Stimulus selection in adaptive psychophysical procedures

David M. Green

Citation: The Journal of the Acoustical Society of America **87**, 2662 (1990); doi: 10.1121/1.399058
View online: https://doi.org/10.1121/1.399058
View Table of Contents: https://asa.scitation.org/toc/jas/87/6
Published by the Acoustical Society of America

Stimulus selection in adaptive psychophysical procedures

David M. Green
Psychoacoustics Laboratory, Psychology Department, University of Florida, Gainesville, Florida 32611

(Received 28 September 1989; accepted for publication 10 January 1990)

In adaptive psychophysical procedures, the stimulus should be presented at a relatively high level rather than near the middle of the psychometric function, which is often defined as the "threshold" value. For some psychometric functions, the optimal stimulus placement level produces 84% to 94% correct responses in a two-alternative forced-choice task. This result is disquieting because the popular two-down one-up rule tracks a relatively low percentage of correct responses, 70.7%. Computer simulations and a variety of psychometric functions were used to confirm the validity of this analysis. These simulations also demonstrate that the precise form of the psychometric function is not critical in achieving the high efficiencies. Finally, data from human listeners indicate that the standard deviation of threshold estimates is indeed larger when the stimulus presented on each trial is at a stimulus level corresponding to 70.7% rather than 94% correct responses.

PACS numbers: 43.66.Yw [WAY]

INTRODUCTION

This paper compares different adaptive psychophysical procedures. Special emphasis is given to the question of how the selection of the stimulus value used on each trial of the adaptive procedure affects the variability of the threshold estimate. Recent work by Laming and Marsh (1988), as well as Watson and Pelli (1983), using the maximum-likelihood procedure has suggested that the selection of a particular stimulus value (called the sweetpoint) should be at a relatively high stimulus level (SL). In fact, for a simple Gaussian psychometric function, the sweetpoint corresponds to an error rate of only 6% in a two-alternative forced-choice task. This is very different from the error rate maintained by most up–down procedures. The widely used procedure employing a two-down one-up rule yields an error rate of about 30%.

We investigated how the selection of a stimulus value influences the threshold estimate for a variety of psychometric functions, using computer simulations to assess the effects of a mismatch between the psychometric functions actually used by the observer and the psychometric functions assumed by the experimenter. We found, in general, that a mismatch does not seriously compromise the effectiveness of the maximum-likelihood procedures. We also used simulations to investigate how changes in error rates generated by selecting different up–down rules affect the standard deviations of threshold estimates using these techniques. These simulations revealed that variability in the threshold estimates was improved if the equilibrium point of the up–down procedure was adjusted to match the sweetpoint of the observer's psychometric function. Finally, we compared the variability of threshold estimates obtained with human observers for experimental conditions that, according to this theory, should have produced quite different variabilities in threshold estimates. These results were in general agreement with the theoretical analyses. The paper begins with an explanation of the maximum-likelihood procedure and how the sweetpoint is determined.

I. MAXIMUM-LIKELIHOOD PROCEDURE

A. Threshold estimate

To understand how a maximum-likelihood estimate is used to determine a threshold value, consider the following example. Suppose there are three functions that are possible candidates for the true psychometric function (i.e., the actual psychometric function for the observer under these stimulus conditions) for a two-alternative forced-choice task. Call the three functions $A(x)$, $B(x)$, and $C(x)$ and assume they are as shown in Fig. 1. For each stimulus value, x_i, we have some probability of being correct, which is given by $A(x_i)$, $B(x_i)$, or $C(x_i)$. Often, we are not interested in the entire psychometric function but only in an arbitrary point along the function called the threshold value.

Table I summarizes some typical data collected on n trials. On each trial, we present a stimulus value x and record whether the response was correct, denoted C, or incorrect, denoted I. The probability of a *correct* response on trial i,

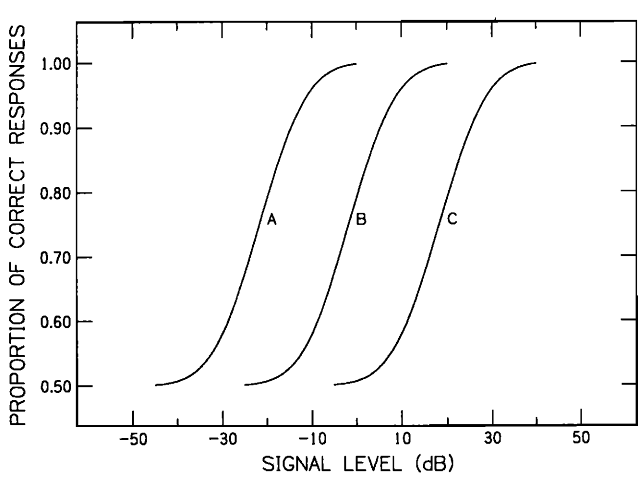

FIG. 1. Three psychometric functions showing the probability of a correct response in a two-alternative forced-choice task as a function of the stimulus level.

 0001-4966/90/062662-13$00.80

TABLE I. Probabilities associated with the different hypothesis after n trials.

Trial No.	Stimulus	Response	Probability for each hypothesis A	B	C
1	x_1	C	$A(x_1)$	$B(x_1)$	$C(x_1)$
2	x_2	I	$1-A(x_2)$	$1-B(x_2)$	$1-C(x_2)$
i	x_i	C	$A(x_i)$	$B(x_i)$	$C(x_i)$
n	x_n	I	$1-A(x_n)$	$1-B(x_n)$	$1-C(x_n)$

given that hypothesis, is $A(x_i)$. The probability of an *incorrect* response on trial i, given that hypothesis, is the complementary probability, $1 - A(x_i)$. In the columns to the right of the table, we entered the probability associated with each of the three hypotheses on every trial.

To calculate the likelihood of these three hypotheses, we must compute the total probability associated with each, given the data of these n trials. That probability for hypothesis A is

$$\text{probability of } A = \prod_{i=1}^{n} A^{c}(x_i)\,[1 - A^{I}(x_i)], \tag{1}$$

where the exponent c is equal to one if the response on trial i is correct and zero otherwise, and the exponent I is equal to one if the response on trial i is incorrect and zero otherwise. There are similar expressions for hypotheses B and C. One of these probabilities must be the largest. That hypothesis, A, B, or C, is the psychometric function that has the greatest likelihood of being correct, given the data contained in the table. This is the essence of the maximum-likelihood procedure.

Of course, it is extremely limiting to have only three hypotheses. In practice, one often entertains many hypotheses about the correct psychometric function. Often, as is true of our example, the functions are all the same except for a single parameter, call it τ, which shifts the function along the stimulus axis. (It is a scale factor in terms of the physical energy of the stimulus, but, because of the logarithmic scale used as the abscissa, changes in τ can be interpreted as lateral shifts in the psychometric function.) We might assume that all the functions are of the form $F(x - \tau)$, where the value of τ can take on many values, say m. One must then calculate m such probabilities. No matter how many hypotheses are considered, one will produce a maximum probability and thereby provide a maximum-likelihood estimate of the true psychometric function, given the observed data. In practice, one simply selects a relatively fine grid size for the stimuli, that is, a value of τ that corresponds to a relatively small stimulus change, say 1/2 or 1 dB. The psychometric function that has the largest likelihood given this grid size is then the goal of the maximum-likelihood procedure.

The threshold value is some arbitrary point on the psychometric function. For our three functions shown in Fig. 1, the reader will note that, for a proportion of about 0.95, the stimulus values are -10 dB for function A, 10 dB for function B, and 20 dB for function C. We could, therefore, call the point where the proportion correct is 0.95 the "threshold" value. A more conventional definition of the threshold value is the stimulus level corresponding to a 0.75 probability of being correct in two-alternative forced-choice tasks—we will call this the conventional threshold value in this paper. In the case of up–down procedures, the threshold is often defined as the equilibrium point in the adaptive procedure (Levitt, 1971). Thus the threshold value of a two-down one-up procedure is the stimulus level corresponding to a proportion of 0.707. If a three-down one-up procedure is used, then the threshold is the stimulus level corresponding to a proportion of 0.794. One seldom worries about the exact decibel difference produced by these two different procedures. Most experiments are designed to reveal trading relations, given a fixed threshold value, for example, how stimulus duration and intensity vary for a fixed threshold value. The exact proportion used to define the threshold value is then ignored.

