The measurement problem in level discrimination

Daniel Shepherd[a] and Michael J. Hautus
The University of Auckland, Auckland, New Zealand

(Received 21 March 2006; revised 21 January 2007; accepted 24 January 2007)

There is disagreement among theorists over the exact measure to be used to quantify auditory level discrimination. It has been proposed that, for level discrimination tasks, the measure that is most linearly related to the sensitivity index, d', will be the correct measure. The level difference (ΔL) and the Weber fraction (Θ) are both candidates, though the latter is sensitive to the physical unit in which it is expressed (e.g., pressure or intensity) while the former is not. Psychometric functions for level discrimination were obtained at a number of pedestal levels for 10-ms sinusoids (either 1000 or 6500 Hz) and broadband noise bursts. These functions were used to assess which of three measures: ΔL, $\Theta=\Delta p/p$, or $\Theta=\Delta I/I$, is most nearly linearly related to d'. The results suggest that $\Delta p/p$ is the measure that comes closest to being linearly related to d'. © *2007 Acoustical Society of America.* [DOI: 10.1121/1.2697628]

PACS number(s): 43.66.Fe, 43.66.Dc, 43.66.Ba, 43.66.Yw, 43.66.Cb [JHG] Pages: 2158–2167

I. INTRODUCTION

Auditory level discrimination refers to the ability of an observer to distinguish between two acoustic waves differing in amplitude. Assume two stimuli are presented to an observer who is to judge which has the greater level. The first is a standard stimulus of magnitude A (the pedestal) and the second a comparison stimulus of magnitude A plus an increment ΔA. Researchers seek to find the smallest value of ΔA that allows the observer to reliably identify the stimulus that contains the increment. This minimum difference is termed the "just noticeable difference" (jnd), and is synomonous with the difference limen (DL). Among theorists, however, there exists no consensus on how the jnd should be measured, and we have termed this difficulty the *measurement problem in level discrimination*.

The measurement problem centers on how the auditory system's discriminatory capabilities should be modeled. More specifically, in an experiment probing how well an observer can discriminate auditory stimuli differing in amplitude, how should a researcher define the dependent variable? For example, should the nature of the measurement be absolute, that is $(A+\Delta A)-A$, or relative, that is, $\Delta A/A$? One absolute measure, called the *level difference* by Buus (1990), and denoted ΔL, is commonly calculated as

$$\Delta L = 20 \log_{10}\left(\frac{p+\Delta p}{p}\right), \tag{1}$$

where p indicates pressures are being used. This measure simply reflects the difference, in decibels, between the pedestal and increment, $p+\Delta p$, and the pedestal alone, p. In units of intensity (I) the level difference is sometimes known as "ΔI in dB" (Grantham and Yost, 1982). Note that the value of ΔL when pressures are used is equal to the value of ΔL when intensities have been selected (Green, 1993). A relative measure is the ubiquitous Weber fraction:

$$\frac{\Delta X}{X} = \Theta, \tag{2}$$

where X can be in units of pressure, or intensity. The Weber fraction, Θ, differs between individuals and sensory dimensions. This measure simply reflects the proportional increase in magnitude, ΔX, needed for a change in level to be detected for a given pedestal value, X.

The measurement problem in audition exists because the candidate jnd metrics, ΔL, $\Delta I/I$, and $\Delta p/p$, are proportional to one another within the typical range of human discriminatory performance. Grantham and Yost (1982) demonstrate this proportionality and Green (1988, 1993) offers approximations between ΔL, $\Delta I/I$, and $\Delta p/p$, and the latter two measures presented in decibels: $10 \log(\Delta I/I)$ and $20 \log(\Delta p/p)$. If the measures are simply transformations of one another then which is correct, and pertinently, is the choice of measure of consequence? Most Weber fractions are small, with $\Delta I/I$ typically between 0.21 and 0.73 (Green, 1993). Therefore it is difficult to distinguish between Eqs. (1) and (2), because $\ln(1+\varepsilon)\approx\varepsilon$, for small ε (Raney *et al.*, 1989). To circumvent this difficulty, experiments must be designed to deliberately inflate jnd values to a region where a nonlinear relationship exists between the jnd measures.

A common psychophysical construct applied in the measurement of discriminatory performance is the psychometric function, which plots the magnitude of an increment normalized to pedestal level (i.e., $\Delta X/X$) as a function of some performance criterion, for example, proportion correct or the sensitivity index d'. The value of ΔX that yields a jnd satisfying some predetermined performance criterion (e.g., 75% correct) and the corresponding value of X are substituted into Eq. (2) to yield a Weber fraction. In relation to the psychometric function, the measurement problem in level discrimination manifests itself as to which of ΔL, $\Delta I/I$, or $\Delta p/p$ should the performance measure be a function of. A solution to the measurement problem in level discrimination is vitally important on theoretical grounds (Doble *et al.*, 2003) because different models of auditory level discrimination pre-

[a]Electronic mail: daniel.shepherd@aut.ac.nz

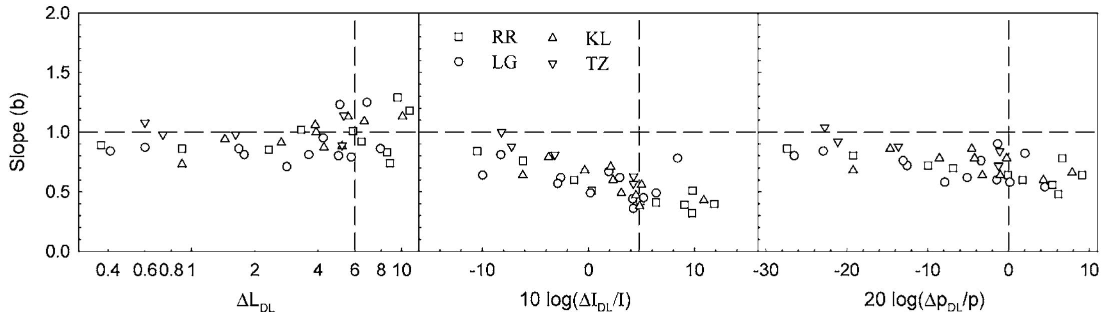

FIG. 1. Slope estimate, b, plotted as a function of the jnd expressed in terms of either ΔL_{DL} (left panel); $10\log(\Delta I_{\mathrm{DL}}/I)$ (center panel), or $20\log(\Delta p_{\mathrm{DL}}/p)$ (right panel). The dashed horizontal line is $b=1$; vertical lines demarcate proportionality (right of line) or not (left of line) between the jnds. Data, for four observers, from Buus and Florentine (1991), Table I (p. 1374).

dict different measures. Furthermore, Buus and Florentine (1991) argue that measures of the jnd based on Eq. (2) distort the relationship between stimulus magnitude and sensitivity, and hence misrepresent the sensitivity of the auditory system to changes in stimulus parameters.

