

QUEST: A Bayesian adaptive psychometric method

ANDREW B. WATSON

NASA Ames Research Center, Moffett Field, California
and Stanford University, Stanford, California

and

DENIS G. PELLI

Institute for Sensory Research, Syracuse University, Syracuse, New York

An adaptive psychometric procedure that places each trial at the current most probable Bayesian estimate of threshold is described. The procedure takes advantage of the common finding that the human psychometric function is invariant in form when expressed as a function of log intensity. The procedure is simple, fast, and efficient, and may be easily implemented on any computer.

A psychometric function describes the relation between some physical measure of a stimulus and the probability of a particular psychophysical response. The physical measure is usually stimulus strength, and the response is "yes" (in a yes/no experiment) or "correct" (in a forced-choice experiment). More generally, there are several possible responses, each with its own psychometric function (e.g., in a rating scale experiment). Psychometric procedures are ways of testing the observer so as to gain information about the psychometric function. The advantages of adaptive procedures, which make use of previous responses to guide further testing, have been discussed by numerous authors (Cornsweet, 1962; Levitt, 1971; Taylor & Creelman, 1967; Wetherill & Levitt, 1965).

Because trials are more effective when they are judiciously placed, an adaptive procedure will be more efficient the more it makes intelligent use of available information. This information is of two sorts: that derived from previous trials (data), and that drawn from the memory of the experimenter, published reports, and so on (prior knowledge). Prior knowledge may be further divided into information about the shape of the psychometric function and knowledge about threshold in the particular condition under study. Several recent methods make effi-

cient use of the data and of knowledge about the shape of the psychometric function by calculating after each trial a maximum likelihood estimate of threshold (Hall, 1968; Klein, 1981; Pavel, 1981; Pentland, 1980; Watson & Pelli, 1979). Our procedure, which we have dubbed QUEST, also makes use of prior knowledge about the location of threshold.

Assumptions

We embody our information about the shape of the psychometric function in an assumption.

- (1) *The psychometric function has the same shape under all conditions when expressed as a function of log intensity. From condition to condition, it differs only in position along the log intensity axis. This position is set by a parameter, T , the threshold, also expressed in units of log intensity.*

This remarkable property of the psychometric function has been noted by numerous authors (Green & Luce, 1975; Nachmias, 1981; Roufs, 1974; Watson, 1979). It will allow us to describe any particular psychometric function, $p_T(x)$, characterized by threshold T , in terms of a canonical form $\Psi(x)$, by the relation

$$p_T(x) = \Psi(x - T), \quad (1)$$

where x is log intensity. The parameter T may be chosen as any convenient landmark in the function Ψ , for example, the 75% point, but we will recommend a particular landmark below. As we shall see, this constraint on the form of the psychometric function greatly simplifies what would otherwise be complex

The QUEST procedure was developed at the Kenneth Craik Laboratory of Cambridge University. A.B.W. was supported by an NIH postdoctoral fellowship, F32 EY05219. D.G.P. was supported by a British Ministry of Defence grant, "Spatial noise spectra and target detection/recognition," to F. W. Campbell. We thank Stanley Klein, Daniel Kersten, and Joseph Hall for helpful comments. A.B.W.'s mailing address is: Department of Psychology, Jordan Hall, Building 420, Stanford University, Stanford, California 94305.

calculations. Although this assumption appears to be upheld under most conditions, it may fail under some conditions (see Nachmias, 1981, and Massof, 1981, for possible examples). This is no great problem, however, since such a failure may be readily detected in the data. Furthermore, simulations suggest that modest departures from this assumption reduce efficiency by only modest amounts.

(2) *The parameter T of the psychometric function does not vary from trial to trial.*

This assumption allows us to use *all* previous trials to estimate T and to guide further testing. This contrasts with most simple procedures, which select the next test point on the basis of only a small number of previous trials. When this assumption is not valid, for example, during the course of adaptation, our method is not appropriate, but such circumstances generally require extraordinary measures.

(3) *Individual trials are statistically independent.*

This is a conventional assumption of psychometric methods. It can usually be insured by interleaving the testing of several different psychometric functions, either of the same or of different stimuli.