B. Optimum stimulus placement

The preceding discussion has established a way to determine the most likely psychometric function or (equivalently) the threshold value of the stimulus after n trials. But, given this knowledge, where should we place the stimulus on the next trial, that is, how should we choose the value of x_i that is used on each trial of the test? A common response is to assert that we should place the stimulus at the threshold value (Simpson, 1989), that is the conventional definition of the threshold, a stimulus value near the middle of the psychometric function. While it is clear that we should avoid a very high or very low stimulus value, our real objective is to minimize the variance associated with these estimates. Let us explore the variability of these estimates in more detail.

Given a single threshold estimate x_0, we will have a range of possible threshold values and a standard deviation associated with our estimate, call it σ. It makes sense to select the stimulus on the next trial to minimize this standard deviation. What may not be obvious is that, in fact, we have enough information to achieve this objective. We can actually minimize the variability of our threshold estimate σ given our assumptions concerning the psychometric function.

To see that this is the case, consider the following. Once we have identified the most likely psychometric function, we know the probability associated with every stimulus value. The variability of each probability estimate p can be estimated because it is a binomial variable; hence, the variance of the probability is proportional to $[p(1-p)]$. In our example, if $A(x-\tau)$ were the most likely hypothesis and n trials were used in the estimate, the variance along the y axis is equal to $\{A(x-\tau)*[1 - A(x-\tau)]/n\}$. To a good first approximation, the variability along the probability scale divided by the variability along the stimulus scale is equal to the slope of the psychometric function. This argument is illustrated in Fig. 2, where we have plotted a psychometric function and illustrated the binomial variability along the y axis for several different probability levels. Thus the variance of the stimulus estimate σ^2 is equal to the binomial variance divided by the slope of the psychometric function squared; that is, because $\sigma_y/\sigma_x = dF/dx$, we have

$$\sigma^2 = [p(1-p)] \quad \left(\frac{dF}{dx}\right)^2, \tag{2}$$

where dF/dx is the slope or derivative of the psychometric

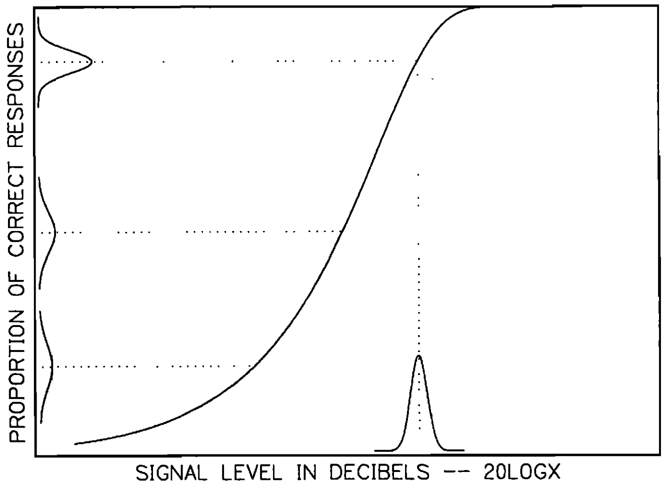

FIG. 2. The relation between the variability in the threshold estimates, shown along abscissa, and the variability in the proportion of correct responses, shown along the ordinate, and their relation to the slope of the psychometric function.

function. Technically, we are assuming that the psychometric function can be approximated over a small region by a linear function, the slope of the function is derivative and, if $y = ax + b$, the variance of y is then equal to a^2 times the variance of x.

Rather than simply select the next stimulus from some arbitrary point along the psychometric function, a more enlightened policy is to choose the next stimulus value so that σ is minimized; that is, minimize the variability in the estimated threshold on each successive trial. We call this a *stimulus selection policy*. The stimulus value chosen on each trial need not be some arbitrarily defined threshold value. As we will see, for most psychometric functions the stimulus value that minimizes the variability of the threshold estimate is *not* near the middle of the psychometric function. For many psychometric functions, a surprising fact is that the sweetpoints are located at stimulus values associated with a relatively high probability of being correct, as illustrated in Fig. 2. For example, some reasonable assumptions about the form of the psychometric functions lead to sweetpoints where the probability of a correct response is in the 0.80–0.94 range.

The stimulus value that minimizes the variability of the threshold estimate has been called the sweat factor by Taylor and Creelman (1967). Our use of this quantity is more similar to the treatment of Laming and Marsh (1988), who calculate the stimulus value that maximizes the inverse of our Eq. (2) and call that quantity the sweetpoint of the psychometric function.

II. MISMATCH BETWEEN ASSUMED AND ACTUAL PSYCHOMETRIC FUNCTIONS

While the preceding theoretical analysis is interesting, its empirical impact rests with the sensitivity of the theory to any errors made about the assumed form of the psychometric functions. As experimenters, we do not know the exact form of the psychometric function. Therefore, we must assume a psychometric function to exercise the maximum-likelihood procedure and to calculate the sweetpoint to be used in the stimulus selection procedure. Regrettably, we know relatively little about the exact form of the psychometric function for most psychophysical tasks. A psychometric function is determined by measuring the proportion of correct judgments at several different stimulus levels. Of course, these proportions are subject to error. The preceding analysis, which requires knowledge of the derivative of the psychometric function, requires that these proportions be measured at a large number of stimulus values and with high accuracy. Obtaining such detailed knowledge is, for the most part, simply an impossible experimental burden. In most cases, our knowledge of the form of the psychometric function amounts to little more than the range of the psychometric function. Typically, there are several equations that are plausible candidates, and each provides a reasonably accurate account of the data. Such uncertainty about the exact form of the psychometric function raises questions about the practical usefulness of the maximum-likelihood approach. Suppose the experimenter is mistaken about the psychometric function and uses the wrong function in the maximum-likelihood calculations. Will such an error between the assumed psychometric function and the real one seriously impair the effectiveness of the procedure?

A. Mismatches between similar psychometric functions

Our first priority, then, is to determine how any error made in the experimenter's assumptions about the form of the psychometric function affects the variability of the threshold estimates. To accomplish this investigation, we selected four different equations to represent the possible psychometric functions. Each of these functions is used as the psychometric function for a simulated "observer" in a psychophysical experiment. We repeatedly determined the threshold value obtained with this observer and measured the variability of such estimates over repeated trials. In separate simulations, we used each of the four functions as the assumed function in the maximum-likelihood procedure, what we call the "assumed" psychometric function, for each observer's psychometric function. Each of these assumed functions had a different stimulus placement policy. For each observer, one of the placement policies was correct, in which case the analysis psychometric function was the same as the observer's. The other three simulations were conducted with a mismatch between the analysis psychometric function and the observer's psychometric function.

The four functions chosen in this analysis were two Gaussian and two logistic probability densities. They are listed in Table II along with the specific parameter values. These values were chosen to make the psychometric functions appear similar when plotted on the same decibel scale (see Fig. 3). No attempt has been made to minimize the deviations among these functions. The specific parameter values of the functions were simply chosen to adjust the function to have roughly the same range and shape as determined visually. All four functions are plausible candidates for the psychometric function in intensity discrimination tasks. The first two assume that the argument of the function

TABLE II. Four psychometric functions. The ordinate is a proportion, $P(c)$. The abscissa is in decibels, SL. $\phi(z)$ is the Gaussian unit normal density $\phi(z) = (1/2\pi)^{1/2}\exp(-z^2/2)$.

