J Am Acad Audiol 12: 113–120 (2001)

Maximum-Likelihood Yes-No Procedure for Gap Detection: Effect of Track Length

Mary Florentine*
Peter Marvit*
Søren Buus†

Abstract

A maximum-likelihood yes-no procedure was used to measure minimum detectable gaps (MDGs) at 1 and 4 kHz using two track lengths (15 and 30 trials). Results from 11 normal listeners show no difference between MDGs for the two track lengths, and variability of the MDGs did not differ significantly for the track lengths. Results from simulations indicate that the variability of MDGs from real listeners is considerably larger than that imposed by the psychophysical procedure. Additional simulations show that random variability of listeners' response criterion is a compelling explanation for the excess variability. These findings indicate that changes in a listener's threshold over time impose a lower bound on the variability obtainable with a yes-no procedure. They imply that increasing the number of trials in a track, beyond the minimum number required to obtain a stable threshold estimate, offers little or no advantage for the test–retest reliability of a clinical threshold measurement.

Key Words: Gap detection, method of maximum likelihood, narrowband noise, psychophysical procedure, test–retest reliability

Abbreviations: CF = center frequency, ERB = equivalent rectangular bands, MDG = minimum detectable gap, MML = method of maximum likelihood, SDT = signal detection theory

This article examines a procedure that will be useful for assessing individual listeners' auditory processing of suprathreshold sounds. These abilities frequently cannot be inferred from a client's audiogram, but knowledge of them is highly likely to be useful for optimal fitting of advanced digital hearing aids. A few examples are noteworthy: The rate of loudness growth with increasing intensity can differ in listeners with nearly identical audiograms (Hellman, 1994; Florentine et al, 1997). The ability to detect a pause in a noise can differ between impaired listeners with identical audiograms and even differ within the same listener tested in different frequency ranges (Florentine and Buus, 2000). Likewise, measurements of tuning curves indicate that some listeners have areas of the basilar membrane in which transduction is severely impaired and/or absent (Thornton and Abbas, 1980; Florentine and Houtsma, 1983; Turner et al, 1983; Moore et al, 2000) and that listeners differ with respect to the functional limits of high-frequency innervation in the cochlea (Buus et al, 1986; Moore et al, 2000). The latter information could be important when recommending high-frequency emphasis in listeners with sloping high-frequency losses (Villchur, 1973; Hogan and Turner, 1998; Turner and Cummings, 1999). In fact, a growing body of data indicates that psychoacoustic abilities and speech perception in noise can differ widely in impaired listeners with the same audiometric configuration. Therefore, suprathreshold psychoacoustic tests need to be developed for clinical use.

A first step in obtaining potentially important information about individual auditory systems is to develop a clinically feasible psychophysical procedure that is applicable to a variety

*Institute for Hearing, Speech, and Language and Department of Speech-Language Pathology and Audiology, †Communications and Digital Signal Processing Center, Department of Electrical Engineering, Northeastern University, Boston, Massachusetts

Reprint requests: Mary Florentine, Institute for Hearing, Speech, and Language and Department of Speech-Language Pathology and Audiology (133 FR), Northeastern University, 360 Huntington Avenue, Boston, MA 02115

of psychoacoustic tasks while being sufficiently fast and yielding a low enough variability to provide useful fitting information in a clinical setting. A maximum-likelihood yes-no procedure is very attractive because a reliable psychophysical measurement can be obtained in relatively few trials when compared with a traditional up-down forced-choice procedure (Green, 1993; Florentine et al, 2000; Leek et al, 2000). In this procedure, a single stimulus is presented, and the listener is instructed to report the presence or absence of the signal. Several catch trials are usually presented in randomly chosen trials to help estimate the false-alarm rate. The procedure employs a large number of candidate psychometric functions and after each trial calculates the probability (or likelihood) of obtaining the listener's responses to all of the stimuli that have been presented given each psychometric function. The psychometric function yielding the highest probability is then used to determine the stimulus to be presented on the next trial. Within 12 to 24 test trials, the procedure usually converges on a stable estimate of the most likely psychometric function from which a threshold can be obtained. At the end of a track, the most likely psychometric function is used to calculate a threshold estimate. (For a general introduction to the maximum-likelihood procedure, see Formby et al [1996]. For a historical review, see Florentine et al [2000].)