Previous attempts to solve the measurement problem (Buus and Florentine, 1989; Raney *et al.*, 1989; Moore *et al.*, 1999) have focused on a popular model of the psychometric function (Egan *et al.*, 1969):

$$d' = aX^b, \tag{3}$$

where d' is the detection-theory index of discriminability (Green and Swets, 1966), a is a scalar that accounts for individual differences, and X can be ΔL, $\Delta I/I$, or $\Delta p/p$. The exponent, b, determines the slope of the psychometric function, though on occasion it is thought of as describing the *shape* of the function (Moore *et al.*, 1999). It has been proposed that the measure of X that exhibits linearity with d' is the correct metric in which to judge auditory level resolution (Buus and Florentine, 1991; Moore *et al.*, 1999). However, for stimuli with small difference limens there already exists a proportionality between the three candidates for X. For stimuli that do not afford such high sensitivity, however, the proportionality between the measures diminishes, and they can be pitted against one another. The region of proportionality is approximately below $\Delta I/I=3(\Delta L=6.02;\Delta p/p=1)$ and so emphasis should be placed on stimuli producing difference limens beyond these values when judging if X is linearly related to d'.

Three previous studies have attempted to determine which metric is linearly related to d' (Buus and Florentine, 1989; Raney *et al.*, 1989; Moore *et al.*, 1999). When Eq. (3) is plotted on log coordinates a value of b close to unity indicates linearity, which is Weber's law. Figures 1 and 2, presenting data from Buus and Florentine (1991) and Moore *et al.* (1999), respectively, plot values of b as a function of jnd: ΔL_{DL}, $10\log(\Delta I_{\mathrm{DL}}/I)$, or $20\log(\Delta p_{\mathrm{DL}}/p)$. The dashed horizontal lines are $b=1$, while the dashed vertical lines discern the zone of proportionality (to the left of the line) or nonproportionality (to the right of the line).

Inspection of Figs. 1 and 2 reveals conflicting conclusions, with Buus and Florentine (Fig. 1) concluding that d' is linearly related to ΔL, while the data of Moore *et al.* (Fig. 2) suggest that d' is linearly related to $\Delta I/I$. The stimulus configurations utilized by the two studies were, however, different, and may account for the lack of agreement between the studies (Laming, 1986). Buus and Florentine (1991) employed a difference discrimination task, while Moore *et al.* (1999) used stimuli more consistent with an increment detection task. Difference discrimination involves discriminating between two stimuli (X and $X+\Delta X$) separated either in space or time. Increment detection involves a continuous uniform stimulus (X) above the level of background noise and the addition of an increment (ΔX). The possibility that different metrics underlie these two tasks is of fundamental importance to those modeling the auditory system.

The findings of Raney *et al.* (1989), using complex profile stimuli, were inconclusive, and their data failed to indicate which of ΔL or $\Delta p/p$ was linearly related to d'. It is clear that further evidence is required to resolve this problem. Consequently, we conducted three difference discrimination experiments in which the stimuli were specifically chosen to yield relatively large jnds.

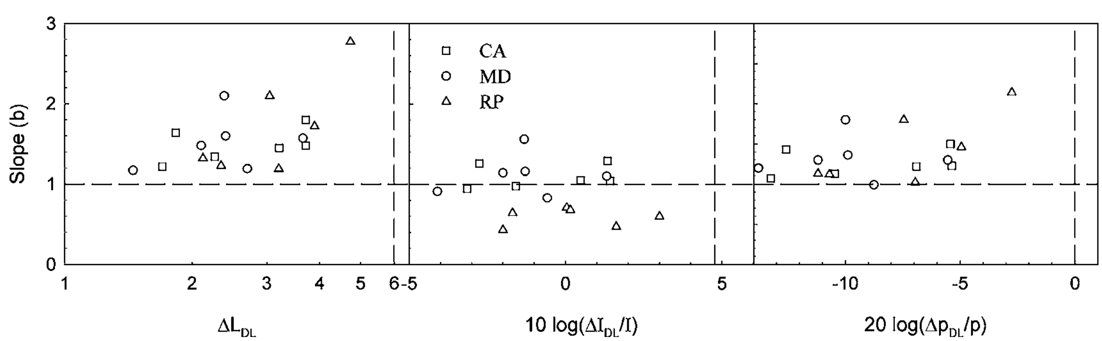

FIG. 2. As for Fig. 1, data, for three observers, from Moore *et al.* (1999).

II. EXPERIMENT 1: 1000 Hz SINSUOIDS IN GATED NOISE

A. Introduction

Among the stimulus configurations employed by Buus and Florentine (1991) were 10-ms 1000-Hz sinusoids. Experiment 1 adopts the same sinusoidal dimensions but additionally includes a broadband-noise background, the purpose of which was to produce larger difference limens through direct masking of the stimuli. The noise also serves to mask spectral splatter associated with short-duration stimuli, a known facilitator of the detection process.

B. Method

1. Observers

Four males: WC (aged 24), MK (aged 29), EL (aged 33), and DS (aged 30) served as observers. All had extensive experience in psychoacoustic tasks and had normal audiometric thresholds in the ear to be tested. All but the author, DS, received a monetary incentive to participate in the experiment.

2. Stimuli

Stimuli were 1000-Hz sinusoids temporally centered in gated broadband noise. The sinusoid had a duration of 10 ms with 1-ms rise and fall times ($\cos^2$) and was generated at a sampling rate of 44.1 kHz. Ten pedestal levels were employed, ranging from 15 to 60 dB sound pressure level (SPL) in 5-dB SPL steps. The waveform of the pedestal, A, differed from the pedestal-plus-increment, $A+\Delta A$, only in the amount of attenuation it was subjected to. The broadband noise (N_0 $=35$ dB SPL) was 200 ms in duration with 10-ms rise and fall times ($\cos^2$).