Function 1	
$P(c) = \int_0^{x/\sqrt{2}} \phi(z)dz$	$\mathrm{SL} = 20 \log (x)$
$P(c) = 0.75$ when $\mathrm{SL} = -0.5$	sweetpoint = 7 dB
Function 2	
$P(c) = \dfrac{1}{1+\exp(-1.2x)}$	$\mathrm{SL} = 20 \log (x)$
$P(c) = 0.75$ when $\mathrm{SL} = -0.8$	sweetpoint = 6 dB
Function 3	
$P(c) = 1/2 + 1/2 \int_{-\infty}^{0.12x+0.2} \phi(z)dz$	$\mathrm{SL} = x$
$P(c) = 0.75$ when $\mathrm{SL} = -1.5$	sweetpoint = 2 dB
Function 4	
$P(c) = \dfrac{1}{1+\exp(-0.21x-0.3)}$	$\mathrm{SL} = x$
$P(c) = 0.75$ when $\mathrm{SL} = -1.4$	sweetpoint = 1 dB

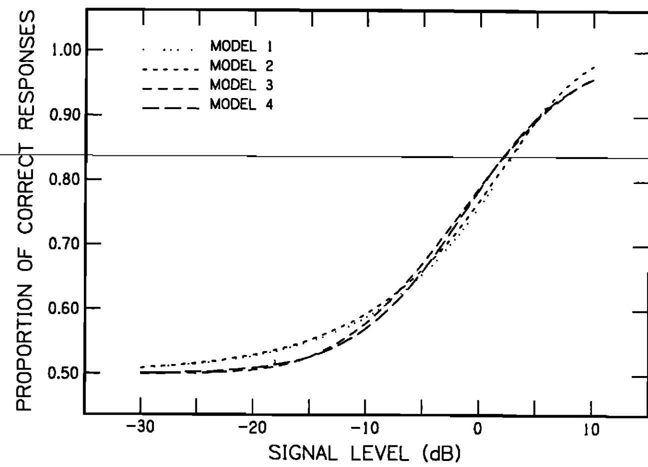

FIG. 3. Four different psychometric functions. Table II gives the equation for each function.

is proportional to the amplitude of the waveform. The second two use the logarithmic transform of the stimulus amplitude as the function's argument and employ a "chance" correction. All of these functions could provide reasonable fits for data obtained when the listener is detecting an increment in the intensity of a sinusoid (Green and Swets, 1988, pp. 195–208). In such cases, the range of the psychometric functions is about 20–25 dB. The general form of this function would also fit data on psychometric functions obtained in profile-analysis tasks (Green, 1988; Raney *et al.*, 1989). Other psychometric functions obtained in intensity discrimination experiments, such as detecting a sinusoidal signal in noise, have a much smaller range, typically about 10–12 dB. Because the variability of threshold estimates is roughly proportional to the range of the psychometric function, the estimates we discuss in this paper should be suitably scaled when interpreting other psychometric functions. For example, if we determine, using our psychometric function with a 20-dB range, that the variability of some threshold estimate is about 3 dB, then you might expect a variability of about 1.5 dB if the psychometric function has a range of 10 dB.

We simulated a two-alternative forced-choice task using all combinations of the four psychometric functions given in Table II. We also systematically varied the placement of the stimulus value on each trial. In different simulations, the stimulus was selected to be at the sweetpoint or at a value as far as 8 dB away from the sweetpoint. The intention was to determine how sensitive the variability in threshold estimates was to changes in the stimulus selection policy. The stimulus threshold was simply the final value of the stimulus after 50 trials. Six hundred such threshold estimates were used to estimate the standard deviation of the thresholds. In all simulations, a 1-dB step size was used, and the initial stimulus value was taken as + 10 dB. Thus the observer is nearly always correct on the first trial (see Fig. 3). Starting the stimulus at a very high level is characteristic of general experimental practice.

Before we present the results, we should comment on the measure we will use to compare the various procedures. We generally use the ratio of the standard deviations of two threshold estimates as our primary means of comparing two different estimation procedures. We are careful to avoid the word "efficiency" in referring to such a ratio because, in its common use, efficiency has the connotation of being a ratio of two energy values or, what is equivalent in this discussion, a ratio of two times needed to complete a given estimate. The quantities that we discuss are ratios of standard deviations and are thus proportional to the square root of the number of trials needed to complete the estimate. The square of our ratios is proportional to the number of trials needed to obtain such an estimate, and such a ratio could indeed be called an efficiency. One could say that the numerical values of our ratios somewhat understate the effects of changes in procedures, and this point should be realized by the reader in evaluating the changes investigated. Nonetheless, most experimentalists are interested in standard deviations of their threshold estimates, and we will base our comparisons on the ratios of these quantities.

Figure 4 displays the obtained estimates of the standard deviation of the measured thresholds for the 16 different conditions. The different columns represent observers using one of the four functions as the psychometric function. The rows represent different selections of the four functions as the experimentally assumed psychometric functions. The figures located along the minor diagonal are those cases where the two psychometric functions are the same. For each of the 16 conditions, the variability of the threshold estimates is plotted as a function of the difference in dB between the sweetpoint of the analysis psychometric function and that used in the simulations. Zero on this scale (indicated by a vertical line) means the stimulus was placed at a value that should minimize the variability in threshold estimates, if the assumed psychometric function were correct. Nonzero values along the abscissa are stimulus placements that were intentionally above or below the sweetpoint of the

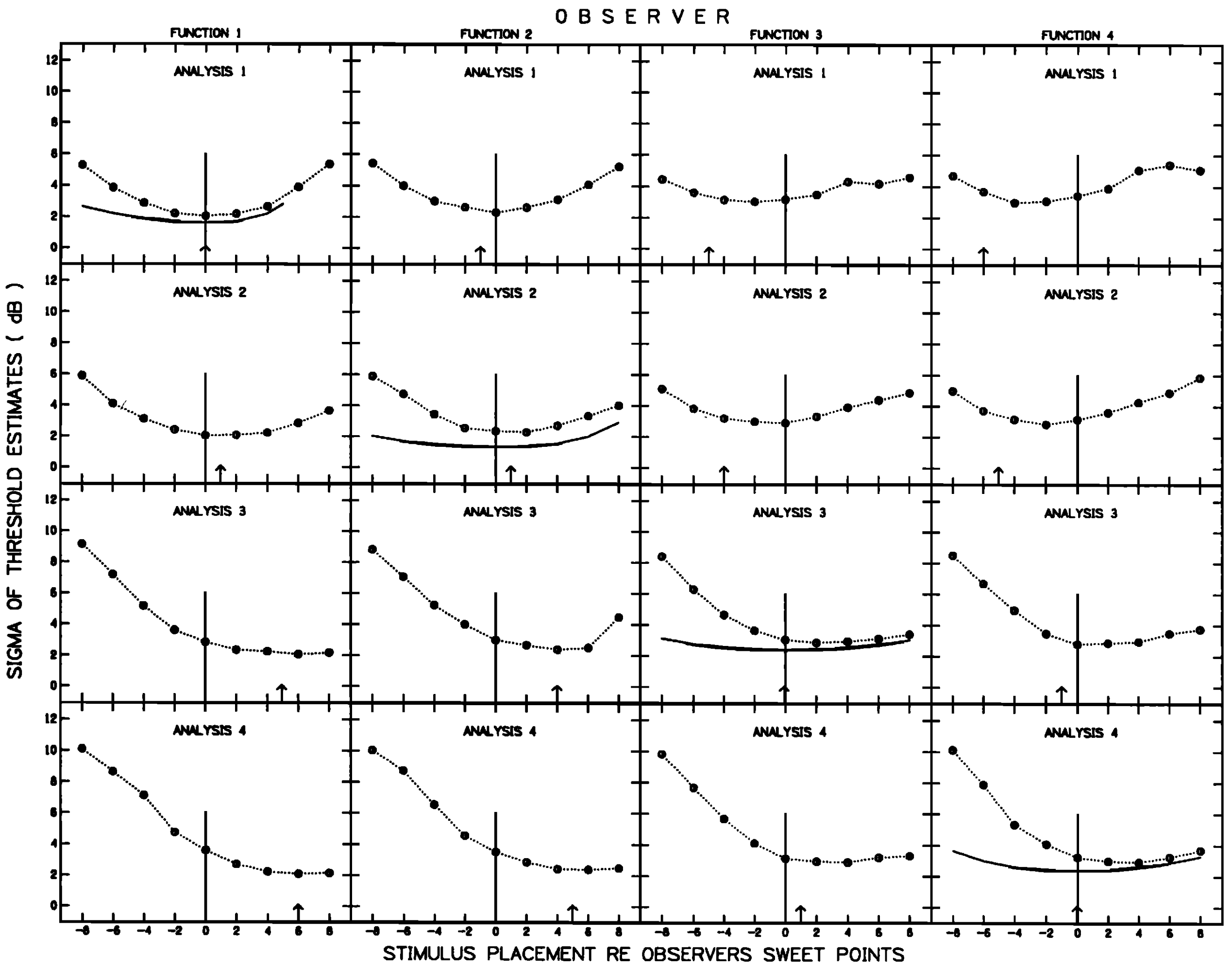

FIG. 4. The standard deviation of 600 threshold estimates as a function of the difference in decibels between where the stimuli were placed and the observer's sweetpoint. The simulations used one of four psychometric functions as the observer (the rows of the figure). The same four functions were used in analyzing each observer's data (the columns of the figure). The solid line drawn in the diagonal panels is the theoretical predictions based on the maximum-likelihood model. The arrow indicates where the stimulus should be placed, given the assumed function for the analysis of the data.

psychometric function assumed in the calculation of the maximum-likelihood procedure. Such values provide information about the effects of incorrect stimulus selection policy. For the first two minor diagonal entries, the minimum in the measured sigmas is very near an abscissa value of zero. This should minimize the variability of threshold estimates, and it apparently does. The minimum value for the sigma does not occur at 0 dB for the combination 3-3 and 4-4. The minimum value, however, is within 10% for the value measured at 0 dB, and we can only attribute this discrepancy to sampling error. In all cases, the changes in variability are not very strong functions of the stimulus selection policy, suggesting that a nonoptimal placement policy is not likely to cause a serious increase in the variability of the resulting threshold estimates.