It is important that the outcome of any clinical procedure is repeatable (i.e., the variability of the data is within reasonable bounds). In theory, the variability of any test can be made as small as desired by increasing the number of trials. However, to make a test efficient, the number of trials should be kept as small as possible. Therefore, it is important to investigate the trade-off between the number of trials and variability. On simple statistical grounds, one might expect the standard deviation across repeated measurements to be inversely proportional to the square root of the number of trials, $\sqrt{N}$. However, Green's (1993) computer simulations showed that the standard deviation decreased faster than the reciprocal of $\sqrt{N}$ for moderate Ns, especially for non-zero false-alarm rates. For a typical false-alarm rate of 0.1, simulations of pure-tone detection thresholds showed a rapid decrease in the variability up to about 24 trials per threshold estimate. Accordingly, Leek et al (2000) advocate that 24 trials (including 4 catch trials) be used to obtain a measurement for general procedures. However, it is unclear whether it is necessary to use 24 trials in a clinical procedure because reasonable threshold estimates have been obtained with only 12 trials (i.e., Green, 1993; Gu and Green, 1994; Florentine et al, 2000). In fact, Florentine et al (2000) proposed a possible procedure for a clinical gap-detection test with only 15 trials (12 test and 3 catch trials). They used a cued maximum-likelihood yes-no procedure that yielded encouraging results. Further support for the idea that long tracks are not necessarily advantageous comes from Green's (1993) observations that the variability obtained in repeated pure-tone threshold measurements in human listeners was larger than that obtained in computer simulations. He concluded that part of the variability obtained in real listeners is likely to reflect day-to-day variability in their thresholds. If listeners' thresholds vary more than the random error caused by the procedure, the repeatability of measurements will benefit little from reducing the variability arising from the psychophysical procedure itself.

Given the time constraints for clinical testing, it is important to determine the minimum number of trials required to obtain stable measurements. However, no systematic investigations of run length have been reported using human listeners. Therefore, the present study compares measurements of narrowband gap detection obtained with two versions of the maximum-likelihood yes-no procedure that differ only in the number of trials used to estimate the gap thresholds.

METHOD

The apparatus and stimuli are identical to those used in our recent studies of gap detection (Florentine et al, 1999, 2000). A detailed description can be found in Florentine et al (2000). An abbreviated description is presented here.

Stimuli

Narrowband noise bursts were presented at 85 dB SPL with center frequencies (CFs) of 1 and 4 kHz. These noises carried the gaps and had a bandwidth equal to three auditory-filter bandwidths (i.e., equivalent rectangular bands [ERBs]) calculated as 0.11 (CF + 165 Hz) (Buus, 1997), with filter slopes essentially infinitely steep. The gap noise bursts were presented with complementary band-stop maskers to prevent listeners from hearing the spectral splatter from the abrupt gating of the gaps. The maskers had a spectrum level 10 dB below that of the gap noiseband. The combined signal noise burst and masker noise was filtered with a nine-ERB filter to keep the spectral splatter inaudible while limiting the bandwidth and loudness of the stimulus. The duration of the noise bursts was 768 msec, and the variable-length gap started 250 msec after the onset of the noise. The noise bursts started and ended with a 20-msec rise and fall. The rise and fall times of the gaps were limited only by the nine-ERB filter. Thus, the rise and fall times specified between 10 percent and 90 percent of the steady-state amplitude were 0.92 msec at 1 kHz and 0.21 msec at 4 kHz. The attenuation of the signal during the gap exceeded 6 dB for a duration equal to the nominal gap duration and exceeded 20 dB for a duration equal to the nominal duration minus the rise and fall times.

The stimuli were digitally synthesized by setting the real and imaginary parts of all frequency components within the passband to random values taken from a Gaussian distribution. All other frequencies were set to zero. The resulting spectrum was then inverse-Fourier transformed into the time domain, and the gap was added to the waveform by setting samples of the noise burst to zero for the duration of the gap. The waveform was Fourier transformed and added to the bandstop masker spectrum, multiplied by the nine-ERB filter function, and then inverse-Fourier transformed and shaped with the 20-msec raised-cosine ramps. A new noise burst was created for each presentation.