Theory

First we consider the representation of prior information. Beyond what is embodied in the assumptions above, this information takes the form of knowledge about the value of threshold in the condition being investigated. This may be derived from previous experiments, hunches, and the like. We represent this knowledge by the *prior* probability density function (pdf) of threshold, $f_T(T)$. This specifies, for each possible intensity, the a priori probability that threshold lies at that point. Typically, this might be a broad Gaussian or rectangular distribution centered on T_{prior} , the best guess of the location of threshold.

The second source of information is the set of trials collected thus far. Let us denote these data by the letter D . The information contained in the data may be expressed in the form of a likelihood function, $f_{D|T}(D | T)$, which is the pdf of the data conditional upon T . This information may be combined with the prior pdf by *Bayes' rule* to form the *posterior* pdf,

$$f_{T|D}(T | D) = \frac{f_T(T)f_{D|T}(D | T)}{f_D(D)} \quad (2)$$

The function $f_D(D)$ is the prior pdf of the data. For a particular set of data, this is a constant which, for most purposes, need not be computed. When necessary, its value may be obtained from the relation

$$f_D(D) = \int_{-\infty}^{\infty} f_T(T)f_{D|T}(D | T)dT. \quad (3)$$

The posterior pdf contains *all* of the available information about threshold: assumptions, prior estimate, and data. It forms a natural basis on which to estimate threshold and on which to select the placement of the next trial. Note that the most probable estimates of threshold is simply the mode (location of the peak) of the posterior pdf, which we represent by T_{post} . If the prior pdf is divided out, what remains is the likelihood function, and its mode is the maximum likelihood estimate of threshold, which we represent by T_{like} .

Priors and prejudice. In deciding what use to make of prior information, we confront a dilemma. On the one hand, we wish to maximize the effectiveness of each trial, and this can only be done by making use of prior information. On the other hand, we do not want to prejudice the experimental outcome, as we will if we allow prior information to influence the final estimate of threshold. QUEST adopts the following solution. Prior information (in the form of the prior pdf) is used along with the data (in the form of the likelihood function) to guide the placement of trials. But, in deciding when to terminate the experiment and in producing a final estimate of threshold, we divide out the prior density and use only the likelihood function.

A placement rule. Where do we place the next trial? To estimate the position of the psychometric function, the ideal placement of each trial is at the value of x which minimizes the quantity

$$\frac{p_T(x)[1 - p_T(x)]}{[dp_T(x)/dx]^2}$$

Note that this *ideal sweat factor* (Taylor, 1971) is the square of the ratio of the binomial standard deviation and the slope of the psychometric function at the point in question. For convenience, we define T as the point at which this quantity is minimized. However, to place trials at this ideal point we would need to know T before we could estimate it. Since we are aware of no general solution to the problem of where to place the next trial, QUEST adopts an intuitively appealing procedure: it places the next trial at the current most probable estimate of T . In other words, the trial is placed at T_{post} , the mode of the posterior pdf.

Practice

In the preceding section we developed the idea of the posterior pdf of threshold, and argued that it is the appropriate basis on which to estimate threshold and to position further trials. Here we develop a sim-

ple and fast method of computing the posterior pdf. The method is sufficiently rapid for it to be computed after every trial without noticeably slowing the progress of the experiment.

We begin by considering the three terms on the right side of Equation 2. The last term, $f_D(D)$, can be derived from the product of the first two, as shown by Equation 3. Thus, *all* the information about the posterior pdf is contained in the product of the first two terms, $f_T(T)f_{D|T}(D | T)$. This product is equal to the joint density of T and D , $f_{T,D}(T,D)$.

Furthermore, since all the information about the posterior pdf is contained in this joint density, it will likewise be present in the log of the joint density, which we write $Q(T)$,

$$Q(T) = \ln f_T(T) + \ln f_{D|T}(D | T). \quad (4)$$

We call this the "quest function." Its first part, the \log_e of the prior density, is supplied by the experimenter at the start of the session. We turn now to the calculation of the second part, $\ln f_{D|T}(D | T)$, which is the \log_e of the likelihood function.