The small vertical arrow along the abscissa indicates the value of the stimulus placement that should have been used, that is, the stimulus value corresponding to the sweetpoint for the observer's psychometric function. In general, the minimum sigma occurs at or near the value indicated by the arrow for observers 1 and 2, and to a lesser extent for observers 3 and 4. Thus, although only of academic interest, it appears that one could achieve smaller relative variability in threshold estimates if the correct, that is, the observer's, sweetpoint is employed even if the incorrect psychometric function is assumed.

We can summarize the effects of a mismatch between the observer's psychometric function and that used in the analysis of the maximum-likelihood procedure by calculating the ratio of the sigma measured at the zero point on the abscissa (what we would achieve given our assumed psychometric function) to the optimum sigma that could be achieved for that observer (what we could achieve in theory). The optimum sigma is the square root of the value calculated in Eq. (2) divided by 50, because 50 is the number of trials used in the simulation. These ratios are presented in Table III. For observers 1 and 2, the ratio is highest when the correct psychometric function is used in the analysis procedure and is near 0.8. For the third and fourth observers, the ratio is nearly independent of the psychometric function

TABLE III. Sigma measured at sweetpoint for analyzing psychometric function to optimum sigma for observer.

		Observer's psychometric function			
		1	2	3	4
Analysis psychometric function	1	0.78	0.80	0.76	0.68
	2	0.77	0.80	0.83	0.77
	3	0.56	0.63	0.79	0.81
	4	0.48	0.53	0.72	0.72

used in the analysis and is nearly 0.8 for all conditions. The only serious effects of a mismatch occur when data from observers 1 or 2 are analyzed using functions 3 and 4. In that case, the ratio drops to nearly 0.5—a decline from 0.8 of nearly 40%. As a general summary, the effects of a mismatch are not likely to cause a loss of more than 50% in the expected size of the standard deviation of the threshold estimate compared to the optimum possible standard deviation. The flatness of the functions seen in Fig. 4 suggests that using the wrong psychometric function in the maximum-likelihood procedure is not likely to lead to a large decline in relative variability of the threshold estimates.

A final question that needs addressing is why the ratio is as low as 0.8 (see Table III) when the correct psychometric function is used in the analysis. The short answer is that we did not use enough trials in the simulations. We believe this ratio occurs because the maximum-likelihood process had not stabilized on a stimulus value near the sweetpoint. Thus the variability did not reduce the threshold by the square root of the number of trials. This conclusion is based on an additional set of simulations in which we varied the number of trials from 25 to 200 and computed the variability in the threshold estimates after each set of 25 trials. The analysis and the observer both used the same psychometric function, and we ran 500 simulations. Figure 5 shows the sigma for the threshold estimates as a function of the number of trials in the block used to estimate the threshold. The solid line is the theoretical sigma expected when the stimulus value is chosen at the sweetpoint value of the observer, as is the case in this simulation—it is simply the square root of Eq. (2) divided by the number of trials used to estimate the threshold. That line decreases as the reciprocal of the square root of the number of trials, as we would expect from a binomial process with a fixed probability of a correct response on each trial. As can be seen, when few trials are used to estimate the threshold, there is a large distance between the line and points. The ratio between the observed and expected sigmas is only about 0.5 when the threshold is based on 25 trials. As the number of trials increases, this ratio is nearly 0.95 after 200 trials. For 50 trials, the value used in our previous simulation, the ratio is about 0.8.

This trend, displayed in Fig. 5, is understandable if we assume that the stimulus level is being adjusted over a wide stimulus range in the early trials of a block. In that case, the probability of a correct response is not constant and does not represent a binomial process. After about 100 trials, the stimulus values are nearly the same on each successive trial,

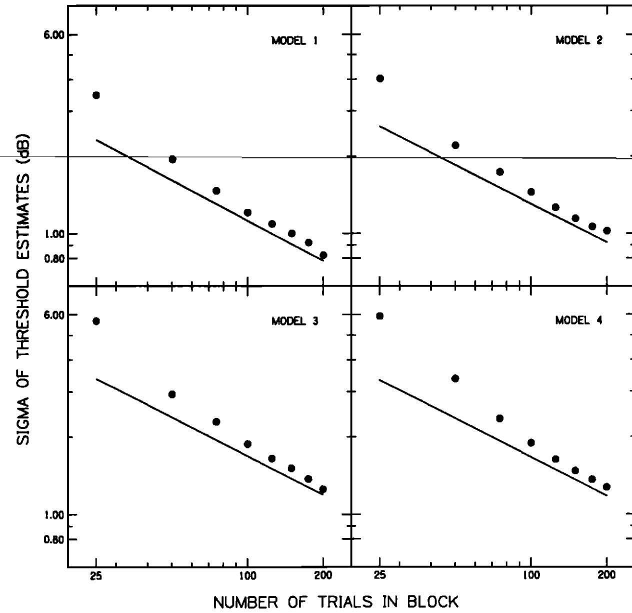

FIG. 5. The standard deviation of 500 maximum-likelihood estimates as a function of the number of trials used to estimate the threshold. The solid line shows the sigma decreasing as the reciprocal of the square root of the number of trials. Four different psychometric functions are used as indicated on the panels. The analysis psychometric function is the same as the observer's psychometric function. Their equations are given in Table II.

and the standard deviation of the threshold estimates decreases by the square root of the number of trials on which it is based.

B. Mismatch between quite different psychometric functions

The previous analysis has indicated that, if the experimenter uses a psychometric function that closely resembles that used by the observer, the effectiveness of the maximum-likelihood procedure is not greatly impaired. This result leads naturally to the next question, namely, what if the two psychometric functions are quite different? To analyze this situation, we used both the Gaussian and logistic functions (functions 1 and 2 of Table II), but used a power function of the stimulus variable to greatly alter the form of the psychometric function. For the Gaussian function, we used a stimulus variable that was proportional to x^k, where $k = 1, 2$, or 3. Figure 6(a) shows the three psychometric functions that result. The functions have been scaled so that they have the same stimulus value at 75% correct. As k increases, the slope of the psychometric function increases and the range diminishes. The set of psychometric functions shown in Fig. 6(a) would encompass all the psychometric functions found in auditory experiments. The sweetpoint of all three psychometric functions is at the stimulus corresponding to a proportion of 0.94 correct judgments. Three different logistic functions were also investigated. They are shown in Fig. 6(b). Again, the functions were adjusted to have the same stimulus level at a proportion of 0.75. Their sweetpoint corresponds to a proportion of 0.92. They were constructed to