Apparatus

Each listener was tested individually in a sound-attenuating booth. A PC-compatible computer with a signal processor (TDT AP2) generated the stimuli, recorded the listeners' responses, and executed the psychophysical procedures. The digitally synthesized noise bursts with gaps were output from a D/A converter (TDT DD1, sample rate = 41.67 kHz) and fed to a low-pass filter (TDT FT5, f_c = 20 kHz, 135 dB/octave). The analog signals were then fed to a programmable attenuator (TDT PA4) and a headphone amplifier (TDT HB6), which presented the stimuli through one earpiece of a Sony MDR-V6 headphone.

Listeners

Eleven listeners participated in this study. None had any history of hearing difficulties, and all had normal, pure-tone audiometric thresholds of less than 15 dB HL with respect to ANSI (1989) standard audiometric thresholds. The listeners were 10 women and 1 man, ranging in age from 19 to 43 years. Ten of the 11 listeners had no previous experience with psychophysical tasks, and no listeners were experienced in gap-detection tasks.

Procedures

The minimum detectable gaps (MDGs) were measured using a method of maximum likelihood (MML) cued yes-no procedure. Each trial had two noise burst intervals separated by 500 msec. The first interval (the "cue") never contained a gap, and the second interval always contained a gap. Catch trials with an inaudible 0.5-msec gap were included to provide an estimate of the false-alarm rate. Listeners responded "gap" if it was heard and "no gap" if it was not. No feedback was provided because the correct answer depended on the listener's perception and could not be defined objectively. Using a cue without a gap makes the yes-no procedure similar to a same-different task, without encouraging any bias in the listener's decision that might result from comparisons of the signal gap to an arbitrarily chosen gap. A complete discussion of the rationale behind this method may be found in Florentine et al (2000).

Each run used to produce an MDG estimate had a total of either 18 or 33 trials (referred to as short or long track length). The first three trials of a run were "warm-up" trials, which had clearly audible gaps (75 msec for the 1-kHz CF, 30 msec for the 4-kHz CF). These trials were not used for the maximum-likelihood estimation of the MDG. In the subsequent 15 or 30 trials, the catch trials were randomly interspersed with variable-gap signal trials with a 20 percent probability. In total, each run had 3 warm-up trials, 3 or 6 catch trials, and 12 or 24 test trials.

Track-length order was counterbalanced across listeners and sessions, and the order of CF was randomly chosen within a track length. One testing session lasted about 45 minutes, and each session repeated the complete set of conditions three times. Listeners could take breaks at any time and were required to take a break midway through a session. Listeners were tested in two sessions separated by about a week. Thus, there were six MDG estimates for each of four conditions (1- or 4-kHz CF, 18 or 33 trials) taken over two sessions. A run was rejected if the false-alarm rate exceeded 25 percent, and that condition was rerun at the end of the session. Due to experimenter error, two runs (of

264) exceeded 25 percent but were included in the data; their inclusion did not change the subsequent analysis. Only 7 of 264 runs exceeded a 20 percent false-alarm rate, and a total of 26 exceeded a 10 percent false-alarm rate.

The MML procedure used to adjust gap duration during a run and then estimate an MDG has been extensively described in another article from this laboratory (Florentine et al, 2000) and is similar to that described by Green (1993). Briefly, for a given run, a set of possible psychometric functions is computed with a logistic function (Florentine et al, 1999) of the form

$$P_{yes} = a + \frac{1-a}{1 + \exp(-k[\log\langle t\rangle - \log\langle m\rangle])}, \qquad (1)$$

where a is the false-alarm rate, t is the gap duration, m is the midpoint of the psychometric function, and k is a free parameter that determines the slope (k = 11 at 1 kHz, k = 16 at 4 kHz, as determined by Florentine et al [1999]). The MML procedure estimates threshold as the midpoint, m, of the most probable function among 135 candidate psychometric functions that vary in midpoint and false-alarm rate.

Data Analysis

The statistical analysis was performed on log-transformed MDGs because the distribution of MDGs is approximately normal on a logarithmic scale. In addition, the psychometric function maintains its shape when plotted as a function of the logarithm of the gap duration (Green and Forrest, 1989; Florentine et al, 1999), and the MML's step size is on a log scale.