3. Apparatus

The gated noise and 1000-Hz sinusoid were generated independently using National Instruments LABVIEW 6.1, and then converted from a digital to an analog representation (NI PCI-6052E). The noise was directed to a pair of static attenuators (TDT, PA4) whose level of attenuation remained constant across the experimental block. The sinusoids were directed through two programmable attenuators (TDT, PA5), set up in series, and then added to the noise in a signal mixer (TDT, SM5). Once combined, the noise and sinusoid were delivered to a headphone buffer (TDT, HB7) and from there to an earphone (TDH 49P, No. 30195). All stimuli were presented monaurally to the observer's left ear.

In addition to generating the stimuli, the LABVIEW software also controlled the programmable attenuators, presented instructions to the observer on a terminal positioned within the sound-attenuating chamber (Amplaid Model E) housing the observer, and, through an auxiliary keyboard, recorded the observer's responses, and controlled feedback lights.

4. Procedure

A two-alternative forced-choice (2-AFC) adaptive three-down, one-up staircase procedure (Levitt, 1971) was used to measure difference limens for each of the ten pedestal levels. This provided the observers practice as well as facilitating in the selection of increments to be used in a subsequent difference discrimination task. Each difference limen was based upon three blocks of trials, each consisting of 15 reversals. The first three reversals changed the pedestal-plus-increment level by ±3 dB SPL, while for subsequent reversals this change was ±1 dB SPL. Any single block returning a standard deviation greater than 2 dB was discarded and repeated. The adaptive procedure initially estimated the difference limen (DL) expressed in terms of the level difference: $\Delta L_{DL}=20\log[(p+\Delta p_{DL})/p]$, where Δp_{DL} is the sound pressure increment (Buus and Florentine, 1991). To calculate difference limens in pressure, Δp_{DL}, the mean of the last 12 reversals for each block were averaged and then converted to the difference limen by solving Eq. (1) for Δp.

Once difference limens had been estimated from the adaptive data, a variation of the method of constant stimuli (Moore *et al.*, 1999) was used to collect psychometric functions for ten pedestal levels. Observers were presented two intervals per trial with one of those intervals containing a pedestal, and the other a pedestal plus an increment. The interval containing the increment was determined randomly with an equal *a priori* probability (i.e., $p=0.5$). The observer was instructed to indicate on a keypad located in the experimental chamber the interval that contained the increment. Trial-by-trial feedback was provided contingent on response.

Empirical psychometric functions were collected with each function based on five increment levels that ranged from −10 to +10 dB SPL in 5-dB SPL steps with reference to the observer's difference limen (i.e., Δp_{DL}). Each psychometric function was based on five blocks of trials, with each of the five points based on 105 trials. The first ten trials of any block were designated practice trials and were omitted from the final analyses.

5. Data analysis

Empirical psychometric functions were constructed for each subject by plotting $\log d'$ vs $\log \Delta L$, $10\log(\Delta I/I)$, and $20\log(\Delta p/p)$ for each of the ten pedestal levels, yielding 30 functions in all. Values of d' were derived from percentage correct scores using an approximation developed by Hacker and Ratcliff (1979). Equation (3) is a basic power function, and linearity between d' and X occurs when the exponent, b, equals one. It is customary to represent Eq. (3) on double-logarithmic coordinates, where the family of power functions is transformed to a family of straight lines. The exponent, b, is the slope of the line, and the scalar, a, is the intercept. Consequently, b is commonly referred to as the *slope* parameter. The parameter, a, is generally not of interest; suffice it to say that it reflects differences in sensitivity between observers.

The psychometric functions were fitted with Eq. (3) using the method of least squares. This provided two parameter estimates for each function: a and b. Next, the value of the jnd was estimated by substituting the estimates of a and b into Eq. (3) and setting d' equal to unity. The value $d'=1$ is conventionally regarded as performance at threshold and is equal to 76% correct in a 2-AFC task (Green and Swets,

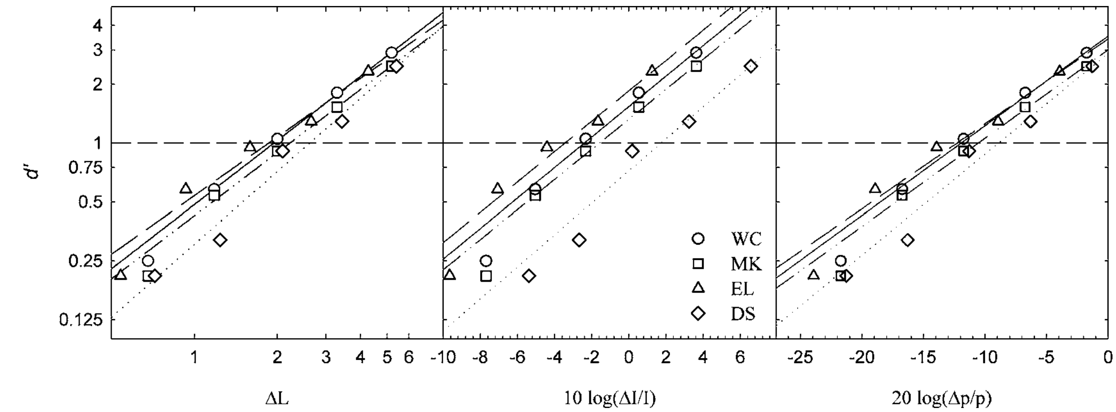

FIG. 3. Psychometric functions plotting the sensitivity index d' as a function of either ΔL (left), $10 \log(\Delta I_{DL}/I)$ (center), or $20 \log(\Delta p_{DL}/p)$ (right) for four observers. The pedestal was a 1-kHz, 10-ms pedestal of 30 dB SPL. The dashed horizontal lines represent performance at threshold (i.e., $d'=1$). The dashed diagonal lines are the best-fitting forms of Eq. (3) for each observer.

1966). Solving Eq. (3) for X provided an estimate of the jnd, against which the slope parameter, b, was subsequently plotted.

C. Results

Figure 3 shows psychometric functions for a single pedestal level: 30 dB SPL. The best-fitting lines, from Eq. (3), sufficiently accounted for data across all three of the experiments, with the goodness-of-fit statistic, R^2, being greater than 0.9 for each psychometric function. Table I lists mean parameter estimates and goodness-of-fit statistics for each observer obtained from fitting Eq. (3) to the data obtained at each pedestal level. The high values of R^2 show that the data were well accounted for by the equation. A one-sample t-test on the data from each subject showed that all estimates of b were significantly different from one ($p<0.05$) with the exception of two cases: observers MK $[t(9)=0.852, p=0.416]$ and DS $[t(9)=0.69, p=0.508]$ did not have estimates of b different from unity when $X=\Delta p/p$.