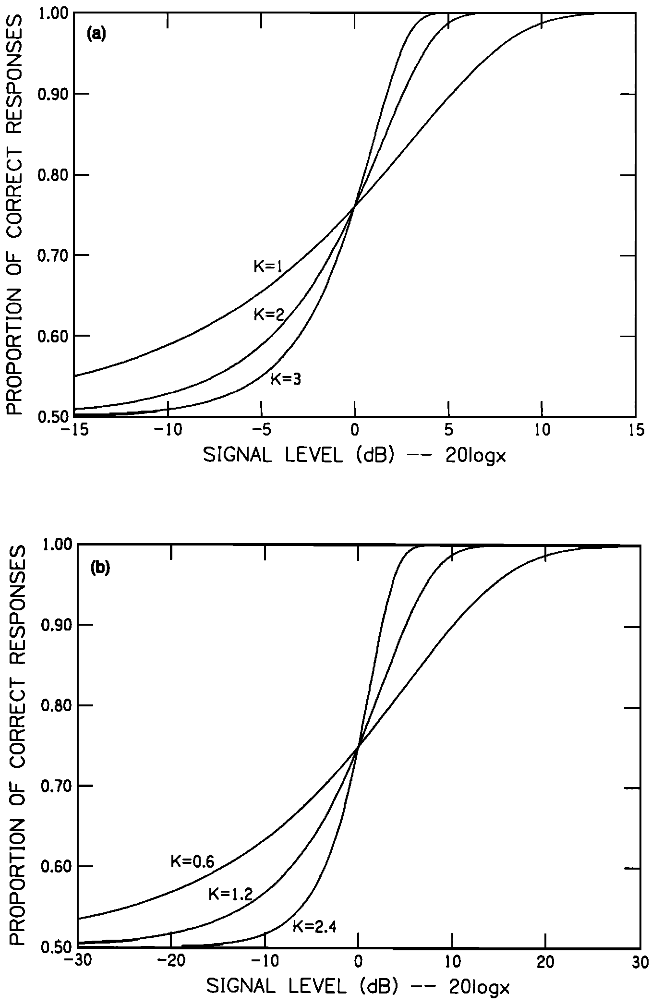

FIG. 6. Three sets of psychometric functions used in the simulations. They differ by raising the stimulus energy to different exponents (i.e., the arguments of the functions are of the form x^K). (a) Gaussian psychometric functions; (b) logistic psychometric functions. In both figures, the functions are scaled so that the same stimulus level produces a proportion of 0.75.

encompass an even larger range of psychometric functions. Note, the abscissa has a range of 60 dB in this figure, compared with the range of 30 dB in Fig. 6(a).

Once more, we ran simulations using each of the three functions as both the observer and the assumed psychometric function used in the maximum-likelihood procedure. Table IV shows the results of 2000 simulations of 50-trial runs that were used to estimate the threshold values. For each of the three observers, three analysis functions were used. The simulations were carried out three times with three different step sizes: 1, 1/2, and 1/4 dB. The major entry in the table is the standard deviation of the threshold estimate. As can be seen in the table, step size is not an important variable. Note, the standard deviations are about twice as large for the logistic function, because the range of the psychometric function is about twice that used for the Gaussian distribution. Each table is organized in three sets: one for each observer's psychometric function. One would expect the variability of the threshold estimate to be the least when the analysis psychometric function is equal to that used by the observers. For the shallower psychometric function, this is true. But, for the steeper psychometric function, the minimum is often obtained with a slightly less steep psychometric function. I have no good explanation for this result.

TABLE IV. Standard deviation of threshold estimates obtained from 2000 simulations. The value in parentheses is the ratio of the sigma for that condition to the minimum sigma obtained for that observer.

Psychometric function used by observer	analysis	Step size 1/4	1/2	1
Gaussian ($k =$)				
1	1	1.98	2.02	1.97
		(1.00)	(1.00)	(1.00)
1	2	2.42	2.42	2.33
		(1.22)	(1.20)	(1.18)
1	3	2.72	2.79	2.87
		(1.37)	(1.38)	(1.45)
2	1	1.13	1.11	1.15
		(1.04)	(1.00)	(1.04)
2	2	1.09	1.11	1.11
		(1.00)	(1.00)	(1.00)
2	3	1.35	1.28	1.27
		(1.24)	(1.15)	(1.14)
3	1	0.85	0.88	0.89
		(1.12)	(1.17)	(1.06)
3	2	0.76	0.76	0.84
		(1.00)	(1.00)	(1.00)
3	3	0.92	0.82	0.99
		(1.21)	(1.09)	(1.18)
Logistic ($k =$)				
0.6	0.6	3.61	3.60	3.54
		(1.00)	(1.00)	(1.00)
0.6	1.2	3.92	3.91	4.02
		(1.09)	(1.09)	(1.14)
0.6	2.4	4.48	4.69	4.55
		(1.24)	(1.30)	(1.29)
1.2	0.6	2.05	2.03	2.05
		(1.08)	(1.03)	(1.07)
1.2	1.2	1.89	1.98	1.92
		(1.00)	(1.00)	(1.00)
1.2	2.4	2.27	2.22	2.36
		(1.20)	(1.12)	(1.23)
2.4	0.6	1.32	1.34	1.39
		(1.25)	(1.18)	(1.26)
2.4	1.2	1.06	1.14	1.10
		(1.00)	(1.00)	(1.00)
2.4	2.4	1.17	1.15	1.13
		(1.10)	(1.01)	(1.03)

Below each entry is the ratio of the standard deviation for the minimum of that set of three compared to the other two values. The ratio seldom exceeds 1.3, indicating that little harm is done in using a psychometric function with a very different range from that used by the observer. In general, the results indicate that less increase in the standard deviation results when the assumed psychometric function is shallower than the true psychometric function. Overall, the conclusion is that variability of the threshold estimates is not strongly affected by enormous mismatch between the observer's psychometric function and that used in the maximum-likelihood analysis. This result was very surprising to the author.

Before leaving this topic, I should sound one note of caution. Using the wrong psychometric function introduces sizable bias in the estimate of the threshold value. The maximum-likelihood procedure presents the stimulus at a high stimulus level corresponding to 0.94 for the Gaussian or 0.92 for the logistic function. The stimulus is adjusted so that the observer achieves this percentage correct. The threshold, defined as the 75% point, is then not at the correct distance from the sweetpoint, except when the analysis and the observer's psychometric functions are equal. Biases of nearly 4 dB for the Gaussian functions and 8 dB for the logistic functions were observed. The direction and amount of the bias could be explained by assuming that both functions were aligned to be equal at the sweetpoint. The threshold estimate and the bias could be explained as the difference between the sweetpoint and the stimulus corresponding to 75% correct. Such a bias may not be serious, because, if all conditions of the experiment produce the same listener's psychometric function, then the bias could be regarded as a constant error. Obviously, it would be desirable to select a function for the analysis that has roughly the same range as that used by the observer. In this way, the constant error could be minimized.

III. RESULTS OBTAINED WITH UP–DOWN ADAPTIVE PROCEDURES

Although the preceding simulations suggest that the maximum-likelihood procedure is robust to errors about the assumed psychometric functions, some experimenters still prefer up–down adaptive procedures. They prefer these procedures because they are simple and make essentially no assumption about the form of the psychometric function—other than it is monotonic increasing. Unfortunately, the most common up–down procedure, the two-down one-up procedure, has an equilibrium probability of $1/\sqrt{2} = 0.707$. The probability at the sweetpoint of most psychometric functions is a much higher value: 0.81–0.94 is the range for our four functions. This led us to consider whether the up–down procedures could achieve better threshold estimates if they were modified to achieve higher equilibrium probabilities, values closer to the probability associated with the sweetpoint of the particular psychometric function.

An easy way to change the equilibrium probability in an up–down procedure is to change the number of successive correct responses needed to move the stimulus to a lower value. If K successive correct responses are required to move down and any error leads to an increase in the stimulus value, the probability of moving down or up is then

$$P(\text{down}) = p^K \quad P(\text{up}) = 1 - p^K,$$

where p is the probability of being correct. The equilibrium probability is obtained when these two forces are equal, or

$$P(\text{equilibrium}) = (1/2)^{1/K}.$$

We carried out simulations using $K = 2$, 3, 4, and 5. The equilibrium probabilities for the different values of K are 0.707, 0.794, 0.841, and 0.871, respectively. We did not use higher values of K because, as K increases, the process becomes increasingly sluggish and fewer turnarounds are achieved. In up–down procedures, an average of the last several pairs of turnarounds is typically used as an estimate of threshold. In our simulations, we used an initial step of 4 dB changing to a 2-dB step after the first four turnarounds. The threshold value for the stimulus was computed as the average of the final pairs of turnarounds, excluding the first four. This procedure has been shown to give reasonable results in our previous simulations of the up–down process (Green *et al.*, 1989).