RESULTS

Figure 1 shows mean MDGs for the two track lengths and two CFs across six repetitions of a condition. The mean MDGs at 1 kHz (13.34 msec) are greater than those at 4 kHz (7.12 msec) and are in quantitative agreement with previous studies from this laboratory (e.g., Florentine and Buus, 1983; Buus and Florentine, 1985; Florentine et al, 1999). The results are also in qualitative agreement with other studies (e.g., Fitzgibbons, 1983; Shailer and Moore, 1983; Fitzgibbons and Gordon-Salant, 1987; Snell and Frisina, 2000), although the present thresholds may be somewhat longer due to differences in stimuli and/or threshold criteria. Results show no difference in MDGs between the short and long track lengths within a CF (1 kHz: short = 12.98 msec, long = 13.71 msec; 4 kHz: short = 6.98 msec, long = 7.35 msec). Additionally, there appears to be no systematic difference in variance between the two track lengths, even though the long condition has twice as many test trials as the short condition.

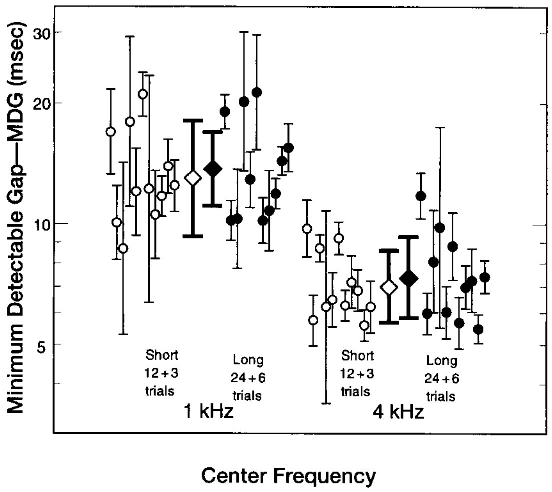

Figure 1 Overall MDGs (diamonds) and the 11 individual subjects' MDGs (circles) are shown for 1- and 4-kHz CFs. Short (unfilled symbols) and long (filled symbols) track lengths are also shown. The error bars on the individuals' MDGs are standard deviations of the logarithmic MDGs. The error bars on the overall mean MDGs are the mean standard deviations across subjects for that condition. The order of individuals is the same from left to right across conditions.

Although most of the listeners were inexperienced, learning across the six repetitions for a condition does not appear to contribute to the variance because no systematic effect of repetition on the MDGs is present. A four-way analysis of covariance for repeated measures (track length [two levels: long or short] × CF [two levels: 1 and 4 kHz] × repetition [continuous] × listener [random factor: 11 levels]) on the logarithms of the MDGs confirmed that repetition was not a significant factor and that there was no learning effect apparent in the data. Therefore, subsequent analyses were pooled across repetitions. As expected, MDGs differed significantly between the 1- and 4-kHz CFs (F [1, 10] = 50.2, $p < .0001$). However, there was no significant effect of track length (F [1, 10] = 0.03, $p = .87$) nor any significant interaction between track length and frequency (F [5, 50] = 0.005, $p = .95$).

Of particular interest to this study was the reliability of MDG estimates for different track lengths. Therefore, variances were computed from the logarithms of the MDGs for each level of track length and CF within listener. A three-way analysis of variance (track length [two levels: short and long] × CF [two levels: 1 and 4 kHz] × listener [random factor: 11 levels]) of the logarithms of the MDGs revealed neither significant main effects of track length or CF nor any significant interactions. That is, there was no significant difference in the variance of the different conditions. However, inspection of the data revealed a trend for variance to be greater in the short track length at 1-kHz CF as is apparent in Figure 1, but it did not reach significance. Thus, the present data indicate that little or no benefit for the test–retest reliability is obtained by increasing the number of trials from 15 to 30 in the MML yes-no procedure.

DISCUSSION

Given the large decrease in the variability of thresholds obtained by Green (1993) when the number of trials was increased from 12 to 24 in simulations of the MML procedure, the empirical finding of little or no benefit of increasing the number of trials is somewhat surprising. To examine whether the difference between Green's simulation results and the present data might be due to the difference in the task (pure-tone thresholds vs gap detection) and the concomitant difference in the listeners' psychometric functions, a simulation of the MML procedure for gap detection was performed. As shown in Figure 2, the variability is larger for the 1-kHz CF than for the 4-kHz CF and increases as the false-alarm rate increases. This holds whether the MDGs are estimated from short or long tracks. Increasing the number of signal trials from 12 (plus 3 catch trials) to 24 (plus 6 catch trials) roughly halves the standard deviations. The decrease is somewhat smaller for the 0 percent false-alarm rate and somewhat larger for greater false-alarm rates. Overall, the present simulations follow the pattern established by Green (1993). These findings show that the small effect of increasing the number of trials used to estimate MDGs in real listeners is not due to the different task and different underlying psychometric functions in the present experiment.