Following n trials, the data set (D) consists of a sequence of responses, r_i at log intensities x_i , where $i = 1, \dots, n$. Each response is either a success ($r_i = 1$) or a failure ($r_i = 0$). The probability of a success at log intensity x is given by the psychometric function, $p_{S|T}(x) = p_T(x)$, and the probability of a failure is $p_{F|T}(x) = 1 - p_T(x)$. Since the responses are statistically independent, the likelihood of the sequence of n trials is

$$f_{D|T}(D | T) = \prod_{i=1}^n p_{r_i|T}(x_i). \quad (5)$$

Taking the \log_e of this function and substituting it into Equation 4, we get as an expression for the quest function after n trials:

$$Q_n(T) = \ln f_T(T) + \sum_{i=1}^n \ln p_{r_i|T}(x_i). \quad (6)$$

This expression reveals a very convenient fact: the quest function after trial n is equal to the quest function on trial $n-1$, plus one or the other of the functions $\ln p_T(x_n)$ or $\ln[1 - p_T(x_n)]$, depending on whether the trial was a success or a failure. This convenience is increased when we note that the two possible addends need not be recomputed for each value of T , but merely displaced, since, from Equation 1,

$$\begin{aligned} \ln p_T(x) &= \ln \Psi(x - T) \\ \ln[1 - p_T(x)] &= \ln[1 - \Psi(x - T)]. \end{aligned} \quad (7)$$

To emphasize this point, we define the success and failure functions,

$$\begin{aligned} S(x) &= \ln \Psi(-x) \\ F(x) &= \ln[1 - \Psi(-x)]. \end{aligned} \quad (8)$$

The development so far can then be summarized by the equations

$$\begin{aligned} Q_n(T) &= Q_{n-1}(T) + \begin{cases} S(T - x_n) & \text{if success} \\ F(T - x_n) & \text{if failure} \end{cases} \\ Q_0(T) &= \ln f_T(T) \end{aligned} \quad (9)$$

These equations show how the quest function is accumulated over the course of trials. If trial n is a success, we shift the success function horizontally a distance x_n and add it point by point to the existing quest function, $Q_{n-1}(T)$. If the trial is a failure, we shift and add the failure function instead. Before any trials are collected, the quest function is just the \log_e of the prior density.

Finally, we note that, in practice, the psychometric function is tested at only a finite number of intensities. Here we suppose that possible test intensities are spaced at intervals of Δx (the grain) along the log intensity axis. Since we will specify intensity only in increments of Δx , it will be sufficient to evaluate $Q(T)$ only at these possible intensities. Using square brackets to enclose the index of a function sampled at intervals of Δx , we can rewrite Equation 9 in the form

$$Q_n[j] = Q_{n-1}[j] + \begin{cases} S[j - k] & \text{if success} \\ F[j - k] & \text{if failure} \end{cases} \quad (10)$$

where $T = T_j$ and $x = x_k$. In other words, the quest function can be obtained after each trial by simply displacing, by k values, a table of values of the success or failure function and adding it to the existing table of values of the quest function. The tables of success and failure functions should, of course, be calculated prior to the experiment.

Termination rules. A natural termination rule is to stop when a confidence interval for the location of threshold is smaller than a specified size. This interval may be constructed in the following way. The quest function, minus the \log_e of the prior density, is the \log_e of the likelihood function of T . Calling this function $L(T)$,

$$L_n(T) = Q_n(T) - Q_0(T). \quad (11)$$

Let T_{like} be the location of the maximum of this function, which is the maximum likelihood estimate of T . To test whether any particular value of T , say T_j ,

is an acceptable estimate of threshold, we may use the ratio of likelihoods of this constrained hypothesis, $T = T_j$, and an unconstrained hypothesis. In the latter hypothesis, we can replace the unconstrained parameter by its maximum likelihood estimate, T_{like} (Hoel, Port, & Stone, 1971). Under certain regularity conditions, minus two times the \log_e of the likelihood ratio is asymptotically chi-square, with degrees of freedom (df) equal to the difference in the number of parameters in the two hypotheses (Wilks, 1962). In this case, there is 1 df. Therefore, we can reject any hypothesis $T = T_j$ for which