Table V presents our results. For each of the four psychometric functions listed in Table II, we simulated different up–down criteria. The different up–down rules are indicated by the different values of K in the first column of the table. The second full column shows the average number of reversals for each up–down rule based on 100 trials in the two-alternative forced-choice procedure. This average number of reversals decreases as K increases, as we would expect. The third column of the table shows the standard deviation of the threshold estimates. This standard deviation is compared with the standard deviation expected from theory as given in the fourth column of the table. This quantity is the square

TABLE V. Simulations using different adaptive rules (observed sigma based on 25 000 simulations). Adaptive rule: If K successive correct move DOWN; any error move UP (for each function, equilibrium probability and expected sigma, given 100 trials, are indicated in parentheses). Function No. (sweetpoint probability, sigma at sweetpoint). Initial stimulus value was threshold for that function; there were 100 trial blocks; and ratio *re*: sweetpoint is ratio of sigma at sweetpoint/observed sigma.

Rule $=K$	Av. No. rev	Sigma Observed	Sigma Expected	Ratio	Ratio *re*: sweetpoint
		function 1 (0.942, sigma = 1.11)			
2 (0.707)	30.5	2.54	2.10	0.83	0.44
3 (0.794)	22.3	1.86	1.50	0.81	0.60
4 (0.840)	17.6	1.70	1.31	0.77	0.65
5 (0.871)	14.6	1.61	1.22	0.76	0.69
		function 2 (0.917, sigma = 1.31)			
2	30.6	2.62	2.17	0.83	0.50
3	22.2	1.98	1.59	0.80	0.66
4	17.4	1.81	1.43	0.79	0.72
5	14.4	1.70	1.30	0.76	0.77
		function 3 (0.845, sigma = 1.68)			
2	30.6	2.34	1.96	0.84	0.72
3	22.0	2.03	1.72	0.85	0.83
4	17.2	2.06	1.69	0.82	0.82
5	14.1	2.14	1.70	0.79	0.79
		function 4 (0.811, sigma = 1.67)			
2	30.6	2.24	1.88	0.84	0.75
3	22.1	1.93	1.67	0.86	0.87
4	17.2	1.95	1.69	0.87	0.86
5	14.1	2.08	1.74	0.84	0.80

root of Eq. (2) divided by 100, the number of trials used in this simulation. The value of p used in this calculation is the equilibrium probability corresponding to the K of the up–down rule. This p value is listed in parentheses next to K for the first psychometric function. The fifth column of the table is the ratio of the numbers listed in columns 4 and 3. The ratio hovers around 0.8, indicating that the theory provides a reasonable approximation to the empirical estimates of variability obtained in the simulations. Finally, the sixth, and last, column of the table expresses the ratio of the observed standard deviation to the standard deviation that could be achieved had the stimulus been placed at the sweetpoint for each psychometric function. For the first two psychometric functions, whose probability at the sweetpoint is larger than any equilibrium probability studied, the ratios are largest at $K = 5$ and decrease for smaller values of K. For the third and fourth psychometric functions, where the sweetpoint probability is near the equilibrium value corresponding to $K = 4$, the best ratio is achieved near that value, and lesser values are obtained as we change K.

The conclusion to be drawn from these simulations depends almost entirely on what one believes to be the correct form of the observer's psychometric function. If functions 1 or 2 are correct, then the maximum-likelihood procedure could produce threshold estimates whose standard errors are half what will be produced by the up–down procedures. Such a difference would suggest that the up–down procedure is unacceptable. If functions 3 or 4 are correct, however, then the relative change in standard deviation is only 0.8, which is probably an acceptable loss in precision.

In two other simulations, we explored the sensitivity of these simulations to certain parameter values. First, we changed the initial starting stimulus to a higher value, 10 dB, rather than starting at the threshold value. This higher value reflects the common practice using the up–down procedure. The new simulations produced little change in the observed sigmas; the largest absolute change was 0.1 dB and the relative change in sigma was always less than 4%. The pattern of results was almost identical to that seen in Table V. An additional alteration was to change the number of trials from 100 to 50. Once more, the simulation started with a stimulus value of $+10$ dB. The decrease in the number of trials naturally led to an increase in the observed variability in threshold estimates. The average increase in sigma was somewhat more than the square root of 2, being 1.49 on average. Again, the general pattern of results, the changes in sigma with K, was very similar. The notable difference was the number of threshold runs that did not produce enough turnarounds (fewer than four) to achieve a valid threshold estimate. All 5000 threshold estimates were valid for $K = 2$ and less than 0.3% failed for $K = 3$. But for $K = 4$, the number of invalid threshold runs was 3.3% and, for $K = 5$, 13.6% of the 50-trial blocks failed to produce more than four turnarounds. Thus, if the higher values of K are to be used, it appears that the experimenter must use more than 50 trials in a block to estimate threshold successfully.

Some experimenters use an up–down procedure and stop after a fixed number of turnarounds. While such a practice avoids the problem of invalid threshold estimates, I consider such a procedure questionable. Turnarounds are generated by sequences of wrong and right responses. The total number of trials used to estimate the threshold will therefore be reduced if the observer is somewhat inconsistent and occasionally makes a wrong response to a very intense (easy) stimulus level. The temptation to shorten a very tedious task such as observing in a psychophysical experiment is a strong one in practically all listeners. For that reason, we have always used a fixed number of observations in a trial block. In that way, we at least provide no incentive for inconsistent behavior on the part of the listener.

IV. HOW THE SWEETPOINT CHANGES FOR INATTENTIVE OBSERVERS

The sweetpoints determined by the maximum-likelihood procedures means that, for most psychometric functions, the observer will make very few errors in a threshold determination. A common query about this analysis is what will happen if the observer has momentary lapses of attention and produces occasional errors unrelated to those inherent in the probabilistic nature of the psychometric function. Will not these additional errors seriously impair the variability of the threshold estimates and also suggest that a stimulus placement policy that leads to a higher probability of error would be a more efficient procedure? Basically, the argument is that the impact of occasional errors will be lessened if a lower proportion of correct responses is followed in the estimation process, because, then, more "real" errors will occur. In short, occasional lapses on the part of the observer must suggest that the sweetpoint occurs at a lower proportion of correct responses.

To analyze this issue, we assume the simplest form of this inattention hypothesis and determined the effect of this hypothesis on the variability of threshold estimates and the value of the sweetpoint. The assumption that the observer has occasional lapses of attention is tantamount to assuming a new form for the psychometric function. Let us assume that, during some portion of the time g the observer has a lapse in attention, and, hence, in a two-alternative forced-choice task, the observer will achieve a 0.5 proportion of correct responses, independent of the stimulus level. For the remaining proportion of trials $1 - g$, the "true" psychometric function $\Phi(x)$ will obtain. This psychometric function will depend on the stimulus value x in the usual way. Thus the observed psychometric function will be a result of the linear mix of the two, equal to $(1 - g)\Phi + 0.5g$. The resultant function will vary from 0.5 for very small values of x, but will asymptote to $1 - 0.5g$ for very large stimulus values. We can now calculate the resultant sweetpoint and variability in the threshold estimates for different values of g. The results of these calculations are shown in Fig. 7 for the Gaussian psychometric function, function 1; see also Table II. As the probability of a lapse in attention g increases, the standard deviation of the threshold estimates will increase. It will double only if g exceeds 0.2. Similarly, the probability value at the sweetpoint will diminish from a value of 0.94, at $g = 0$, to a value of 0.8, at $g = 0.2$. A value of $g = 0.2$ is probably higher than experimenters would admit. More typical esti-

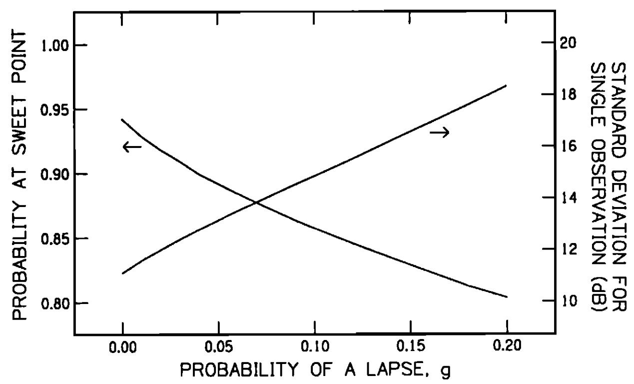

FIG. 7. The effect of changes in the probability of a lapse in attention g on the probability value of the sweetpoint, left ordinate, and the standard deviation of the threshold estimates, right ordinate.

mates are less than 0.05, in which case the probability at the sweetpoint would still exceed 0.9. The larger values for g shown in the graph were included to give some indication of the effect of this variable in a clinical situation where strong extraneous factors might well be present. For most psychophysical experiments, we may assume $g < 0.05$. In that case, the lapses in attention do influence the stimulus policy placement, but not in a major way.