Figure 4 plots b as a function of jnd, with each plot consisting of 40 points (four observers by ten pedestal levels). The slope, b, is different across the three jnd measures $[F(2,119)=93.23, p<0.001]$ with mean values, across observers and pedestals, being 1.22 (s.d. $=0.2$) for ΔL, 0.75 (s.d. $=0.14$) for $\Delta I/I$, and 0.97 (s.d. $=0.12$) for $\Delta p/p$. The slope estimates for $\Delta L[t(39)=7.26, p<0.001]$ and $\Delta I/I/[t(39)=-11.24, p<0.001]$ were significantly different from unity, but the slope parameter for $\Delta p/p[t(39)=-1.56, p=0.126]$ was not. These results indicate that the best measure of level discrimination for brief 1000-Hz sinusoids is the Weber fraction expressed in units of pressure.

III. EXPERIMENT 2: BROADBAND NOISE WITH 3-AFC

A. Introduction

Buus and Florentine (1991) utilized noise that covered the audible frequency range ($f_c=22$ kHz), had a noise power density of 20 dB SPL, and was 500 ms in duration. A direct comparison between the limens they obtained with noise to

TABLE I. Estimates of best-fitting parameters and of the corresponding fit statistic, R^2, for equations (a) $d'=a\Delta L^b$; (b) $d'=a(\Delta I/I)^b$, and; (c) $d'=a(\Delta p/p)^b$ in Experiment 1. Asterisks signify the means differ significantly from unity ($p<0.05$).

	a		b		R^2	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
(a) $d'=a\Delta L^b$						
WC	0.311	0.151	1.164*	0.173	0.988	0.009
EL	0.356	0.135	1.096*	0.104	0.982	0.012
MK	0.276	0.148	1.285*	0.249	0.981	0.016
DS	0.164	0.084	1.350*	0.135	0.984	0.001
(b) $d'=a(\Delta I/I)^b$						
WC	1.109	0.666	0.706*	0.099	0.990	0.005
EL	1.257	0.524	0.712*	0.122	0.987	0.008
MK	1.111	0.627	0.811*	0.165	0.982	0.009
DS	0.782	0.345	0.778*	0.153	0.988	0.006
(c) $d'=a(\Delta p/p)^b$						
WC	2.312	0.786	0.923*	0.071	0.990	0.006
EL	2.536	0.876	0.894*	0.074	0.980	0.018
MK	2.418	0.764	1.039	0.146	0.982	0.010
DS	1.854	0.805	1.024	0.110	0.988	0.007

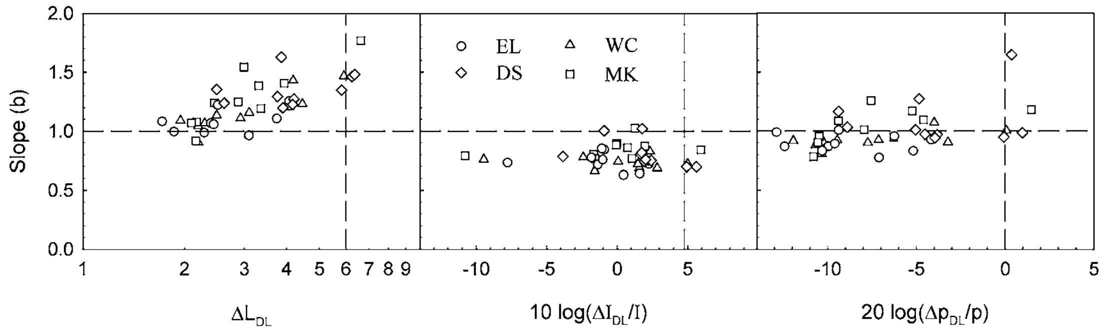

FIG. 4. Slope estimate, b, plotted as a function of the jnd expressed in terms of either ΔL_{DL} (left panel); $10\log(\Delta I_{\mathrm{DL}}/I)$ (center panel); or $20\log(\Delta p_{\mathrm{DL}}/p)$ (right panel). The dashed horizontal line is $b=1$; vertical lines demarcate proportionality (right of line) or not (left of line) between the jnds. Stimuli were 10-ms bursts of 1000-Hz sinusoids presented in noise. Data for four observers.

those they derived using 500-ms 1000-Hz sinusoids showed the noise thresholds were higher than the sinusoid thresholds. We hypothesize that such a difference will also hold between noise and sinusoids of 10-ms duration. Moore *et al.* (1999) implied that difference limens could also be increased by making the observer's task more difficult. The addition of an extra observation interval should increase the uncertainty inherent in the observer's decision, which has long correlated with task difficulty (Green and Swets, 1966). Confusion over when a signal occurs forces an observer to attend to an increased number of noise (i.e., pedestal alone) channels in an effort to detect the increment (Nachmias and Kocher, 1970). As the number of noise channels needing to be monitored increases, the psychometric function becomes shallower, increasing the difference limen. For example, uncertainty about stimulus duration (Dai and Wright, 1998) and temporal occurrence (Watson and Nichols, 1975) serve to introduce nonlinearities into the psychometric function. Consequently, in addition to adopting 10-ms bursts of Gaussian noise as stimuli, we increased the difficulty of the task by employing a 3-AFC procedure.

B. Method

1. Observers

Three observers undertook Experiment 2; two had participated in Experiment 1 (EL and DS), while one had not (MF, a 37 year old female). MF had extensive experience in auditory psychophysical tasks. MF had normal audiometric thresholds (re: ISO Standard, 1975) at all frequencies tested, bar 6000 Hz. MF's audiogram exhibited no thresholds greater than 20 dB HL. Only EL received a financial incentive to participate.

2. Stimuli

Stimuli were 10-ms broadband noises low-pass filtered at 8000 Hz. Filtering was undertaken with a fourth-order Butterworth filter. The ten noise pedestals had spectrum levels of −15, −10, −5, 0, 5, 10, 15, 20, 25, and 30 dB SPL. The noise had rise and fall times ($\cos^2$) of 1 ms. Masking noise was absent throughout.