Why, then, does the variability of MDGs in real listeners not decrease appreciably when the number of trials is increased? One possible answer may be derived from Green's (1993) observation that "some of the variability inherent in testing over several days actually reflects changes in threshold over time." Clearly, a sufficiently large variation in the listeners' "true" MDGs could swamp any reduction in variability that should be obtained by increasing the number of trials. Indeed, the standard deviations obtained in the simulations generally are well below those obtained from the real listeners. Given that the false-alarm-rate estimates for the listeners averaged about 0.03, the simulations indicate that the variability inherent in the MML procedure should produce standard deviations ranging from about 8 percent for 24 + 6 trials at the 4-kHz CF to about 22 percent for 12 + 3 trials at the 1-kHz CF. (Because the listeners' actual false-alarm rates are likely to be about twice that estimated by the MML procedure [Gu and Green, 1994], these standard deviations are those obtained for a simulated false-alarm rate of 0.05.) Thus, it appears that variability from a source other than the psychophysical procedure contributes substantially to the overall variability obtained for real listeners. In fact, Figure 3 shows that the variability obtained in the real listeners is well accounted for if it is assumed that a constant variance corresponding to a standard deviation of 22 percent adds independently to the variance inherent in the MML procedure. This finding indicates that the failure to obtain significantly reduced variability by increasing the number of trials is likely to be caused by variations of the

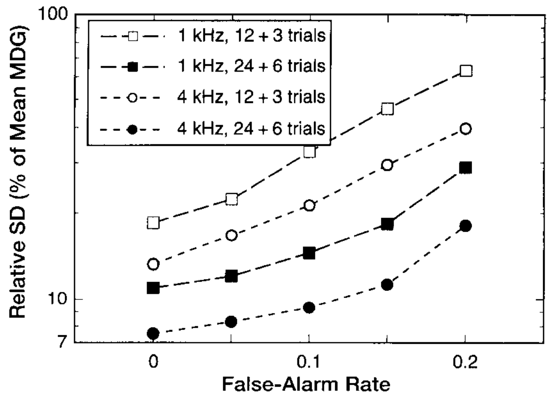

Figure 2 Variability of MDGs in simulated listeners with stationary psychometric functions. The standard deviation, expressed as a percentage of the MDG, is plotted as a function of the false-alarm rate. Data are shown for psychometric functions corresponding to center frequencies of 1 kHz (squares) and 4 kHz (circles). The unfilled symbols show data obtained with 12 signal trials and 3 catch trials per run; the filled symbols show data obtained with 24 signal trials and 6 catch trials per run. Each data point is based on 10,000 simulated runs of the MML procedure.

