined as $20\log(1+\Delta p/p)$, where p is the sound pressure of the standard and Δp is the just-noticeable change in pressure], decreases by about a factor of 2 for every tenfold increase in duration (Henning, 1970; Florentine, 1986). In contrast, Buus *et al.* (1997) reported that loudness difference limens, when comparing tones of different durations, seemed to be independent of duration. Also, DLs for level discrimination generally decrease monotonically with increasing stimulus level, at least for band-limited stimuli, giving rise to the well-known ''near-miss'' to Weber's law (McGill and Goldberg, 1968). For loudness discrimination at 5 kHz, however, DLs seemed to be maximal at medium sound levels and decrease somewhat at both low and high sound levels, as predicted by the slope of the derived loudness function (Buus *et al.*, 1997).

Most previous studies have assumed that level discrimination is based on an internal variable related to loudness (e.g., Durlach and Braida, 1969), and much effort has gone into relating the form of the loudness function to level DLs (e.g., Hellman and Hellman, 1990; Allen and Neely, 1997). The suggestion by Buus *et al.* that the form of the loudness function may underlie loudness discrimination, but not level discrimination, is contrary to the basic tenet of these studies

a)Present address: Research Laboratory of Electronics, Rm. 36-763, Massachusetts Institute of Technology, Cambridge, MA 02139. Electronic mail: oxenham@mit.edu

and, if true, has important theoretical consequences.

While the distinction between loudness and level discrimination is a possible interpretation of the data of Buus *et al.* (1997), other interpretations exist. For instance, Durlach, Braida, and colleagues (e.g., Durlach and Braida, 1969; Lim *et al.*, 1977; Braida *et al.*, 1984) have proposed that intensity discrimination is limited by a number of different internal noise sources, including sensation noise and various other noise sources related to memory processes or comparisons of unlike stimuli. Sensation noise is assumed to reflect the fundamental coding inaccuracy at an early stage in the auditory system, and should not be affected by changes in experimental paradigm. In fact, in the models of Durlach and colleagues, sensation noise is assumed to be constant in all conditions. It may be, however, that sensation noise could be more accurately modeled by allowing it to be dependent on certain stimulus parameters, such as stimulus duration (representing the increasingly accurate neural representation of level with increasing duration) and level (by assuming that the near-miss to Weber's law represents a fundamental improvement in neural coding at high levels through, for example, the recruitment of off-frequency auditory-nerve fibers). Memory and other more ''central'' noise, on the other hand, should not depend on the stimulus itself, but would be expected to depend on experimental parameters, such as the interstimulus interval or whether the stimuli are roved in level across trials. Which of these two noise categories dominates will depend on the exact conditions tested. Thus, it is at least conceivable that different paradigms will reveal different dependencies on parameters such as stimulus duration, without it being necessary to postulate different underlying mechanisms.

The purpose of this study was to investigate further the differences between level discrimination and loudness discrimination by measuring DLs as a function of duration and level for both paradigms. Three conditions were tested. The first condition was a simple level-or intensity-discrimination task (Jesteadt *et al.*, 1977; Florentine, 1986) in which the listener was instructed to select the more intense of two otherwise identical tones, and the pedestal level was kept constant within a run.

The second condition was again level discrimination, but with the overall level of the stimuli roved over a range of 20 dB from trial to trial. While the effects of stimulus duration and level have been previously investigated for fixed-level discrimination, very few studies have employed a roving-level paradigm to examine these parameters. Berliner *et al.* (1977) measured performance in fixed-level and roving-level conditions for durations between 200 and 1250 ms and found no noticeable difference in the dependence of sensitivity on duration. However, it seems that the dependence of sensitivity on duration may be greater at durations below about 200 ms (Henning, 1970; Florentine, 1986), and no data comparing fixed-level with roving-level discrimination seem to exist for these durations. A number of studies have found improved performance at the ends of a given level range (both high and low) in roving-level conditions (Berliner and Durlach, 1973; Berliner *et al.*, 1977), and this has been ascribed to a ''perceptual anchor'' effect (Braida *et al.*, 1984): performance at the extreme intensities of a given range seems to be governed primarily by context coding (i.e., a long-term memory representation of the maximum and minimum levels of a stimulus range), while mid-range performance may be governed primarily by a trace mode, which is related to a rapidly degraded short-term memory representation of a stimulus. So far, however, no studies have examined the effect of level, independently of stimulus range, for roving-level conditions.

The third condition was designed to require loudness, rather than level, discrimination, as defined by Buus *et al.* (1997). In each trial two tones, one at 1 kHz and the other at 2 kHz, were presented in random order and listeners were asked to select the louder one. Our use of two frequencies separated by an octave should preclude the use of within-channel level cues and should force listeners to use a cue related to overall loudness. In this condition, the overall level across trials was also roved to rule out the possibility that listeners based their judgments solely on the within-trial level of one of the two frequencies, while ignoring the other (Lim *et al.*, 1977). Based on the findings of Buus *et al.* for tones of different durations, one might expect DLs for tones of different frequencies also to be independent of stimulus duration. Also, according to the theory of Buus *et al.*, the DL for loudness discrimination is inversely proportional to the slope of the loudness function. Thus, according to the data of Florentine *et al.* (1996), the DL for loudness discrimination should be maximal at medium sound levels and should decrease at low and high levels, in contrast to what is generally found for level discrimination.

In summary, this study is designed to test the hypothesis that loudness discrimination and level discrimination reflect different underlying mechanisms. If the hypothesis is correct, DLs for across-frequency comparisons should be independent of duration and should depend on level in a way that is different from the level dependence observed in level discrimination experiments. If, on the other hand, both tasks reflect the same underlying mechanism, it may be possible to account for all the results within a framework of two noise classes, as described above. Specifically, changes in performance in roving-level and across-frequency conditions may be accountable in terms of additional, stimulus-independent noise, representing the additional load of trace memory and/or across frequency comparisons.

I. EXPERIMENT 1: LEVEL AND LOUDNESS DISCRIMINATION AS A FUNCTION OF STIMULUS DURATION

A. Stimuli and apparatus

The stimuli were tone bursts at either 1 or 2 kHz, gated on and off with 5-ms raised-cosine ramps. Difference limens were measured for stimulus durations of 5, 15, 50, 150, and 500 ms, measured at the half-amplitude points of the envelope. The stimuli were generated digitally at a sampling rate of 50 kHz and played out via a TDT digital-to-analog converter (DD1). The stimuli were low-pass filtered at 20 kHz (TDT FT5) and passed through a programmable attenuator

(TDT PA4) before being presented via a headphone amplifier (TDT HB6) to one earpiece of a Sony MDR-V6 headset.