V. EMPIRICAL RESULTS

A. Naive listeners

To test some of these theoretical concepts, we used human listeners and determined their thresholds, using a maximum-likelihood procedure, for the increment in intensity of a single sinusoidal component of a complex auditory stimulus (profile stimulus). From previous research (Raney *et al.*, 1989), we know that the psychometric function for this detection task is reasonably approximated by a cumulative Gaussian function, such as function 1 (see Table II). Thirty-two threshold estimates were made using two different stimulus placement policies. Each threshold estimate was based on 30 two-alternative forced-choice trials. The stimulus was adjusted in level by 1-dB steps, and the possible psychometric functions differed by the same amount. One placement was at the sweetpoint of the psychometric function, a probability of a correct response corresponding to 0.94. The other placement was a signal level 9 dB below that value, corresponding to a probability of 0.707. The initial stimulus level was 15–20 dB above the presumed threshold; thus the response on the first trial was invariably correct. The threshold was the stimulus level corresponding to a probability of a correct response equal to 0.76, as estimated on the final, 13th trial. The tests were conducted in four sets of eight threshold determinations, each set alternating between the two stimulus placements. The 64 estimates were completed in a pair of 2-h sessions conducted on successive days.

We tested 13 listeners, mostly naive subjects. They were recruited at the beginning of the semester, prior to their participation in a number of other experiments. Our primary interest is in the standard deviations computed over the 32 threshold estimates. Table VI shows the results of the experiment.

The obtained results are nicely in accord with theoretical expectations. The variability of the threshold estimates when the sweetpoint (0.94) is used for placing the next stimulus value is nearly half the value when the stimulus is placed at a level corresponding to the equilibrium probability in a two-down one-up procedure (0.707). So far as we know, this is the first clear demonstration that the variability in threshold estimates using one psychophysical procedure differs in

TABLE VI. Results of 32 threshold estimates based on 30 trials.

	Stimulus placement policy				
	Sweetpoint (94%)		9 dB below SP (70.7%)		
Obs.	σ in dB	$P(c)$	σ in dB	$P(c)$	Ratio of σ
1	3.7	0.90	6.9	0.69	0.54
2	4.4	0.87	7.6	0.63	0.58
3	3.8	0.88	2.8	0.70	1.33
4	2.5	0.88	5.9	0.69	0.43
5	3.6	0.89	4.5	0.71	0.79
6	4.9	0.87	10.0	0.64	0.49
7	3.9	0.87	9.2	0.64	0.42
8	3.6	0.88	7.7	0.65	0.47
9	3.5	0.83	7.4	0.62	0.47
10	3.4	0.89	7.8	0.69	0.44
11	3.6	0.80	10.7	0.62	0.33
12	4.9	0.85	10.0	0.65	0.49
13	3.1	0.87	4.8	0.67	0.64
Av	3.75	0.87	7.34	0.66	0.51
Theoretical	2.03	0.94	3.80	0.707	0.53
Ratio of theoretical σ to obtained σ	0.54		0.52		

any significant respect from those obtained using an alternative procedure. This success probably occurred because we used stimulus values that were expected, theoretically, to produce a factor of 2 differences in variability. Previous experimental efforts have not been as closely guided by theoretical analysis (Madigan and Williams, 1987; Shelton *et al.*, 1982; Simpson, 1989).

The estimated probability of a correct response is somewhat lower than the target probability of the placement policy. This reflects the fact that the stimulus level is being adjusted to determine threshold, and the initial stimulus settings are often well below the final estimates. We will discuss this matter in more detail shortly. Computer simulations of the maximum-likelihood procedure using only 30 trials gave average percent correct values of 86% and 66% for the two stimulus placement policies used in the experiment, remarkably close to the average values we obtained experimentally. This initial search to find the correct stimulus level also influences the effectiveness of the procedure. The obtained standard deviations for both stimulus placements are only about half the theoretical expectation. That ratio would probably improve if we increased the total number of trials used to estimate threshold, as we demonstrated for the simulation results (see Fig. 5).

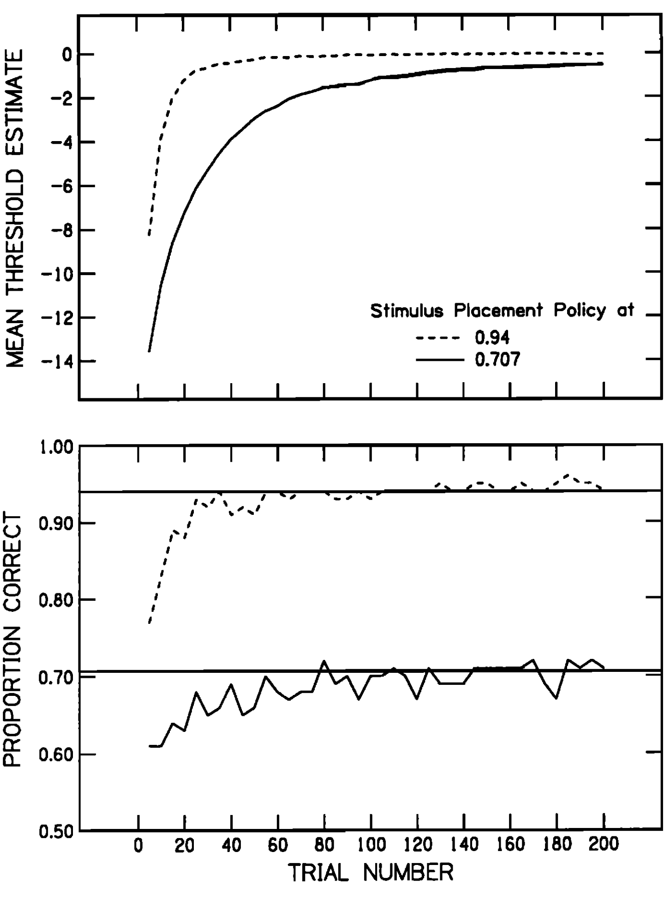

FIG. 8. The change in the mean threshold estimate and the probability of a correct response as estimated as a function of trial number in 1000 simulations. Function I (see Table II) was used as the analysis and observer's psychometric function. Two placement policies are shown on the graph, a stimulus level corresponding to a probability of a correct response equal to 0.94 (the sweetpoint for this function) and 0.707.

Finally, we must comment on some data not shown in the table, namely, the mean threshold values estimated in these experiments. The threshold value for the stimulus should, in theory, be independent of the stimulus placement procedure. The estimated thresholds were noticeably different for the two empirical threshold estimates. The thresholds were slightly more than 4 dB different, being -15.8 dB when the stimulus was placed at the sweetpoint and -20.2 dB for the lower stimulus placement. Thus the threshold estimated with a stimulus placed at the lower percent correct value would suggest that the listener is more than 4.4 dB more sensitive than when the stimulus was placed at the sweetpoint. The difference in threshold was in the same direction for all 13 listeners, and the range of differences was 1.5–8.6 dB.

We ran some computer simulations to determine if these differences in the threshold estimates were a characteristic of the maximum-likelihood procedure or of the human observers. Maximum-likelihood simulations were conducted with the two placement procedures used in the human experiments, that is, the sweetpoint (0.94) and a value 9 dB below that level (0.707). The number of two-alternative forced-choice trials in a single run was set at 200. The threshold value and the proportion of correct responses were estimated, based on 1000 simulations, at five-trial intervals. Figure 8 shows the results. The upper panel shows the mean threshold estimates. The nominal (0.76) threshold value was 0 dB in these simulations. For the stimulus placement at the sweetpoint, the estimated threshold is within 1 dB of the correct value after 20 trials. The lower placement value takes about 100 trials to achieve that level of accuracy. At the end of 30 trials, the number of trials used in the human experiments, the bias is -5.3 dB for the lower placement level and about -0.7 dB for the placement at the sweetpoint. The difference of 4.6 dB is close to the value observed in the human experiment.

In addition to the mean data, we have also plotted the probability of a correct judgment for the two placement policies as a function of the trial number in the lower part of Fig. 8. Recall that the stimulus is initially 15–20 dB above threshold, so that the first response is nearly always correct. If the first response is correct, then the next stimulus is placed at the lowest possible level, which generates a probability of 0.5. Gradually, the probability makes its way back to the nominal value, 0.94 for the case of placement at the sweetpoint and 0.707 for the lower placement value.