$$[L(T_{\text{like}}) - L(T_j)] \geq .5 \chi^2_{\alpha, df}, \quad (12)$$

where α is the confidence level and $df = 1$. This allows us to construct an interval outside of which the hypothesis $T = T_j$ is rejected. We terminate when the interval is less than a certain size. Note that to implement this rule only a modest number of subtractions and comparisons are required. If $L(T)$ is assumed to be monotonic on either side of its peak (usually a safe bet), then one need only perform the test at the bounds of the desired interval. This rule observes the convention adopted in the Priors and Prejudice section above: no prior information is used in the decision on when to stop.

Another possible termination rule is to stop after a fixed number of trials. While this may sacrifice something in efficiency, it has the advantage of simplicity and fits well with the block structure of conventional psychometric experiments.

The final estimate. The final estimate of threshold produced by QUEST is T_{like} , the location of the peak of $L(T)$, following the last trial. This is the maximum likelihood estimate of threshold based upon all the data. Confidence limits about the final estimate may be constructed by the method described above. Note that both the final estimate and confidence limits make no use of prior information.

Unbounded estimates. If an experiment consists of a fixed number of trials, there is a small but finite probability that the likelihood function will fail to have a single peak, and that the range over which it is at its maximum will include one or the other endpoint of the range of possible intensities. This outcome will occur, for example, when every trial in the session is correct. When this occurs, the estimate should be discarded and a new experiment conducted. It should be noted that a maximum likelihood estimate can always be unbounded, no matter how the data were collected.

An example

Many of the ideas in the preceding discussion may be clarified through an example. To this point, we have tried to be general, particularly as regards the

experimental task and the assumed psychometric function. Here we consider a specific task, two-alternative forced choice, and a specific psychometric function based on the Weibull (1951) distribution.

In the remainder of this paper, we will express intensity in decibels (dB), where 1 dB is a factor of $10^{1/20}$. Our reasons for using this scale will become evident below. The Weibull psychometric function may be written

$$w_T(x) = 1 - (1 - \gamma) \exp[-10^{(\beta/20)(x - T + \epsilon)}]. \quad (13)$$

The parameter γ specifies the probability of a success at zero intensity: for two-alternative forced choice it is .5, for n -alternative forced choice it is n^{-1} ; for yes-no, it is the false alarm rate. The parameter β specifies the "slope" of the psychometric function; this may depend somewhat on procedure and conditions (Massof, 1981; Nachmias, 1981), but for two-alternative forced choice it may be set to a value of 3.5. The parameter ϵ is introduced so that T will be the ideal testing point. As noted in A Placement Rule above, this is the point on the psychometric function at which the ideal sweat factor is least. For two-alternative forced choice ($\beta = 3.5$, $\gamma = .5$), ϵ should be 1.5 dB; for yes/no ($\beta = 3.5$, $\gamma = 0$), ϵ should be 1.15 dB. For any other case, ϵ can be determined empirically. With these settings, threshold intensity (T) gives probabilities of success of .920 and .796, respectively. Note that β , γ , and ϵ are not free parameters, and must be chosen before the start of the experiment.

Defining T more conventionally as the 75% correct point for two-alternative forced choice (instead of 92%) or the 50% point for yes/no (instead of .796) would increase the sweat factor by a factor of 1.8 or 1.4, respectively.