B. Procedure

Three different conditions were tested. These are referred to as the fixed-level, roving-level, and across-frequency conditions, and are described in detail below.

1. Fixed-level and roving-level conditions

In the fixed-level condition, ΔL_{DL} was measured using a fixed pedestal level of 65 dB SPL. In the roving-level condition, ΔL_{DL} was measured with the pedestal level roved across trials over a range from 55 to 75 dB SPL, uniformly distributed in steps of 1 dB. For these two conditions, ΔL_{DL} was measured at frequencies of 1 and 2 kHz using a 2IFC method with a three-down one-up adaptive interleaved tracking procedure. Within each run, a given trial was selected at random from four independent tracks. Each trial consisted of two intervals, marked by lights on the response box, separated by a silent interstimulus interval (ISI) of 700 ms. Both intervals contained the pedestal; in one interval, chosen at random, a second sinusoid (the signal) with the same duration, frequency, and phase as the pedestal was added to the pedestal to produce a level increment. The listener's task was to choose the interval with the increment, and correct-answer feedback was provided after each trial. Initially, the signal level in each track was 5 dB above that of the pedestal, resulting in an initial ΔL of 8.9 dB. For the first two reversals in each track, the signal level was varied in steps of 5 dB. Thereafter, the step size was reduced to 2 dB for the final four reversals. The threshold for each track was defined as the mean of the last four reversals, and the overall threshold for the run was defined as the mean threshold level across the four tracks. Once the threshold for the run had been determined, it was transformed into units of ΔL, using the equation $\Delta L = 20\log[(p_1+p_0)/p_0]$, where p_1 and p_0 are the sound pressures of the signal and the pedestal, respectively. Three such estimates of the threshold value of $\Delta L(\Delta L_{\mathrm{DL}})$ were obtained on different days for each duration and frequency combination. The final threshold estimate was the geometric mean of the three estimates of ΔL_{DL}.

2. Across-frequency condition

The across-frequency condition, was designed to prevent listeners from using within-channel cues in performing loudness discrimination. In each trial, one interval contained a tone with a nominal frequency of 1 kHz, while the other interval contained a tone with a nominal frequency of 2 kHz. Although the excitation patterns of these two tones would not be expected to overlap greatly, it is at least conceivable that one auditory filter, centered at the point where the excitation patterns cross, would carry reliable information as to the relative levels of the two tones. To reduce the reliability of such a cue, the frequencies of the tones were roved from trial to trial by $\pm 5\%$. Except for the frequency rove and the different frequencies across intervals, the stimuli were generated and presented in the same way as in the second condition. Which of the two nominal frequencies was presented first was randomized from trial to trial. As in the second condition, a 20-dB level rove was imposed on the stimuli, with the 1-kHz pedestal level ranging from 55 to 75 dB SPL. The pedestal level of the 2-kHz tone was set so as to approximate the loudness of the 1-kHz pedestal for each listener individually, as described below. In two of the four tracks, the standard interval was the 1-kHz tone; in the other two tracks, the standard interval was the 2-kHz tone. The interleaving ensured that listeners could not tell which was the standard based solely on the frequency. Again, the averages of the third to the sixth reversals in each track were used to estimate the DL. In contrast to the other conditions, a track did not terminate after six reversals, but its probability of presentation was reduced. If a track with six or more reversals was selected, the random selection procedure was repeated. If the same occurred on the second selection, a third random selection was made. The third selection was then used without regard to the number of reversals already made on the selected track. Thus, only after three selections would a track with six or more reversals be presented. The reason for this procedure was to avoid the possibility that all but one track would terminate, leaving a single track for which it would be possible (using the feedback provided) to distinguish between the standard and the signal intervals based simply on frequency.

The signal levels at threshold for the two tracks with the 1-kHz tone as pedestal were averaged and converted to units of ΔL, and the same was done for the two tracks with the 2-kHz tone as pedestal. Then, the two estimates of ΔL_{DL} were averaged to provide an unbiased estimate of ΔL_{DL}, as described in the Appendix. The preliminary experiment, described below, provided an estimate of equal loudness levels for the 1- and 2-kHz tones for each listener. However, the accuracy of the estimated DL does not depend critically on the accuracy of the equal-loudness balances. As shown in the Appendix, by measuring DLs for both the 1 and 2-kHz tones as pedestals, any bias towards selecting one or other frequency can be eliminated from the estimate of the overall sensitivity.

Prior to collecting data for the across-frequency condition, a preliminary experiment was run to determine for each listener separately the level difference necessary to make the 1- and 2-kHz tones sound equally loud within the level range of 55 to 75 dB SPL. Signal durations of 5, 50, and 500 ms were tested. In this experiment, six independent tracks were interleaved within each run. In three of the tracks, the level of the 1-kHz tone was fixed and the level of the 2-kHz tone was varied; in the other three tracks, the level of the 2-kHz tone was fixed and the 1-kHz tone was varied. The fixed level of the three tracks at each frequency was set to 55, 65, and 75 dB SPL. For each trial, one of the tracks was chosen at random, as was the order of presentation (fixed or varied tone first). Thus, listeners had no way of distinguishing the fixed from the varied tones. Listeners were asked to select which of the two intervals contained the louder tone. No feedback was provided, as the correct answer depended on the listener's perception of loudness. The starting level of the varied tone in each track was set randomly over a range of ± 10 dB relative to the fixed tone. An adaptive 2IFC procedure with a one-down one-up rule was used in each track to determine the point of subjective equality. The level of the varied tone was initially altered in steps of 4 dB. After the first five reversals, the step size was reduced to 2 dB for the final four reversals. Threshold was defined as the mean of the last four reversals in each track. Each condition was repeated three times, and the mean of all conditions, pooled across durations and levels for each subject, was taken as the level difference required to produce equal loudness for 1- and 2-kHz tones.