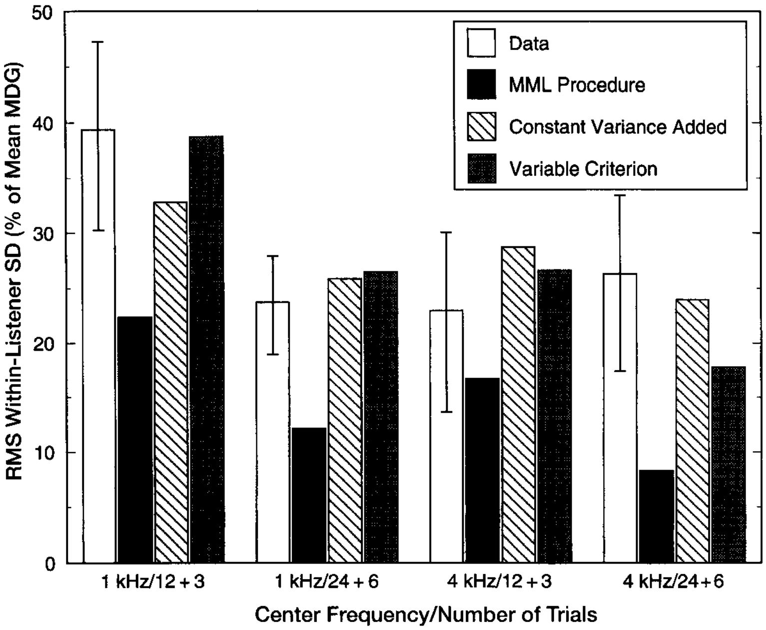

Figure 3 Comparison of standard deviations, expressed as percentage of the MDG, for real and simulated listeners. Each group of bars shows data for a different frequency and track length. The white bars show the standard deviations obtained in the real listeners; the error bars indicate the range encompassed by plus and minus one standard error of the mean variance. The black bars show the standard deviations obtained by simulating the MML procedure for a listener with a stationary psychometric functions and a false-alarm rate of 0.05; these data are replotted from Figure 2. The striped bars show the standard deviation obtained by adding a constant variance to that due to the procedure. The gray bars show the standard deviations obtained by simulating the MML procedure with listeners who change their response criterion randomly from run to run.

listeners' MDGs over time. Note that these variations may occur over relatively small intervals of time as evidenced by the current data showing within-listener standard deviations that are only slightly larger across than within days. This finding contrasts with Green's (1993) finding of a substantially larger variability of absolute threshold across than within days. Whether the different relationships between across- and within-day variability obtained in the two studies is due to task or schedule of measurements (i.e., two measurements on each of 5 days for Green and three measurements on each of 2 days in the present study) is unclear. However, both studies clearly show that thresholds vary over time.

One reason for listeners' MDGs varying over time may be that their criterion for responding that the gap was heard varies over time. As the criterion increases, the number of "gap" responses to both signal and catch trials decreases. The MML procedure makes some adjustment for this effect by using the midpoint of the psychometric functions to estimate the MDG (i.e., the MDG corresponds to 50% "gap" responses if the false-alarm rate is 0% but to 60% "gap" responses if the false-alarm rate is 20%). However, using signal-detection theory (SDT), it is easy to show that this adjustment is likely to be insufficient to maintain a threshold that is independent of the listener's response criterion. SDT holds that the listener's response on a given trial is determined by the value of an internal variable, the decision variable, which varies randomly around a mean value that is determined by the stimulus. If the decision variable on a given trial exceeds some response criterion set by the listener, the response will be "gap"; otherwise, it will be "no gap." If it is assumed that the decision variables derived from a given stimulus are normally distributed and have the same variance for all stimuli, the listeners' performance can be characterized by the distance between the means of the distributions for stimuli without a gap (catch trials) and for stimuli with a gap of some fixed duration. This distance is the sensitivity (d′), which is measured in units of the standard deviation of the normal distribution. Because d′ does not depend on the listener's criterion, it provides a measure of a listener's gap-detection ability that should be unaffected by the response criterion. (For a thorough introduction to SDT, see Green and Swets [1966] and Macmillan and Creelman [1991].)

These simple considerations permit the effect of the listeners' response criterion on their psychometric functions for gap detection to be investigated. To this end, it may be assumed that the relation between d′ and gap duration is fixed for a given listener and CF and corresponds to the psychometric function obtained by Florentine et al (1999). Using typical parameters for the psychometric function at 4 kHz (i.e., false-alarm rate = 0.1 and MDG = 6 msec), one can calculate d′ as a function of gap duration as shown in the inset of Figure 4. Given this function, the probability of a "gap" response can be

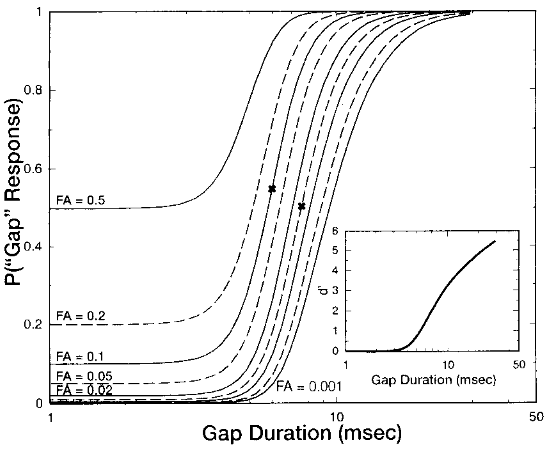

Figure 4 The effect of response criterion on the psychometric functions at the 4-kHz CF. The proportion of "gap" responses is plotted as a function of gap duration. As indicated by the labels, the parameter is the false-alarm rate, which varies from 0.001 for the right-most psychometric function to 0.5 for the left-most psychometric function. The inset shows the relationship between gap duration and the signal-detection sensitivity, d′, used to derive all of the psychometric functions. It is derived from the psychometric function measured by Florentine et al (1999), which had a false-alarm rate of 0.1. The crosses give an example of shifts in the midpoint of the psychometric function because of changes in the false-alarm rate (see text).