3. Apparatus

The apparatus was identical to that employed in Experiment 1. Noise pedestals and increments were generated independently using National Instruments LABVIEW 6.1. The pedestals and increments were directed to the pair of static attenuators (TDT, PA4) and the pair of programmable attenuators (TDT, PA5), respectively, and they were combined at the signal mixer (TDT, SM5).

4. Procedure

The procedure was identical to that employed in Experiment 1 except that there were, on any one trial, three, as opposed to two, observation intervals. The adaptive three-down, one-up 3-AFC procedure located the difference limen that corresponds to 79% correct for each of the ten pedestals (Levitt, 1971). The variant of the Method of Constant Stimuli employed also had three observation intervals. Five increment levels were employed, defined with respect to the difference limen estimates obtained with the adaptive procedure. They ranged from −10 to +10 dB re: Δp_{DL}, in 5-dB steps.

C. Results

Estimates of a and b, averaged across the ten pedestal levels for each observer, are presented in Table II for each jnd measure. The goodness-of-fit statistics, R^2, again indicate that Eq. (3) provides an acceptable fit to the data ($R^2 > 0.97$). A one-sample t-test on the data from each observer showed that, for ΔL and $\Delta I/I$, all mean estimates of b were significantly different from one ($p<0.001$). For $\Delta p/p$, results for two observers [MF($t(9)=-0.153, p=0.882$]; EL[$t(9)=-0.209, p=0.839$)] were not significantly different from one, whereas that for DS was [$t(9)=4.967, p<0.001$].

Inspection of Fig. 5 suggests that the slope parameter, b, depends on the jnd measure. Analysis-of-variance confirms that this is the case [$F(2,89)=253.8, p<0.001$], with a *post hoc* test (Bonferonni) indicating that all three means were significantly different from each other ($p<0.001$). These mean estimates, obtained by averaging across both observer and pedestal level, are $b=1.694$ (s.d.$=0.163$) for ΔL, $b=0.769$ (s.d.$=0.0689$) for $\Delta I/I$, and $b=1.040$ (s.d.$=0.1$) for $\Delta p/p$. These means were significantly different from unity for $\Delta L[t(29)=15.82, p<0.001]$ and $\Delta I/I[t(9)=$

TABLE II. Estimates of best-fitting parameters and of the corresponding fit statistic, R^2, for the equations (a) $d' = a\Delta L^b$; (b) $d' = a(\Delta I/I)^b$, and; (c) $d' = a(\Delta p/p)^b$ in Experiment 2. Asterisks signify the means differ significantly from unity ($p < 0.05$).

	a		b		R^2	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
(a) $d' = a\Delta L^b$						
MF	0.085	0.037	1.498*	0.164	0.983	0.011
EL	0.113	0.051	1.420*	0.184	0.988	0.006
DS	0.126	0.043	1.491*	0.142	0.993	0.006
(b) $d' = a(\Delta I/I)^b$						
MF	0.507	0.099	0.702*	0.084	0.973	0.019
EL	0.546	0.104	0.763*	0.069	0.983	0.009
DS	0.619	0.141	0.842*	0.050	0.985	0.010
(c) $d' = a(\Delta p/p)^b$						
MF	1.135	0.278	0.995	0.099	0.979	0.015
EL	1.294	0.212	1.019	0.104	0.987	0.006
DS	1.688	0.381	1.104*	0.066	0.989	0.008

$-14.27, p < 0.001$], but not significantly different for $\Delta p/p[t(29) = 1.82, p = 0.079]$. Thus, of the three candidate measures: ΔL, $\Delta I/I$, or $\Delta p/p$, the evidence again favors $\Delta p/p$.

There is evidence that employing noise and increasing task complexity increased ΔL's, possibly through the added uncertainty in the decision-making process. The difference in ΔL's, averaged across pedestal levels and observers, between Experiment 1 ($\bar{x} = 3.38$, s.d. $= 1.33$, $n = 40$) and Experiment 2 ($\bar{x} = 4.94$, s.d. $= 1.11$, $n = 30$) are significantly different ($\alpha = 0.05$, one-tailed) from those estimated in Experiment II [$t(68) = -5.21$, $p < 0.001$]. For the two observers that participated in both experiments, DS and EL, the difference between Experiment 1 ($\bar{x} = 3.58$, s.d. $= 1.37$) and Experiment 2 ($\bar{x} = 4.59$, s.d. $= 0.8$) is not as pronounced but is still significantly different ($t(38) = -2.84, p = 0.004$).

IV. EXPERIMENT 3: 6500 Hz SINUSOIDS IN BANDSTOP NOISE

A. Introduction

Buus and Florentine (1991) demonstrated that for sinusoids between 1 and 14 kHz the difference limen increases with frequency. Thus, larger jnds are obtained for high frequency sinusoids. Additionally, the severe departure to Weber's law manifests itself as inflated difference limens for midlevel ($\approx 35-55$ dB SPL) high-frequency sinusoids (>5000 Hz).

Carlyon and Moore (1984) reported that the addition of bandstop noise boosted the difference limen for a sinusoidal signal centered in the spectral notch, expressed in units of intensity, by approximately $\Delta I_{DL} = 5$ dB. Because this increase in the difference limen was evident only for midlevel pedestals, it appears the addition of bandstop noise serves to enhance the severe departure to Weber's law.

These findings suggest that relatively large jnds can be attained for the discrimination of high-frequency sinusoids in bandstop noise. Our choice of a 6500-Hz sinusoid was, in part, determined by the response characteristics of the headphones, which effectively acted as a low-pass filter with a cut-off of 8000 Hz. Additionally, Carlyon and Moore (1984, 1986) utilized sinusoids at this frequency, and both studies utilized identical bandstop masking noise.

B. Method

1. Observers

There were two participants, WC and DS, both of whom had also participated in Experiment 1. WC received financial compensation.

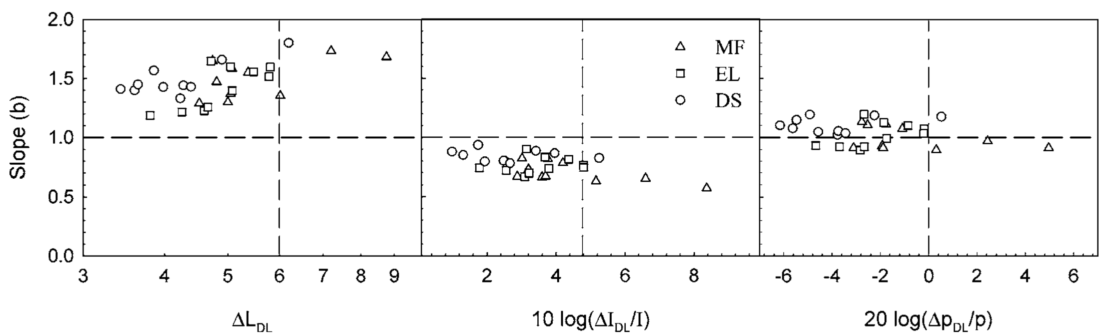

FIG. 5. As for Fig. 4. Stimuli were 10-ms broadband noise bursts. Data for three observers.