The results from this preliminary experiment were used in the across-frequency condition to adjust the relative levels of the 1- and 2-kHz tones for each listener individually. The 1-kHz pedestal was set to the nominal level in each trial, as determined by the roving procedure, and the 2-kHz tone was adjusted by the difference found in the preliminary experiment. The level difference between 1- and 2-kHz tones judged to be equally loud ($L_{2k}-L_{1k}$) ranged from -4.7 dB for S7 to $+5$ dB for S2. The average across listeners was 0.1 dB. A repeated-measures analysis of covariance with pedestal level and log(pedestal duration) as continuous variables showed a main effect of subject ($F_{6,350}=265$; $p<0.05$) and a significant interaction between subject and pedestal duration ($F_{6,350}=2.61$; $p<0.05$) and subject and pedestal level ($F_{6,350}=2.39$; $p<0.05$), but no main effects of duration or level. This shows that duration and level could have significant effects on loudness balancing, but that the effects were not consistent across listeners. These relatively small effects were ignored and only the mean difference for each subject was used.

C. Subjects

Seven female listeners participated as subjects. All had absolute thresholds of 15 dB HL or less at octave frequencies between 250 and 8000 Hz, and none reported any history of hearing disorders or difficulties. The ages of the listeners ranged from 19 to 44 years (mean age 25.4 years). All were college students who were paid for their participation. Listeners received 1–2 h training in each condition before data were collected.

Data were collected individually in 2-h sessions. For subjects 1 and 2, the conditions were tested in an interleaved manner, such that all conditions were completed at approximately the same time. For the remaining five listeners, the fixed-level condition was completed first, followed by the roving-level condition, and finally the across-frequency condition. The reasoning behind this ordering was that listeners, having been exposed to roving-level and across-frequency conditions, may use the same, nonoptimal, strategy in discriminating stimuli in the fixed-level condition. Richards (1992) found that prior exposure to roving-level conditions impaired listeners' performance in a tone-in-noise detection task, and we hypothesized that the same might apply to our conditions.

D. Results

The results from the individual listeners are plotted in Fig. 1. Filled symbols represent the fixed-level conditions and open symbols represent the roving-level conditions. Squares and circles represent measurements using a 1- and 2-kHz tone, respectively. The circles-in-squares represent the across-frequency conditions. The results are reported in units of ΔL, plotted on a logarithmic axis. This is consistent with the method used by Florentine (1986) and also has some theoretical justification. In studies of a wide range of level discrimination conditions in both normal-hearing and hearing-impaired listeners, Buus and Florentine (1991) and Buus *et al.* (1995) found that d' can be reasonably approximated as being proportional to ΔL, and that this relationship holds over a wider range for ΔL than for either ΔI or Δp. As changes in performance should be stated as ratios, and not differences, of ΔL or d', it is appropriate to plot ΔL on a logarithmic axis. For a more detailed discussion, see Buus and Florentine (1991) and Buus *et al.* (1995).

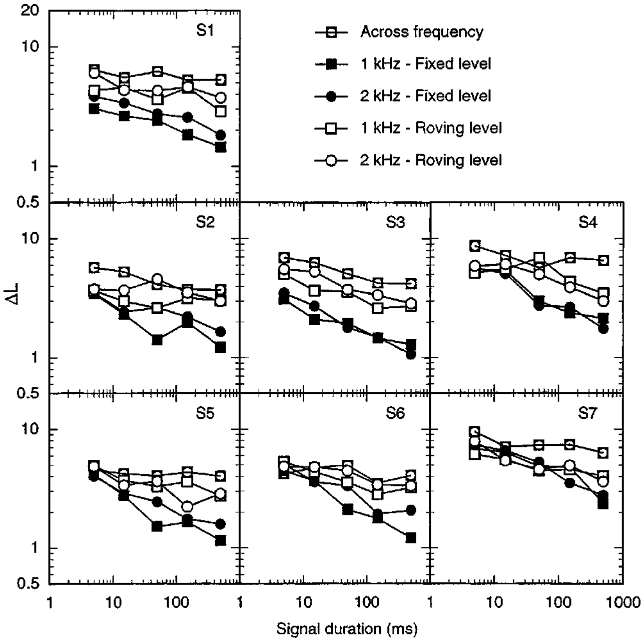

FIG. 1. Individual data from experiment 1. Difference limens in ΔL are shown as a function of pedestal duration for the fixed-level (filled symbols), roving-level (open symbols), and across-frequency (circles in squares) conditions. Squares and circles represent pedestal frequencies of 1 and 2 kHz, respectively.

Generally, the results are rather similar across the seven subjects. In particular, the results from subjects 1 and 2 do not seem very different from those of the other five subjects, suggesting that the order of presentation of the conditions did not have a strong effect on the results. The trends in the individual data are also reflected in the geometric mean data, shown in Fig. 2. Error bars represent $\pm$ one standard error across listeners. Regression slopes [log(ΔL) against log(duration)] for the mean data are shown in Table I.

Consider first the data from the fixed-level conditions (filled symbols). Thresholds decrease with increasing duration over the range of durations tested. Listeners seem slightly, but consistently, more sensitive at 1 kHz than at 2 kHz. The regression slopes for the two frequencies of -0.22 and -0.21, respectively, are shallower than the slope of -0.38 found by Florentine (1986) at 1 kHz for a 65-dB SPL pedestal. However, the slope estimates in that study varied in a relatively nonsystematic way with pedestal frequency and

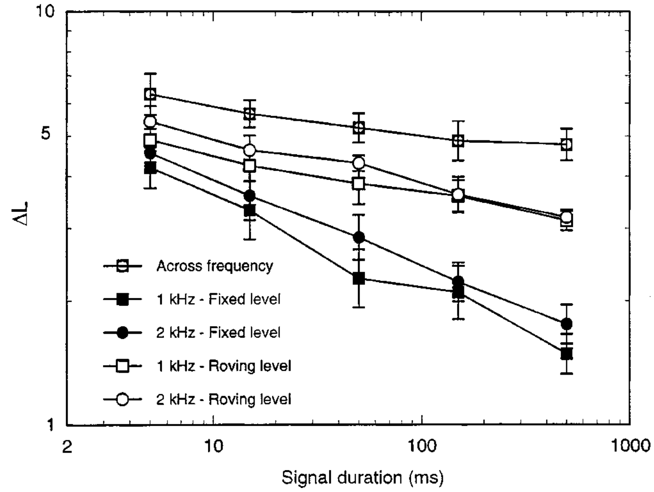

FIG. 2. Mean data from experiment 1. The axes and symbols are the same as in Fig. 1. Error bars represent ±1 standard error of the mean across the seven listeners.

level, with estimates ranging from -0.2 to -0.38. Thus, the slope estimates for the fixed-level conditions found in this study are at least within the range of those found previously.