calculated for any gap duration and false-alarm rate. As shown in the main graph of Figure 4, the resulting psychometric functions show that the midpoint, which is used as an estimate of the MDG by the MML procedure, shifts systematically as the false-alarm rate changes. The lower the false-alarm rate, the larger the MDG. The change in MDG is far from negligible. For example, as the false-alarm rate changes from 0.1 to 0.01, the MDG increases from 6 to about 7.5 msec, or about 25 percent as indicated by the crosses in Figure 4. This shows that criterion changes well within the range of criteria that listeners may adopt can lead to a variability equal to or even greater than the variability imposed by the MML procedure. Thus, it seems quite possible that the reason the variability in real listeners exceeds that obtained in simulations of listeners with stationary psychometric functions is that the real listeners change their response criteria over time.

To examine how variability in the response criterion may affect the MDGs at various center frequencies, an analytical expression for the MDG as a function of response criterion is useful. Given that the MDG is defined as the midpoint of the logistic psychometric function, it can be shown that the logarithm of the MDG, M, obtained with a given response criterion, C, is equal to

$$M = M_o - \frac{\ln\left(\frac{1 - F(M_o - C_o + C)}{F(M_o - C_o + C) + F(C) - 1}\right)}{k}, \quad (2)$$

where M_0 is the logarithm of the MDG (6 msec in Fig. 4) obtained with the reference response criterion, C_0, $F(\cdot)$ is the cumulative normal function, and k is the slope parameter for the logistic function describing the psychometric function. (The reference criterion in Figure 4 corresponds to a false-alarm rate of 0.1—i.e., $0.1 = 1 - F[C_0]$.) Equation 2 shows that the relative variability of the MDG caused by a criterion variability of some fixed magnitude increases as the slope decreases. If variability of the listeners' response criterion is the reason for the excess variability of real listeners' MDGs, one would expect its effect to be relatively larger at the 1-kHz CF than at the 4-kHz CF in comparison with the simple model with a constant added variance discussed above.

Although the variability of the measured within-listener standard deviations is too large to allow a clear distinction between the simple addition of a constant variance and the effect of a constant variability in the listeners' response criteria, it is noteworthy that simulations of the MML procedure for listeners who vary the response criterion randomly from run to run also provide a good fit to the data. If the response criterion (C) is assumed to be normally distributed with a mean of 1.8 (which yields an average estimated false-alarm rate of about 0.03) and a standard deviation of 0.47, the standard deviations of the simulated MDGs appear quite similar to those of the real listeners' data, as shown in Figure 3. Therefore, random variability of the listeners' response criterion seems to be a compelling explanation for the finding that the variability of real listeners' MDGs is considerably larger than the variability imposed by the procedure and decreases relatively little when the number of trials is increased from 15 to 30.

The implications are clear: If listeners change their response criteria over time, adding trials beyond the minimum number required to obtain a stable threshold estimate offers little or no advantage for the test–retest reliability of a clinical threshold measurement. If psychoacoustic measurements require a precision greater than that obtainable with this minimum number of trials, simply adding trials is unlikely to be a sufficient remedy. Time may be better spent on controlling the listeners' response criteria or compensating for their changes, although the latter requires a reasonably precise assessment of response criteria. It should be noted that determining the commonly obtained low false-alarm rates with sufficient precision to reduce appreciably the variability of the estimated MDGs is likely to require many more trials than is feasible for a clinical test. Until methods to control the response criterion in a clinical setting are developed, the usefulness of psychoacoustic tests will have to be evaluated within the context of the limited precision that is ultimately imposed by the variation of the listener's threshold over time.

Acknowledgment. This work was supported by T grant no. NIH/NIDCDR01 DC00187.

REFERENCES

American National Standards Institute. (1989). *Specifications for Audiometers.* (Vol. ANSI S3.6-1989). New York: ANSI.

Buus S. (1997). Auditory masking. In: Crocker MJ, ed. *Encyclopedia of Acoustics.* Vol. 3. New York: Wiley, 1427–1445.