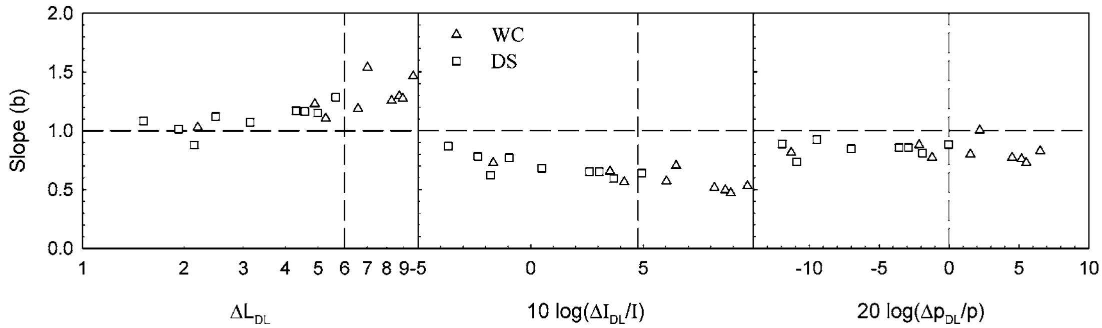

FIG. 6. As for Fig. 4. Stimuli were 10-ms bursts of 6.5-kHz sinusoids presented in bandstop noise. Data for two observers.

2. Stimuli

Stimuli were 6500-Hz sinusoids embedded in a bandstop noise background. Two bands of noise ($W=1950$ Hz), one centered on 4875 Hz and the other on 8125 Hz were produced with fourth-order Butterworth filters, and then added. The notch width was therefore 1300 Hz, and the 6500-Hz sinusoid fell in the middle of the notch. The extremities of the higher frequency band of noise extended beyond the frequency response of the observer's ear piece (≈8000 Hz). The nine pedestals had levels of: 20–60 dB SPL in 5-dB steps. The five increment levels ranged from −10 to +10 dB SPL in 5-dB steps, with reference to the observer's difference limen, Δp_{DL}. Both the noise and the sinusoids were of 10-ms duration, including 1-ms onsets and offsets ($\cos^2$).

3. Apparatus and procedure

The apparatus and procedures used were identical to those employed in Experiment 1.

C. Results

Figure 6 plots the exponent b as a function of jnd for the two observers. Table III provides the mean parameter estimates for each observer and each measure. The goodness-of-fit is again very good, with Eq. (3) sufficiently accounting for the data ($R^2 > 0.97$). Two one-sample t-tests performed on each subject's data showed that values of b were significantly different from unity: ($p < 0.001$) regardless of the measure used to represent the jnd. Examination of the data for these two observers shows that, for the 36 estimates of b associated with $\Delta I/I$ and $\Delta p/p$, only one was greater than unity. For the 18 estimates of b associated with ΔL, only one was less than unity. This latter finding reflects the results obtained for ΔL with 1000-Hz sinusoids (see Experiment 1) and noise (Experiment 2).

The mean estimates of b, obtained by averaging across both observer and pedestal level, were $\Delta L=1.25$ (s.d. $=0.22$), $\Delta I/I=0.655$ (s.d. $=0.09$), and $\Delta p/p=0.881$ (s.d. $=0.71$), and are significantly different $[F(2,55)=58.16, p<0.001]$. Bonferonni *post hoc* analyses indicate that all three means are different from each other ($p<0.001$). Additionally, these mean slope estimates are all significantly different from unity $[\Delta L(t(17)=5.101, p<0.001]$; $\Delta I/I[t(17)=-17.063, p<0.001]$; $\Delta p/p[t(17)=-7.53, p<0.001)]$. From this analysis it must be concluded that none of the three candidate measures, ΔL, $\Delta I/I$, and $\Delta p/p$, obtained strict linearity with d'. All measures produced slope estimates significantly different from unity, with the measures based on the Weber fraction ($\Delta I/I$ and $\Delta p/p$) yielding slopes less than unity, while ΔL produced slopes greater than unity. Of the three measures, $\Delta p/p$ is the closest to unity.

V. DISCUSSION

In order to ascertain which of ΔL, $\Delta I/I$, or $\Delta p/p$ achieves a linear relationship with d', a series of three experiments were undertaken using 1000-Hz sinusoids (Experiment 1), broadband noise (Experiment 2), or 6500-Hz sinusoids in bandstop noise (Experiment 3). All stimuli were presented monaurally for a duration of 10 ms. Figure 7 illus-

TABLE III. Estimates of best-fitting parameters and of the corresponding fit statistic, R^2, for the equations (a) $d'=a\Delta L^b$, (b) $d'=a(\Delta I/I)^b$, and (c) $d'=a(\Delta p/p)^b$ in Experiment 3. Asterisks signify the means differ significantly from unity ($p<0.05$).

	a		b		R^2	
	Mean	s.d.	Mean	s.d.	Mean	s.d.
(a) $d'=a\Delta L^b$						
WC	0.114	0.133	1.367*	0.213	0.994	0.006
DS	0.329	0.187	1.091*	0.119	0.990	0.007
(b) $d'=a(\Delta I/I)^b$						
WC	0.498	0.331	0.616*	0.068	0.981	0.011
DS	1.041	0.508	0.687*	0.104	0.973	0.016
(c) $d'=a(\Delta p/p)^b$						
WC	1.0102	0.750	0.878*	0.065	0.989	0.009
DS	2.329	1.306	0.881*	0.090	0.986	0.008