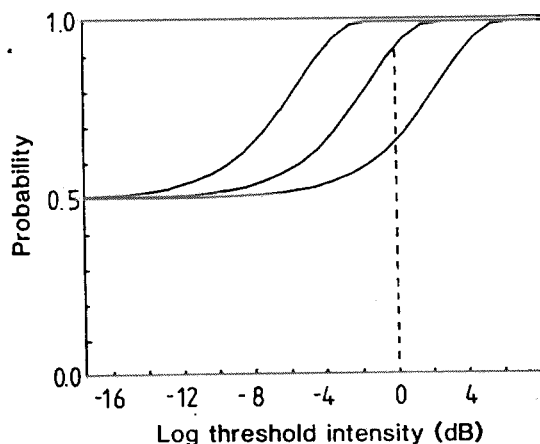


Figure 1. Three examples of a Weibull psychometric function. Center curve is the canonical form, flanking curves are for thresholds of -4 and 4 dB. The parameters of all functions are: $\beta = 3.5$, $\gamma = .5$, and $\epsilon = .01$.

Table 1
Table of Success and Failure Functions for a
Weibull Psychometric Function

x (in Decibels)	$\Psi(-x)$	$S(x) =$ $\ln \Psi(-x)$	$F(x) =$ $\ln[1 - \Psi(-x)]$
> 17	.500*	-.693*	-.693*
16	.501	-.690	-.696
15	.502	-.689	-.697
14	.503	-.687	-.700
13	.505	-.684	-.703
12	.507	-.679	-.708
11	.511	-.672	-.715
10	.516	-.662	-.726
9	.524	-.647	-.742
8	.535	-.625	-.766
7	.552	-.595	-.802
6	.575	-.553	-.856
5	.608	-.497	-.937
4	.653	-.426	-1.058
3	.710	-.342	-1.240
2	.779	-.249	-1.511
1	.853	-.159	-1.916
0	.920	-.084	-2.523
-1	.968	-.033	-3.432
≤ -2	.990*	-.010*	-4.605*

Note—Parameters are $T = 0$, $\beta = 3.5$, $\gamma = .5$, $\delta = .01$, and $c = 1.5$.
*Truncated at $\Psi(x) = \gamma$ or $1 - \delta$.

The interesting theoretical properties of the Weibull psychometric function have been reviewed elsewhere (Green & Luce, 1975; Nachmias, 1981; Quick, 1974); here we note only that it is an excellent empirical description of psychometric data in a wide variety of conditions (Nachmias, 1981; Watson, 1979).

However, this function is inaccurate in one respect. It says that the probability of a failure well above threshold is vanishingly small. But, in real life, observers occasionally make errors regardless of how intense the stimulus. The possibility of these blinks, "finger mistakes," or whatever can be accommodated by specifying an upper asymptote, $1 - \delta$, to the probability of a success. The assumed psychometric function then becomes

$$p_T(x) = \min[1 - \delta, w_T(x)]. \quad (14)$$

The parameter δ , like β and γ , must be chosen before the experiment, and should be set to the estimated probability of a failure well above threshold. We have typically used a value of .01. Figure 1 shows three examples of the resulting psychometric function, one with threshold 0 dB (the canonical form) and the other two with thresholds of -4 and 4 dB.

It is clear from the relationship between the quantities x and T in Equation 13 and from examination of Figure 1 that this psychometric function is translation invariant on a log intensity axis, and therefore satisfies Assumption 1.

Figure 1 shows that, with a β of 3.5, the psychometric function varies significantly over a range of only about 20 dB of intensity. Outside this range,

it is essentially constant. Simulations suggest that it is efficient to measure at intervals of about 1/20th of this range, or 1 dB. With this grain, the success and failure functions can be specified by tables of 20 numbers each, outside the bounds of which each is equal to the value at the bound. These tables are given in Table 1 and are plotted in Figure 2. Note that the size of the range of variation is inversely related to β . For some other β , the grain should be scaled accordingly.