In comparing overall thresholds across different studies, care must be taken to equate levels of performance. In most previous studies, an adaptive procedure tracking the 70.7% correct point on the psychometric function has been used, whereas the present experiment tracked the 79.4% correct point. Assuming unbiased responses, these two scores correspond to d''s of 0.78 and 1.16, respectively. As ΔL is approximately proportional to d' (Buus *et al.*, 1995), equal performance would be reflected by a ΔL_{DL} 1.5 times greater for the 79.4% conditions than for the 70.7% conditions. At a pedestal duration of 500 ms, the mean value of ΔL_{DL} at 1 kHz found here is 1.49 dB. This is very comparable with the ΔL's of 1.42 and 1.06 dB found at 1 kHz for levels of 60 and 70 dB SPL, respectively, by Florentine *et al.* (1987). If the different levels of performance are taken into account, by dividing our ΔL_{DL} by 1.5, our listeners performed somewhat better on average. In contrast, using the same pedestal level of 65 dB SPL, Florentine (1986) found a mean ΔL_{DL} at 1 kHz of 0.4 dB (equivalent to about 0.6 dB at 79.4% correct). Jesteadt *et al.* (1977) found a ΔL_{DL} of about 1 dB for a 1-kHz pedestal at a sensation level of 40 dB, which was presumably lower than 65 dB SPL. The reasons for such discrepancies have been discussed in detail by Florentine *et al.* (1987), and seem to be primarily due to amount of practice and motivation. Despite the reduced overall sensitivity relative to some earlier studies, the dependence of ΔL_{DL} on duration for the fixed-level conditions is consistent with previous studies (Henning, 1970; Florentine, 1986).

TABLE I. Slopes of $\log(\Delta L_{\mathrm{DL}})$ as a function of log(duration) for the mean data in experiment 1. All slopes are significantly greater than zero ($p<0.01$).

Frequency (kHz)	Fixed level	Roving level	Across frequency
1	−0.22	−0.092	−0.062
2	−0.21	−0.11	

Consider next the roving-level conditions (open symbols). As expected (Berliner and Durlach, 1973; Berliner *et al.*, 1977), DLs are higher than in the fixed-level conditions. Interestingly, however, the deterioration in performance is much less marked at short durations than at long durations. This pattern of results is observed in all listeners and is reflected in shallower slopes for the function relating $\log(\Delta L)$ to $\log(\text{duration})$, as can be seen in Table I.

Finally, performance in the across-frequency condition (circles in squares) shows even less dependence on duration than in the roving-level condition. Nevertheless, some improvement with increasing duration is evident in nearly all listeners, and the mean ΔL_{DL} decreases from 6.3 to 4.8 dB as duration is increased from 5 to 500 ms. As indicated in Table I, the slope for the mean data of -0.062 is still significantly different from zero ($p<0.01$).

Mean discrimination thresholds in the across-frequency condition never exceed 7 dB, even for the shortest durations. The value of 7 dB is the best an optimal detector could achieve if the judgment was based simply on a single-frequency, across-trial analysis, without comparing the two frequencies within one trial (Green, 1988, p. 19). Thus, it appears that listeners were in fact comparing levels across frequency, as desired.

E. Discussion

The finding that the across-frequency condition results in a greatly reduced dependence on duration relative to the fixed-level conditions is consistent with Buus *et al.*'s (1997) distinction between loudness discrimination and level discrimination, as described in the introduction. However, this distinction does not predict the observed reduction in duration dependence due to level roving. Despite the level roving, it should still be possible to compare excitation patterns across the two intervals, and so duration dependence should not be affected. A possible explanation for the results from all three conditions was outlined in the Introduction, and assumes the presence of at least two independent internal noise sources (Durlach and Braida, 1969). The first noise source is associated with the sensory coding of the stimulus, and has been referred to as ''sensation noise.'' In the various versions of this model to date, this noise has been assumed to be constant. However, it seems reasonable to assume that it would depend on stimulus duration—a parameter not yet accounted for in this framework. The second noise source is independent of stimulus characteristics and instead depends on experimental conditions, such as whether level roving is used (Berliner and Durlach, 1973; Berliner *et al.*, 1977) or across-frequency comparisons are made (Lim *et al.*, 1977). This noise may consist of several components, including various short- and long-term memory processes, and inaccuracies in comparing level across frequency. For our purposes, it is sufficient to group all these sources together in one stimulus-independent component, termed ''central noise.''

Thus, we may restrict ourselves to defining two internal noise sources, with zero means and variances of $\sigma_s^2(T)$ and σ_c^2, where $\sigma_s^2(T)$ is the variance associated with sensation noise, σ_c^2 is the variance associated with central noise, and T

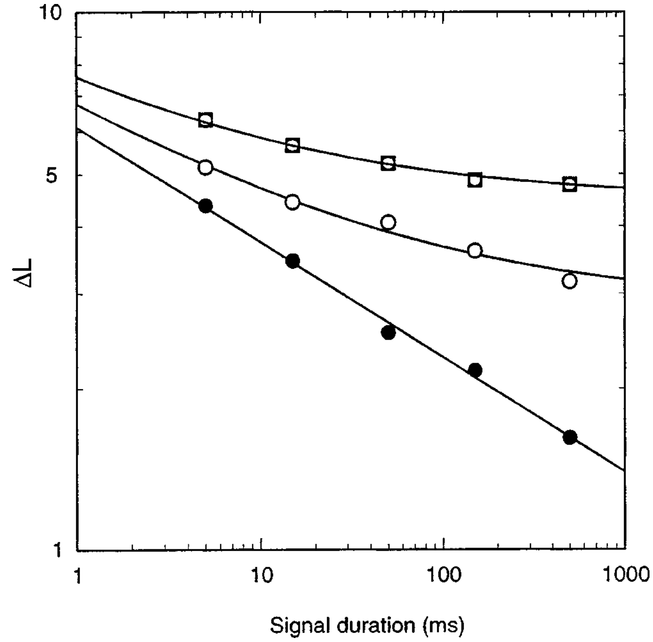

FIG. 3. Mean data from experiment 1, averaged across the two pedestal frequencies. The filled and open circles represent the fixed-level and roving-level conditions, respectively. The circles in squares represent the across-frequency condition. The curves are fits using the model described in the text.

is the stimulus duration in ms. If ΔL is proportional to d' (Buus *et al.*, 1995), it follows that, for a given d', ΔL at threshold (ΔL_{DL}) is proportional to σ, the total internal noise standard deviation. This is because $d' = k\Delta L/\sigma$, where k is a constant. The total internal noise variance for the fixed-level condition can be derived by letting $k=1/d'$. Thus $\sigma^2 = \Delta L_{DL}^2$. This overall σ^2 can be thought of as the sum of $\sigma_s^2(T)$ and σ_c^2. In the roving-level and across-frequency conditions, we assume that the value of $\sigma_s^2(T)$ remains the same as in the fixed-level condition, while the value of σ_c^2 increases, reflecting the increased difficulty of the task.