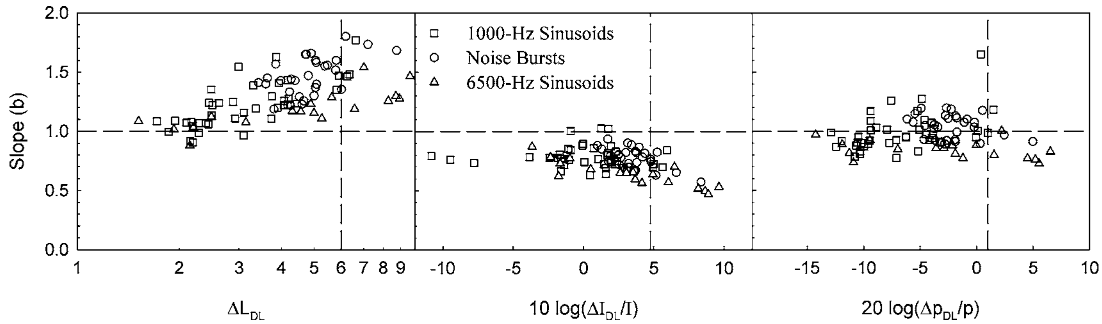

FIG. 7. Slope estimate, b, plotted as a function of the jnd expressed in terms of either ΔL_{DL} (left panel), $10\log(\Delta I_{DL}/I)$ (center panel), or $20\log(\Delta p_{DL}/p)$ (right panel). The symbols represent three different configurations of stimuli.

trates the slope estimates, b, obtained in all three experiments, for each of the jnd measures. Data points clustering around the dashed horizontal lines indicate linearity between d' and the jnd. From these data there is evidence that, when the jnd is represented as the Weber fraction, $\Delta X/X$, d' is linear to pressure ($\Delta p/p$). If $\Delta I/I$ is selected as the jnd, then b appears to be consistently lower than unity. When the measure of the jnd is taken to be ΔL, the slope exponent b progressively increases as ΔL increases.

Further support for the claim that d' is linearly related to $\Delta p/p$ comes from analysis of the entire data set. Averaging all estimates of b across the three experiments gives the following means: ΔL=1.312 (s.d.=0.22), $\Delta I/I$=0.736 (s.d. =0.12), and $\Delta p/p$=0.977 (s.d.=0.12). Only the mean for $\Delta p/p$ is not significantly different from unity [$\Delta p/p(t(89)=-1.873, p=0.064]$; $\Delta L[t(89)=13.457, p<0.001]$; $\Delta I/I[t(89)=-20.587, p<0.001)]$. This conclusion is consistent with Laming's (1986) sensory analytical model that predicts $\Delta p/p$ to be the correct measure of auditory level discrimination.

The relationship among ΔL, $\Delta I/I$, and $\Delta p/p$ was discussed in Sec. I. It was stressed that for $\Delta I/I<3$, a proportionality exists among ΔL, $\Delta I/I$, and $\Delta p/p$. Beyond this region, however, the proportionality no longer holds, and it is argued (e.g., Buus and Florentine, 1991; Moore *et al.*, 1999) that stimuli falling into this region (i.e., $\Delta I/I>3$) should be given heavier weighting when judging the measure obtaining linearity with d'. The dashed vertical lines in Fig. 7 represent the value of the jnd where $\Delta I/I$ equals three. This occurs at ΔL=6.02, $10\log(\Delta I/I)$=4.77, and $20\log(\Delta p/p)$=0. Adopting this criterion and examining the data to the right of the vertical lines in Fig. 7, it is again apparent that of the three measures $\Delta p/p$ is the candidate that obtains the most convincing linearly relationship with d'.

In contrast Buus and Florentine (1991), on the basis of their data (see Fig. 1), concluded that d' is linearly related to ΔL. They did, however, report departures from linearity, notably for stimuli possessing large jnds, where b consistently exceeded unity. They explained this inflation of b in terms of bias inherent both in data analytical procedures and attention lapses on the part of the observer. An increase in b with higher values of ΔL_{DL} has been found consistently during the course of the current investigation. Given the value placed upon stimuli producing large jnd measures in the elucidation of the correct measure to employ in level discrimination, Buus and Florentine's conclusion is in need of further empirical support. The data and subsequent interpretations presented by Moore *et al.* (1999) are also not reflected in the present study. They found that when ΔL and $\Delta p/p$ were selected as the jnd all estimates of b were above unity (see Fig. 2). In contrast the estimated exponents for $\Delta I/I$ fell on either side of unity. However, a comment on stimulus context is warranted. The data reported by Buus and Florentine (1991) and the current study were obtained from a traditional difference discrimination task, while that of Moore *et al.* (1999) are reported to have come from an increment detection task. Thus the evidence suggests that it is unlikely that a single jnd metric will be able to account for data obtained using both forms of stimulus configuration. This fundamental difference in the way the auditory system resolves level with respect to stimulus context has also been found with psychometric functions (Green and Sewall, 1962; Laming, 1986).

Estimates of b as a function of pedestal level for each of the three experiments are displayed in Fig. 8. Inspection of these figures reveals that the metric associated with the greatest amount of variability in b is ΔL (Fig. 8, left columns), though the data did not permit meaningful significance testing to be undertaken. However, it does appear that, for jnds expressed in terms of $\Delta p/p$ (Fig. 8, right columns) and $\Delta I/I$ (Fig. 8, center columns), b is relatively stable across pedestal levels.

VI. CONCLUSION

Three experiments employing 10-ms 1000-Hz sinusoids, broadband noise, and 6500-Hz sinusoids indicate that d' is most linearly related to $\Delta p/p$. These results differ fundamentally to those described by Buus and Florentine (1991) and Moore *et al.* (1999). That these differences exist among the three independent studies is of interest, and it is clear that further investigation is called for. One promising direction is suggested by Ward and Davidson (1993), who showed that large Weber fractions can be obtained from pedestals of low frequency and level.