Buus S, Florentine M. (1985). Gap detection in normal and impaired listeners: the effect of level and frequency. In: Michelsen A, ed. *Time Resolution of Auditory Systems.* London: Springer, 159–179.

Buus S, Florentine M, Mason CR. (1986). Psychoacoustical tuning curves and absolute thresholds at high frequencies. In: Moore BCJ, Patterson RD, eds. *Auditory Frequency Selectivity.* New York: Plenum, 344–350.

Fitzgibbons PJ. (1983). Temporal gap detection in noise as a function of frequency, bandwidth and level. *J Acoust Soc Am* 74:67–72.

Fitzgibbons PJ, Gordon-Salant S. (1987). Temporal gap resolution in listeners with high frequency sensorineural hearing loss. *J Acoust Soc Am* 81:133–137.

Florentine M, Buus S. (1983). Temporal resolution as a function of level and frequency. *Proc 11th Int Congr Acoust* 3:103–106.

Florentine M, Buus S. (2000, February). *Temporal-resolution deficits can vary with frequency in cochlear hearing losses.* Presented at the Mid-Winter Meeting of the Association for Research in Otolaryngology, St. Petersburg, FL.

Florentine M, Buus S, Geng W. (1999). Psychometric functions for gap detection in a yes-no procedure. *J Acoust Soc Am* 106:3512–3520.

Florentine M, Buus S, Geng W. (2000). Toward a clinical procedure for narrowband gap detection I: a psychophysical procedure. *Audiology* 39:161–167.

Florentine M, Buus S, Hellman RP. (1997). A model of loudness summation applied to high-frequency hearing loss. In: Jesteadt W, ed. *Modeling Sensorineural Hearing Loss.* Mahwah, NJ: Erlbaum, 187–198.

Florentine M, Houtsma AJM. (1983). Tuning curves and pitch matches in a listener with a unilateral, low-frequency hearing loss. *J Acoust Soc Am* 73:961–965.

Formby C, Sherlock LP, Green DM. (1996). Evaluation of a maximum-likelihood procedure for measuring pure-tone thresholds under computer control. *J Am Acad Audiol* 7:125–129.

Green DM. (1993). A maximum-likelihood procedure for estimating thresholds in a yes-no task. *J Acoust Soc Am* 93:2096–2105.

Green DM, Forrest TG. (1989). Temporal gaps in noise and sinusoids. *J Acoust Soc Am* 86:961–970.

Green DM, Swets JA. (1966). *Signal Detection Theory and Psychophysics.* New York: Wiley.

Gu X, Green DM. (1994). Further studies of a maximum-likelihood yes-no procedure. *J Acoust Soc Am* 96:93–101.

Hellman RP. (1994). Relation between the growth of loudness and high-frequency excitation. *J Acoust Soc Am* 96:2655–2663.

Hogan CA, Turner CW. (1998). High-frequency audibility: benefits for hearing-impaired listeners. *J Acoust Soc Am* 104:432–441.

Leek MR, Dubno J, He N-J, Ahlstrom JB. (2000). Experience with a yes-no single interval maximum-likelihood procedure. *J Acoust Soc Am* 107:2674–2684.

Macmillan NA, Creelman CD. (1991). *Detection Theory: A User's Guide.* Cambridge, UK: Cambridge University Press.

Moore BCJ, Huss M, Vickers DA, Glasberg BR, Alcantara JI. (2000). A test for the diagnosis of dead regions in the cochlea. *Br J Audiol* 34:204–225.

Shailer MJ, Moore BCJ. (1983). Gap detection as a function of frequency, bandwidth and level. *J Acoust Soc Am* 74:467–473.

Snell KB, Frisina DR. (2000). Relationships among age-related differences in gap detection and word recognition. *J Acoust Soc Am* 107:1615–1626.

Thornton AR, Abbas PJ. (1980). Low-frequency hearing loss: perception of filtered speech, psychophysical tuning curves, and masking. *J Acoust Soc Am* 67:638–643.

Turner CW, Burns EM, Nelson DA. (1983). Pure tone pitch perception and low-frequency hearing loss. *J Acoust Soc Am* 73:966–975.

Turner CW, Cummings KJ. (1999). Speech audibility for listeners with high-frequency hearing. *Am J Audiol* 8:47–56.

Villchur E. (1973). Signal processing to improve speech intelligibility in perceptive deafness. *J Acoust Soc Am* 53:1646–1657.