The question of whether simply increasing the value of σ_c^2 can account both for the increase in thresholds and the decreasing dependence on duration can be answered by finding a value of σ_c^2 that best fits the data, while keeping the value of $\sigma_s^2(T)$ fixed for a given duration across conditions. This was done by first fitting the data from the fixed-level conditions. The value for σ^2 as a function of pedestal duration was estimated by fitting a curve to the mean values of ΔL_{DL}^2. The data from the roving-level and across-frequency conditions were then predicted by increasing the value of σ_c^2 (which is independent of T) to best match the data.

The predictions using this model are shown as curves in Fig. 3. The data are taken from Fig. 2, averaged across the two frequencies in the fixed-level and roving-level conditions. The predicted curve for the fixed-level condition is a simple power function that described the data well, namely, $\Delta L_{DL}^2 = e^{3.62}T^{0.427}$. As the parameters for the curve were fitted to the data, it is not surprising that the fit is good. The other two curves are more interesting. They represent the effect of increasing the duration-independent component of the noise variance. It can be seen that by increasing the duration-independent central noise variance, it is possible to account well for both the increase in thresholds and the decrease in duration dependence for both the roving-level and across-frequency conditions. The constants added to the values of ΔL_{DL}^2 in order to fit the data from the roving-level and across-frequency conditions were 8.2 and 20.1, respectively.

The data and the results from the modeling suggest that it may not be necessary to invoke two different mechanisms to account for level discrimination and loudness discrimination. Instead, these two tasks may form part of a continuum: in some conditions, with low uncertainty and high similarity of stimuli, sensation noise may dominate, producing a strong dependence on stimulus duration. In other conditions, central noise may dominate, in which case the duration dependence is much reduced.

In this experiment, both roving-level and across-frequency conditions resulted in a reduction in the dependence of ΔL_{DL} on duration. The pattern of results was well described by assuming that performance is limited by two internal noise sources, termed sensation noise and coding noise (Durlach and Braida, 1969). If sensation noise depends on the stimulus characteristics, such as duration, then it may also be responsible for changes in level discrimination with pedestal level. As suggested by McGill and Goldberg (1968), the ''near-miss'' to Weber's law, whereby performance improves with increasing level, may be explained in terms of improved coding accuracy, or a relative decrease in sensation noise, with increasing level. Thus, measuring level discrimination as a function of overall level provides another opportunity to attempt to separate the effects of sensation noise and coding noise. Also, as described below, different predictions follow from the different theories under consideration.

II. EXPERIMENT 2: LEVEL AND LOUDNESS DISCRIMINATION AS A FUNCTION OF STIMULUS LEVEL

In the second experiment we examined the dependence of ΔL_{DL} on stimulus level in the same three conditions tested in experiment 1. Using tones with the same frequency (5 kHz) but different durations, Buus *et al.* (1997) found that difference limens were generally highest at medium levels and lower at low and high levels. They related this finding to the shallower loudness function they found at medium sound levels and proposed that, unlike level discrimination (where performance generally improves monotonically with level), loudness discrimination was dependent on the slope of the loudness function. This predicts that ΔL_{DL} should follow the near-miss to Weber's law in the fixed-level and roving-level conditions, but be a nonmonotonic function in the across-frequency condition with ΔL_{DL} being roughly inversely proportional to the slope of the loudness function as found, for instance, by Florentine *et al.* (1996).

In the framework of Durlach, Braida, and colleagues, the near-miss to Weber's law is accounted for by a steepening of the function relating level to internal representation (Braida *et al.*, 1984), while the sensation noise remains constant. This would predict that, while overall performance may depend on various experimental manipulations, the dependence of ΔL on pedestal level should remain unchanged. Thus, according to this model, the three conditions should produce parallel functions when $\log(\Delta L_{DL})$ is plotted as a function of pedestal level.

A third hypothesis can be formulated as follows: The near-miss to Weber's law may be due to a relative decrease in sensation noise with increasing level (McGill and Goldberg, 1968), consistent with the idea that the near-miss is due to additional information being provided by off-frequency channels (Viemeister, 1972; Moore and Raab, 1974; Viemeister, 1974; Florentine and Buus, 1981). In this case, roving the level, or forcing listeners to compare stimuli with different frequency would be equivalent to adding stimulus-independent noise. As in experiment 1, this would predict that the slope of the function relating $\log(\Delta L_{DL})$ to stimulus level should become increasingly shallow, as the stimulus-independent noise begins to dominate.

These three different predictions were tested here using the three conditions from experiment 1, with the pedestal duration fixed at 500 ms and the mean pedestal level varied parametrically.

A. Stimuli and procedure

The method of generating the stimuli was the same as that in experiment 1. As in experiment 1, three conditions were tested: fixed-level and roving-level level discrimination, and across-frequency loudness discrimination. For the first two conditions, only 1-kHz pedestals were used, as experiment 1 showed no great differences between pedestals of 1 and 2 kHz. For the third condition, 1- and 2-kHz tones were compared. The stimuli had a half-amplitude duration of 500 ms and were gated using 5-ms raised-cosine ramps. The fixed-level condition was tested at pedestal levels of 25, 50, and 80 dB SPL (except for S4 who was tested at 30, 50, and 80 dB SPL). The roving-level condition was tested with pedestal levels in the ranges of 15–35 dB (20–40 dB for S4), 40–60 dB, and 70–90 dB SPL. The rove range was thus always 20 dB, but the median level varied from low to high. This is different from most other roving-level experiments, in which the rove range and the overall level range generally covary. The across-frequency condition was tested with the same 1-kHz pedestal levels as in the roving-level condition. The level of the 2-kHz pedestal was adjusted for each listener and level range to provide roughly equal loudness across the two frequencies, as described below. In an attempt to improve the overall performance of listeners, and to further reduce the influence of coding noise in the fixed-level condition, the ISI was reduced to 250 ms for all three conditions.