ACKNOWLEDGMENTS

The authors would like to thank B. C. J. Moore for making data available for further analysis. This research was supported by grants from the University of Auckland Department of Psychology Research Expenses Fund and

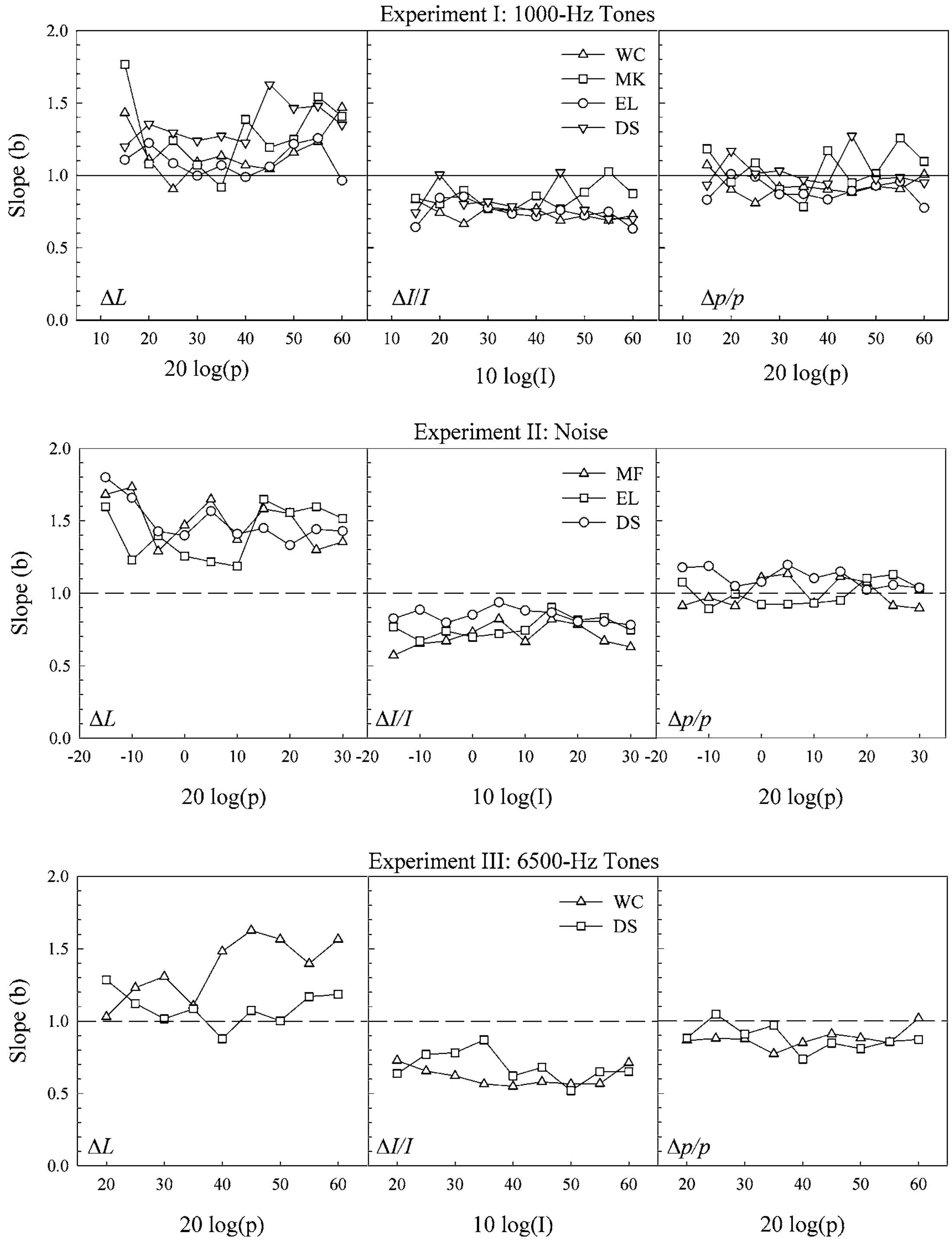

FIG. 8. The slope parameter, b, as a function of pedestal level for jnds expressed as ΔL (left), $10\log(\Delta I_{\mathrm{DL}}/I)$ (center), and $20\log(\Delta p_{\mathrm{DL}}/p)$ (right). The top (Experiment 1), middle (Experiment 2), and bottom (Experiment 3) panels represent the three stimulus configurations used, while the legends indicate the observers.

Conference Travel Fund, and also from the University of Auckland Research Committee. The authors would like to dedicate this paper to the memory of Soren Buus, who kindly provided valuable comments on experimental design and data analysis during its formative stage.

Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643–2654.

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

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.

Carlyon, R. P., and Moore, B. C. J. (**1986**). "Detection of tones in noise and the 'severe departure' from Weber's law," J. Acoust. Soc. Am. **79**, 461–464.

Dai, H., and Wright, B. A. (**1998**). "Predicting the detectability of tones with unexpected durations," J. Acoust. Soc. Am. **105**, 2043–2046.

Doble, C. W., Falmagne, J., and Berg, B. G. (**2003**). "Recasting Weber's law," Psychol. Rev. **110**, 365–375.

Egan, J. P., Linder, W. A., and McFadden, D. (**1969**). "Masking level differences and the form of the psychometric function," Percept. Psychophys. **6**, 209–215.

Grantham, D. W., and Yost, W. A. (**1982**). "Measures of intensity discrimi-

nation," J. Acoust. Soc. Am. **72**, 406–410.
Green, D. M. (**1988**). *Profile Analysis* (Oxford University Press, New York).
Green, D. M. (**1993**). "Auditory intensity discrimination," in *Human Psychophysics*, edited by W. A. Yost, A. N. Popper, and R. R. Fay (Springer, New York).
Green, D. M., and Sewall, S. T. (**1962**). "Effects of background noise on auditory detection of noise bursts," J. Acoust. Soc. Am. **34**, 1207–1216.
Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).
Hacker, M. J., and Ratcliff, R. (**1979**). "A revised table of d' for M-alternative forced choice," Percept. Psychophys. **26**, 168–170.
Laming, D. (**1986**). *Sensory Analysis* (Academic London).
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Moore, B. C. J., Peters, R. W., and Glasberg, B. R. (**1999**). "Effects of frequency and duration on psychometric functions for detection of increments and decrements in sinusoids in noise," J. Acoust. Soc. Am. **106**, 3539–3552.
Nachmias, J., and Kocher, E. C. (**1970**). "Visual detection and discrimination of luminance increments," J. Opt. Soc. Am. **60**, 382–389.
Raney, J. J., Richards, M., Onsan, Z. A., and Green, D. M. (**1989**). "Signal uncertainity and psychometric functions in profile analysis," J. Acoust. Soc. Am. **86**, 954–960.
Watson, C. S., and Nichols, T. L. (**1975**). "Detectability of auditory signals presented without defined observation intervals," J. Acoust. Soc. Am. **59**, 655–668.
ISO Standard (1975).
Ward, L. M., and Davidson, K. P., (**1993**), "Where the action is: Weber fractions as a function of sound pressure at low frequencies," J. Acoust. Soc. Am. **94**, 2587-2594