These initial transients, in both the mean threshold estimate and the probability of a correct response, remind us that runs based on very few observations will, by nature, produce biased estimates. Bias in a threshold estimate, while hardly desirable, is generally not a serious issue. The same bias will often be present in all the threshold estimates and generally the datum of primary interest is the relative change in threshold as a function of some other stimulus parameter. The size of the bias, however, is much less if the sweetpoint is used in the stimulus placement policy.

B. Experienced listeners

The preceding results were obtained from naive listeners who had not previously served in psychoacoustic experi-

TABLE VII. Three practiced listeners in two detection tasks. Entries are standard deviation (σ) in dB. Sigma based on 32 threshold estimates using 30 trials for maximum-likelihood procedures and 50 trials for up–down procedure.

Profile task			
	Sweetpoint (94%)	9 dB below SP (70.7%)	2 down–1 up
Obs. 1	4.0	5.1	4.3 (5.5)
Obs. 7	3.2	5.0	3.4 (4.4)
Obs. 12	3.9	5.2	4.3 (5.5)
Av	3.7	5.1	4.0 (5.2)
Sinusoid in noise task			
	Sweetpoint (94%)	4 dB below SP (70.7%)	2 down–1 up
Obs. 1	2.7	5.6	2.4 (3.1)
Obs. 7	2.7	5.7	3.5 (4.5)
Obs. 12	2.0	4.3	1.9 (2.5)
Av	2.5	5.2	2.6 (3.4)

ments. About 2 months after this experiment, we had the opportunity to retest three of the these listeners. We repeated the original threshold determinations in the profile task. We also measured their thresholds for detecting a sinusoidal signal (1000 Hz) in noise. Once more, 32 threshold estimates were made using the maximum-likelihood procedure in 30 trials of a two-alternative forced-choice task. We used function 1 (Table II) as the psychometric function in the profile task. For the tone-in-noise experiment, we used function 1, except we replaced x with x^2. This steepens the psychometric function so that it covers the 10-dB range that is typical of tone-in-noise detection (Green and Swets, 1988). In addition to the maximum-likelihood estimates, we also made 32 estimates of thresholds using a two-down one-up adaptive procedure. In order to produce sufficient reversals to estimate threshold, we used 50 trials in the up–down task, rather than the 30 trials used in the maximum-likelihood procedure. Fifty trials in a two-down one-up task produces about 10 to 16 reversals.

The subjects had now participated in about 80 h of listening, primarily in a binaural experiment investigating the ability to detect coherence among sinusoidal envelopes. This listening was quite different from either the profile or tone-in-noise detection tasks in which the thresholds were estimated.

Table VII presents a summary of the standard deviation of the threshold estimates measured in the conditions. The listeners were observers 1, 7, and 12 from the original experiment (see Table VI). The maximum-likelihood estimates are based on 30 trials, whereas the up–down procedure is based on 50. We list the standard deviation actually measured in the two-down one-up adaptive task. The number in parentheses is that value increased by a factor of 1.29 $[(50/30)^{1/2}]$, which is an estimate of the standard deviation based on 30 instead of 50 trials.

Let us begin by considering the profile data, since it can also be compared with the earlier threshold estimates (Table VI). The most noticeable change is the decrease in variability of the threshold estimated at the 70.7% level. The average variability estimated when the stimuli were placed at the sweetpoint was 4.2 and is now 3.7 dB, a reduction of 0.5 dB. The average variability when the stimuli were placed at the 70.7% point was 8.7 and is now 5.1 dB, a reduction of 3.6 dB. The ratio of sigmas for the two threshold estimates is now about 0.7, rather than the original ratio of about 0.5. The up–down procedure, which tracks a probability of 0.707, produces a variability in the threshold estimate of about 5.2 dB when corrected for 30 rather than 50 trials. This sigma is nearly equal to that produced by the maximum-likelihood procedure when the stimulus is placed at the same detection level. Note, however, that there is variability among the subjects.

For the tone-in-noise data, the variability should be about half that measured in the profile experiment, because the psychometric function has about half the range. That difference is apparent in only two of the threshold estimates—the sweetpoint estimate and the up–down estimate. The ratio of variability for the two maximum-likelihood estimates is approximately one-half, as theory would predict. The up–down estimates are noticeably better than those seen in the profile experiment. For two of the listeners, the uncorrected sigmas are actually smaller than those obtained with the optimum maximum-likelihood procedure. One can also argue that the sigmas are nearly equal and the maximum-likelihood procedure uses only about half as many trials. For the average data, the corrected variability of the up–down procedure lies about midway between the sweetpoint estimates and the 70.7% estimates.

It would appear that experienced observers do not show the large difference in variability shown by less experienced listeners. Thus some of the advantages of the maximum-likelihood procedure may be diminished if tested with very experienced listeners. Also, the results obtained with the up–down procedures, at least for the tone-in-noise experiment, may not be as inefficient as would be predicted, given the rather low probability that these procedures track.

VI. CONCLUSIONS

We have confirmed, both with computer simulations and with human listeners, that placing the stimulus level at a relatively high level decreases the variability of the threshold estimate in the maximum-likelihood procedure. With naive listeners, the standard deviations of the threshold estimates are nearly equal to those predicted in theory. With more experienced listeners, the gains are somewhat less. These improvements are not very sensitive to the exact form of the psychometric function assumed by the experimenter, as long as the range of the psychometric function is nearly the same as that used by the observer; nor are they very sensitive to momentary lapses in the observer's attention, as long as the probability of such lapses is small. Improvements in the precision of the threshold estimates are also found in up–down procedures that track a higher probability of a correct response. The relative effectiveness of the maximum-likelihood estimates improves, and the bias in estimate decreases as the total number of trials used to estimate the threshold value increases.

I generally agree with the sentiments expressed by

Simpson (1989) in a recent paper, namely, that all psychophysical procedures produce nearly the same result. But most psychophysical experiments take a long time, and all of us are interested in minimizing the standard deviation of our threshold estimates. If the ratio of standard deviations is 0.7, the same precision can be obtained from an experiment that lasts 3 weeks instead of 6.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation and by the Air Force Office of Scientific Research. I would like to thank Dr. J. Donald Laming for his extensive comments on an early draft, and for the comments of an anonymous reviewer on a later draft. Dr. Huanping Dai and Dr. Virginia Richards provided helpful comments. The paper also benefited from a presentation of some of these results to a group that attended the hearing research seminar in our laboratory. Comments and questions by Dr. Steven Colburn, Dr. William Yost, and Dr. Roy Patterson stimulated the analysis of the inattentive observer.

Green, D. M. (**1988**). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U. P., New York).

Green, D. M., and Swets, J. A. (**1988**). *Signal Detection Theory and Psychophysics* (Peninsular, Los Altos, CA).

Green, D. M., Richards, V. M., and Forrest, T. (**1989**). "Stimulus step size and heterogeneous stimulus conditions in adaptive psychophysics," J. Acoust. Soc. Am. **86**, 629–636.

Laming, D., and Marsh, D. (**1988**). "Some performance tests of QUEST on measurements of vibrotactile thresholds," Percept. Psychophys. **44**(2), 99–107.

Levitt, H. (**1971**). "Transformed up–down methods in psychophysics," J. Acoust. Soc. Am. **49**, 467–477.

Madigan, R., and Williams, D. (**1987**). "Maximum-likelihood psychometric procedures in two-alternative forced-choice: Evaluation and recommendations," Percept. Psychophys. **42**(3), 240–249.

Raney, J., Onsan, Z., Richards, V. M., and Green, D. M. (**1989**). "Signal uncertainty and psychometric functions in profile analysis," J. Acoust. Soc. Am. **86**, 954–960.

Shelton, B. R., Picardi, M. C., and Green, D. M. (**1982**). "Comparison of three adaptive psychophysical procedures," J. Acoust. Soc. Am. **71**, 1527–1533.

Simpson, W. A. (**1989**). "The step method: A new adaptive psychophysical procedure," Percept. Psychophys. **45**(6), 572–576.

Taylor, M. M., and Creelman, C. D. (**1967**). "PEST: Efficient estimates on probability functions," J. Acoust. Soc. Am. **41**, 782–787.

Watson, A. B., and Pelli, D. G. (**1983**). "QUEST: A Bayesian adaptive psychometric method," Percept. Psychophys. **33**(2), 113–120.