The same procedures were used as in experiment 1. Again, interleaved tracks, as described in experiment 1, were used to estimate thresholds, and the mean of three runs was defined as the final threshold.

B. Subjects

Three subjects from experiment 1, S3, S4, and S5, participated in this experiment. Their ages were 23, 44, and 21 years, respectively. This experiment was carried out after experiment 1, and so listeners already had considerable experience in level-discrimination tasks. Listeners were given a further 2 h of practice on the new conditions before data were collected.

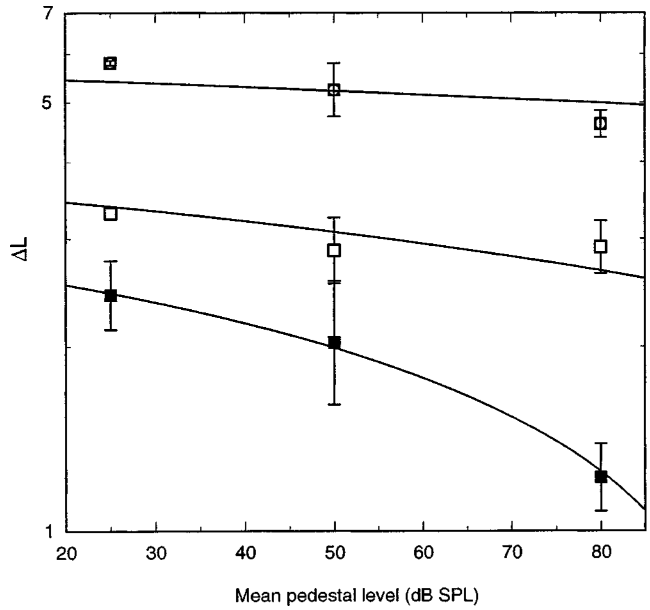

FIG. 4. Mean data from experiment 2. Difference limens in ΔL are plotted as a function of mean pedestal level. The filled and open squares represent the fixed-level and roving-level conditions, respectively. The circles in squares represent the across-frequency condition. Error bars represent ± 1 standard error of the mean across the three listeners. The curves are fits using the model described in the text.

C. Preliminary experiments

Before any data were collected, absolute thresholds at 1 and 2 kHz were measured for the three listeners using a two-interval two-alternative forced choice adaptive procedure with a three-down one-up tracking rule that tracks the 79.4% correct point on the psychometric function. In the actual experiment, the lowest level range was selected such that the lowest level presented was at least 5 dB above absolute threshold at both 1 and 2 kHz. For S3 and S5, this was 15–35 dB. For S4, the levels had to be increased by 5 dB to meet this criterion.

As in experiment 1, a preliminary experiment was conducted to determine points of subjective equality for the loudness of the 1- and 2-kHz tones for the three different level ranges and the three listeners individually. This was done as in experiment 1, except that only the 500-ms durations were tested. The three level ranges were tested separately, using fixed levels of 15, 25, and 35 (20, 30, and 40 for S4), 40, 50, and 60 dB, and 70, 80, and 90 dB SPL. The mean differences for each of the level ranges and each of the listeners were used in the across-frequency condition to set the level of the 2-kHz pedestal.

D. Results and discussion

The data from the three listeners were rather similar and therefore only the mean results are shown in Fig. 4. Error bars represent ± 1 s.e. of the mean. The data at the lowest level are shown as having a mean pedestal level of 25 dB

SPL, although the levels for S4 were in fact 5 dB higher. The fixed-level condition (solid squares) shows a typical decrease in ΔL with increasing pedestal level, often referred to as the ''near-miss'' to Weber's law. Previous studies of the ''near-miss'' have generally quoted slopes in terms of $10\log(\Delta I)$, or intensity increment level, as a function of $10\log(I)$, or pedestal level (McGill and Goldberg, 1968; Viemeister, 1972). In these units, the slope of the mean data from the fixed-level condition is 0.93, which is consistent with results from previous studies (Rabinowitz *et al.*, 1976).

In contrast, both the roving-level (open squares) and the across-frequency (circles in squares) conditions show much less dependence on overall level, with slopes $[10\log(\Delta I)$ as a function of $10\log(I)]$ closer to unity, of 0.99 and 0.97, respectively. The curves in the figure are model fits, described below. As in experiment 1, there seems to be no qualitative difference between the results from level discrimination (fixed and roving level) and loudness discrimination across frequency. Instead, the deterioration in performance due to both level roving and across-frequency comparisons leads to a reduced dependence on overall level.

As with the data in experiment 1, it is possible to derive predictions based on the effects of stimulus-dependent sensation noise and stimulus-independent central noise. Again, the results from the fixed-level condition were used to derive an initial estimate of internal variance. Performance in the other two conditions was predicted by assuming the addition of level-independent central noise. This was achieved as described for experiment 1. The predictions are shown as solid curves in Fig. 4. The equation used to derive the noise variance for the fixed-level condition was $\Delta L^2 = -0.0802L + 7.98$, where L is the nominal pedestal level in dB SPL. The constants added to the values of ΔL_{DL}^2 in order to fit the data from the roving-level and across-frequency conditions were 5.5 and 23.3, respectively. These values are reasonably similar to the ones found in experiment 1 of 8.2 and 20.1, considering that only a subset of the original listener group was used in experiment 2. The similarity of these values lends some support to the idea that roving-level and across-frequency conditions may deteriorate performance in a way that is stimulus independent.

The predictions capture the trend of reduced level dependence in the roving-level and across-frequency conditions. However, the actual data show a smaller-than-predicted effect of level in the roving-level condition, and a greater-than-predicted effect of level in the across-frequency condition. While the fit is not quite as convincing as in experiment 1, the trend is captured sufficiently to suggest that the data can be explained in terms of an increase in stimulus-independent central noise relative to stimulus-dependent sensation noise.