Figure 2 illustrates the accumulation of the quest function over a simulated sequence of trials. Panel A shows the success and failure functions. Note that the failure function has much larger variations than the success function. This means that a failure tells us

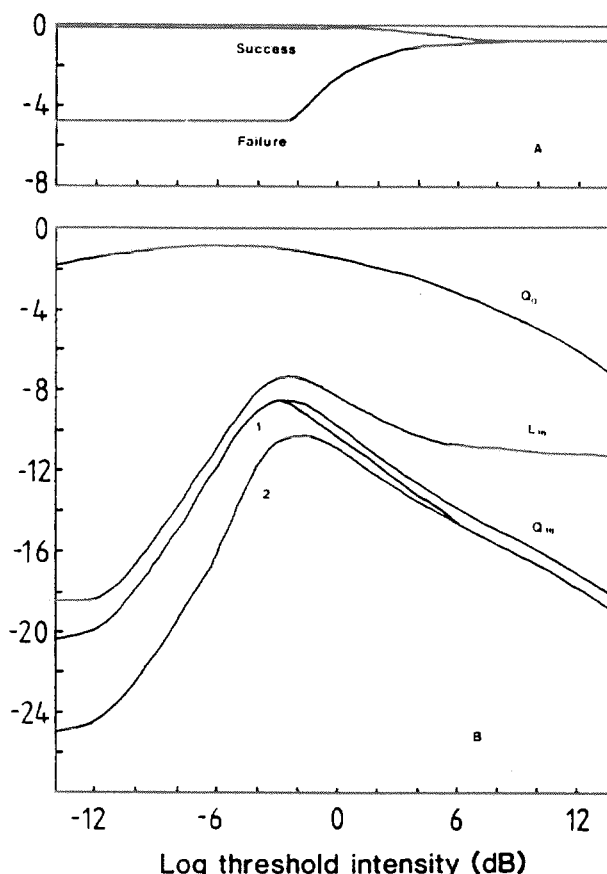


Figure 2. Construction of the quest function. The abscissa denotes possible values of T in steps of 1/20th of a log₁₀ unit (1 dB). Panel A shows the success and failure functions. In panel B, Q_0 is the quest function at trial 0, equal to the log₁₀ of the prior density. In this case, the prior density was specified as a normal distribution with a mean (T_{prior}) of -6 dB and a standard deviation of 8 dB. Curve Q_{16} is the quest function after trial 16. Curve L_{16} is the log₁₀ of the likelihood function of T following the 16th trial, and is equal to the difference between Q_0 and Q_{16} . The curve labeled 1 is the quest function after a successful trial 17 at an intensity of -3 dB. It is obtained by shifting the success function to the left by 3 dB and adding it to Q_{16} . Curve 2 is the quest function after a failure on trial 17, obtained by shifting and adding the failure function to Q_{16} .

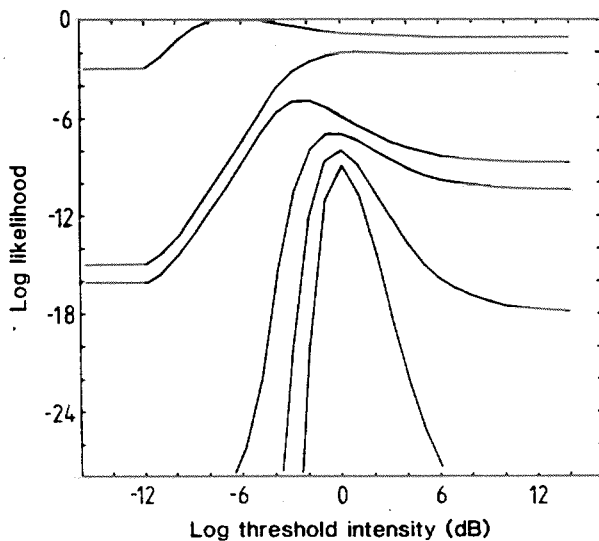


Figure 3. The log likelihood function $L(T)$ following various numbers of trials, indicated on the right. Data are from the same simulated experiment shown in Figure 2. The true threshold is at 0 dB. The curves have been displaced vertically by arbitrary amounts.

more about the location of threshold than does a success. This makes sense, since, in a two-alternative forced-choice experiment, a correct trial is quite probable above or below threshold, while an incorrect response above threshold is quite improbable. In fact, if not for the assumption that observers occasionally make mistakes well above threshold ($\delta = .01$), the failure function would rapidly fall to minus infinity, causing some difficulty in the subsequent computations. This is a further practical reason for introducing the parameter δ .