No deterioration in performance at medium levels was found in the across-frequency condition. Based on the assumption that the loudness DL is inversely proportional to the slope of the loudness function (Buus *et al.*, 1997), and the fact that the loudness function seems to be shallower at medium levels than at high or low levels at 1 kHz, at least when measured using loudness integration (Florentine *et al.*, 1996), a deterioration at medium levels would have been expected. Such a mid-level deterioration in DLs was found by Buus *et al.* (1997) at 5 kHz in a comparison of long- and short-duration tones, which was quantitatively consistent with the change in the slope of the derived loudness function. A number of parameter differences could account for the different pattern of results observed here, including the use of equal-duration but different-frequency tones, the use of feedback, and the lower frequencies tested here. It should be noted, however, that a mid-level deterioration in DLs has been observed before in traditional level discrimination tasks at frequencies above about 2 kHz, but not below (Penner *et al.*, 1974; Carlyon and Moore, 1984; Long and Cullen, 1985; Florentine *et al.*, 1987; Oxenham and Moore, 1995). Thus, the mid-level deterioration observed by Buus *et al.* at 5 kHz may in part be a reflection of an effect also observed in traditional level-discrimination tasks.

The results are also inconsistent with the predictions of the models explored by Durlach, Braida, and colleagues. Generally in these models, sensation noise is assumed to be constant, and when the near-miss to Weber's law has been accounted for, this has been achieved by a steepening of the function relating sound level to an internal sensation variable (e.g., Braida *et al.*, 1984). Thus it seems that their model would predict parallel functions in Fig. 4.

The present results may also have some implications for models that attempt to relate the slope of the loudness function to ΔL_{DL} (Hellman *et al.*, 1987; Hellman and Hellman, 1990; Allen and Neely, 1997). If ΔL_{DL} as a function of level is determined by the slope of the loudness function, changing the experimental parameters should result in a parallel shift in thresholds, as predicted by the models of Durlach and colleagues. The fact that the relation between ΔL_{DL} and level depends on experimental manipulations seems to indicate that there exists a dissociation between the loudness function and level discrimination, which is not accounted for by models that postulate a direct relation between the slope of the loudness function and level discrimination.

III. SUMMARY

In experiment 1, level discrimination was measured for 1- and 2-kHz tones as a function of stimulus duration for three conditions: (1) within-frequency comparisons with a fixed pedestal level of 65 dB SPL; (2) within-frequency comparisons with the pedestal level roved uniformly between 55 and 75 dB SPL; and (3) across-frequency comparisons with the pedestal level roved uniformly between 55 and 75 dB SPL. As found previously (Florentine, 1986), discrimination in the fixed-level condition improved with pedestal duration. Roving the pedestal level produced a greater deterioration in performance for long than for short tones, leading to less dependence of discrimination thresholds on duration. Across-frequency comparisons produced a further deterioration in performance and even less duration dependence. However, there was no qualitative difference between the results of the within-frequency and the across-frequency roved conditions.

In experiment 2, ΔL_{DL}'s for the same three conditions were measured as a function of pedestal level using 500-ms tones. In the fixed-level condition, the expected improvement with increasing pedestal level was observed. Roving the pedestal level over a 20-dB range increased ΔL_{DL}'s and reduced the dependence on overall level, producing a relationship much closer to Weber's law than that observed in the fixed-level condition. Across-frequency comparisons produced a further reduction in performance. Again, there was no qualitative difference between the results of the within-frequency and across-frequency roved conditions.

The results do not provide support for a distinction between level discrimination and loudness discrimination, whereby level discrimination is performed using independent information from multiple frequency channels and loudness discrimination is based on an overall sensation of loudness (Buus *et al.*, 1997). Instead, the results may be understood within a framework similar to that suggested by Durlach, Braida, and colleagues (Durlach and Braida, 1969; Berliner and Durlach, 1973; Berliner *et al.*, 1977; Lim *et al.*, 1977). In this approach, as adapted here, two separate internal noise sources are required. The first, termed sensation noise, represents the variance associated with the initial coding of the stimulus. This noise is assumed to be independent of experimental manipulations, such as roving level or changing the interstimulus interval, but (unlike in the formulation of Durlach and colleagues) is assumed to be dependent on stimulus parameters, such as duration and level. The second noise, termed central noise, is assumed to be the variance added by higher, perhaps more cognitive, processes, such as memory and across-frequency processing (Lim *et al.*, 1977). This noise is not dependent on stimulus parameters, but is dependent on experimental manipulations. The increase in ΔL_{DL}'s and the reduction in their dependence on duration and level in the roving-level and across-frequency conditions can be explained in terms of an increase in the stimulus-independent central noise, which represents the increased difficulty of the tasks, while leaving the initial coding of the stimuli (sensation noise) unaffected.

ACKNOWLEDGMENTS

This work was supported by NIH Grant Nos. R01 DC 00187 and R01 DC 03909. Two reviewers provided helpful comments on an earlier version of this paper.

APPENDIX: DERIVING ΔL

In the across-frequency conditions, ΔL_{DL} was derived from four interleaved tracks, in which two used a 1-kHz standard and a 2-kHz comparison tone, and two used the opposite. If the two tones are not calibrated exactly for equal loudness, the two pairs of interleaved tracks will not converge on the same ΔL. Even in this case, the ''true'' ΔL_{DL} can be calculated quite simply, as follows:

Assume two tones at different frequencies where tone 1 is fixed at level L_{1F} and tone 2 is adjusted to be 1 DL above the fixed tone 1, such that

$$L_{2V}=L_{1F}+K+\Delta L_{\mathrm{DL}}, \tag{A1}$$

where L_{2V} is the (varied) level of tone 2, L_1 is the (fixed) level of tone 1, K is the level difference necessary for equal loudness, and ΔL_{DL} is the ''true'' DL. The quantity $L_{2V}-L_{1F}$ is then the measured DL. Similarly, when tone 2 is fixed, and tone 1 is set to be 1 DL above tone 2,

$$L_{1V}=L_{2F}-K+\Delta L_{\mathrm{DL}}, \tag{A2}$$

where ΔL_{1V} is the varied level of tone 1, L_{2F} is the fixed level of tone 2, and $L_{1V}-L_{2F}$ is the measured DL. Adding A1 and A2 and rearranging the terms gives

$$2\Delta L_{\mathrm{DL}}=(L_{2V}-L_{1F})+(L_{1V}-L_{2F}). \tag{A3}$$

The terms $(L_{2V}-L_{1F})$ and $(L_{1V}-L_{2F})$ are the two measured DLs, ΔL_1 and ΔL_2, respectively, so,

$$\Delta L_{\mathrm{DL}}=(\Delta L_1+\Delta L_2)/2. \tag{A4}$$

Equation A4 shows that the ''true'' ΔL_{DL} can be found as the average of the DLs measured when L_1 is varied and when L_2 is varied, even if the fixed levels did not yield equal loudness. Strictly speaking, then, it would not have been necessary in the experiments to adjust the levels of the tones to approximate equal loudness, as any difference could be dealt with *post hoc*. However, as feedback was provided, it was desirable to have the two pedestals as close to equal loudness as possible.

Allen, J. B., and Neely, S. T. (**1997**). ''Modeling the relation between the intensity just-noticeable difference and loudness for pure tones and wideband noise,'' J. Acoust. Soc. Am. **102**, 3628–3646.

Berliner, J. E., and Durlach, N. I. (**1973**). ''Intensity perception. IV. Resolution in roving-level discrimination,'' J. Acoust. Soc. Am. **53**, 1270–1287.

Berliner, J. E., Durlach, N. I., and Braida, L. D. (**1977**). ''Intensity perception. VII. Further data on roving-level discrimination and the resolution and bias edge effects,'' J. Acoust. Soc. Am. **61**, 1577–1585.

Braida, L. D., Lim, J. S., Berliner, J. E., Durlach, N. I., Rabinowitz, W. M., and Purks, S. R. (**1984**). ''Intensity perception XIII. Perceptual anchor model of context-coding,'' J. Acoust. Soc. Am. **76**, 722–731.

Buus, S., and Florentine, M. (**1991**). ''Psychometric functions for level discrimination,'' J. Acoust. Soc. Am. **90**, 1371–1380.

Buus, S., Florentine, M., and Poulsen, T. (**1997**). ''Temporal integration of loudness, loudness discrimination, and the form of the loudness function,'' J. Acoust. Soc. Am. **101**, 669–680.

Buus, S., Florentine, M., and Zwicker, T. (**1995**). ''Psychometric functions for level discrimination in cochlearly impaired and normal listeners with equivalent-threshold masking,'' J. Acoust. Soc. Am. **98**, 853–861.

Carlyon, R. P., and Moore, B. C. J. (**1984**). ''Intensity discrimination: A severe departure from Weber's Law,'' J. Acoust. Soc. Am. **76**, 1369–1376.

Durlach, N. I., and Braida, L. D. (**1969**). ''Intensity perception. I. Preliminary theory of intensity resolution,'' J. Acoust. Soc. Am. **46**, 372–383.

Florentine, M. (**1986**). ''Level discrimination of tones as a function of duration,'' J. Acoust. Soc. Am. **79**, 792–798.

Florentine, M., and Buus, S. (**1981**). ''An excitation-pattern model for intensity discrimination,'' J. Acoust. Soc. Am. **70**, 1646–1654.

Florentine, M., Buus, S., and Mason, C. R. (**1987**). ''Level discrimination as a function of level for tones from 0.25 to 16 kHz,'' J. Acoust. Soc. Am. **81**, 1528–1541.

Florentine, M., Buus, S., and Poulsen, T. (**1996**). ''Temporal integration of loudness as a function of level,'' J. Acoust. Soc. Am. **99**, 1633–1644.

Green, D. M. (**1988**). *Profile Analysis* (Oxford U. P., Oxford).

Hellman, R. P., Scharf, B., Teghtsoonian, M., and Teghtsoonian, R. (**1987**). ''On the relation between the growth of loudness and the discrimination of intensity for pure tones,'' J. Acoust. Soc. Am. **82**, 448–453.

Hellman, W. S., and Hellman, R. P. (**1990**). ''Intensity discrimination as the driving force for loudness. Application to pure tones in quiet,'' J. Acoust. Soc. Am. **87**, 1255–1265.

Henning, G. B. (**1970**). ''A comparison of the effects of signal duration on frequency and amplitude discrimination,'' *in Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden).

Jesteadt, W., Wier, C. C., and Green, D. M. (**1977**). ''Intensity discrimination as a function of frequency and sensation level,'' J. Acoust. Soc. Am. **61**, 169–177.

Lim, J. S., Rabinowitz, W. M., Braida, L. D., and Durlach, N. I. (**1977**). ''Intensity perception. VIII. Loudness comparisons between different types of stimuli,'' J. Acoust. Soc. Am. **62**, 1256–1267.

Long, G. R., and Cullen, J. K. (**1985**). ''Intensity difference limens at high frequencies,'' J. Acoust. Soc. Am. **78**, 507–513.

McGill, W. J., and Goldberg, J. P. (**1968**). ''Pure-tone intensity discrimination and energy detection,'' J. Acoust. Soc. Am. **44**, 576–581.

Moore, B. C. J., and Raab, D. H. (**1974**). ''Pure-tone intensity discrimination: some experiments relating to the 'near-miss' to Weber's Law,'' J. Acoust. Soc. Am. **55**, 1049–1054.

Oxenham, A. J., and Moore, B. C. J. (**1995**). ''Overshoot and the 'severe departure' from Weber's law,'' J. Acoust. Soc. Am. **97**, 2442–2453.

Penner, M. J., Leshowitz, B., Cudahy, E., and Richard, G. (**1974**). ''Intensity discrimination for pulsed sinusoids of various frequencies,'' Percept. Psychophys. **15**, 568–570.

Rabinowitz, W. M., Lim, J. S., Braida, L. D., and Durlach, N. I. (**1976**). ''Intensity perception. VI. Summary of recent data on deviations from Weber's law for 1000-Hz tone pulses,'' J. Acoust. Soc. Am. **59**, 1506–1509.

Richards, V. M. (**1992**). ''The detectability of a tone added to narrow bands of equal-energy noise,'' J. Acoust. Soc. Am. **91**, 3424–3435.

Viemeister, N. F. (**1972**). ''Intensity discrimination of pulsed sinusoids: the effects of filtered noise,'' J. Acoust. Soc. Am. **51**, 1265–1269.

Viemeister, N. F. (**1974**). ''Intensity discrimination of noise in the presence of band-reject noise,'' J. Acoust. Soc. Am. **56**, 1594–1600.