We suppose that the experimenter has supplied a Gaussian prior density with a standard deviation of 8 dB and a mean or mode (T_{prior}) of -6 dB. The uppermost curve in panel B shows the \log_e of this density, which also serves as the quest function before the

first trial, $Q_0(T)$. The quest function following the 16th trial is shown by the curve labeled Q_{16} . Subtracting $Q_0(T)$ from $Q_{16}(T)$, we get the curve labeled L_{16} , which is the log likelihood function following the 16th trial.

We now apply the likelihood ratio termination rule discussed earlier. We note that one half the chi-square criterion for .975 confidence and 1 df is 2.51. Therefore, we look at the width of $L_{16}(T)$ a distance 2.51 down from its peak. This yields an interval of about 8 dB. If this is greater than a preestablished termination interval, we perform a 17th trial. The 17th trial is placed at T_{post} , the peak of the quest function, which occurs at an intensity of -3 dB. If the trial is a success, the new quest function (labeled curve 1) is obtained by shifting the success function horizontally by -3 dB and adding it to the old (Q_{16}). If the trial is a failure, the failure function is shifted and added to Q_{16} , resulting in curve 2. The experiment would continue in this way until the confidence interval was less than a specified size.

Figure 3 gives another view of the same simulated experiment. The curves show the function $L(T)$ following 4, 8, 16, 32, 64, and 128 trials. With increasing numbers of trials, the maximum likelihood estimate of T moves closer to the true value (0), and the confidence interval about this estimate grows narrower. Note that the upper bound is always less abrupt than the lower bound.

A Program

To make the mechanics of the QUEST procedure as clear as possible, we provide a small BASIC program in Figure 4 which should run without modification on most computers. The variable S is the standard deviation of the prior density, and D , G , and E are δ , γ , and ϵ , respectively. The variable B is $\beta/20$, and M is the number of trials. In lines 110-120, X represents intensity (in dB) relative to the true threshold. "LOG()" in line 120 represents $\log_e()$. Throughout the program, "T" represents possible values of

```

100 DIM P(80),Q(40),Q0(40),S(1,80) \ N=20 \ N2=2*N \ S=12 \ D=.01 \ G=.5 \ B=3.5/20 \ E=1.5 \ M=32
110 FOR X=-N2 TO N2 \ P(N2+X)=1-(1-G)*EXP(-10^(B*(X+E))) \ IF P(N2+X)>1-D THEN P(N2+X)=1-D
120 S(0,N2-X)=LOG(1-P(N2+X)) \ S(1,N2-X)=LOG(P(N2+X)) \ NEXT X
130 FOR T=-N TO N \ Q0(N+T)=-(T/S)^2 \ Q(N+T)=Q0(N+T) \ NEXT T
140 PRINT "Prior estimate (+/-";S;"dB)"; \ INPUT T0 \ PRINT "Actual threshold"; \ INPUT T1
150 FOR K=1 TO M
160 GOSUB 220 \ R=INT(P(N2+X+T0-T1)+RND(0)) \ PRINT "Trial";K;"at ";X+T0;"dB has response";R
170 FOR T=-N TO N \ Q(N+T)=Q(N+T)+S(R,N2+T-X) \ NEXT T
180 NEXT K
190 FOR T=-N TO N \ Q(N+T)=Q(N+T)-Q0(N+T) \ NEXT T
200 GOSUB 220 \ PRINT "Maximum likelihood estimate is ";X+T0;"dB"
210 STOP
220 X=-N \ FOR T=-N TO N \ IF Q(N+T)>Q(N+X) THEN X=T
230 NEXT T \ RETURN

```

Figure 4. Small BASIC program to implement QUEST procedure.

threshold (in decibels) relative to the prior guess. "T0." In lines 160-170, "X" is the Bayesian estimate of threshold relative to "T0." The possible settings run from T0-N to T0+N in steps of 1 dB. In line 160, an observer with a threshold of T0 is simulated. In an actual experimental program, this instruction would be replaced by stimulus presentation and collection of a response from a real observer. The program has not been optimized in any particular way, and a real experimental program would doubtless be expanded to accommodate several stimuli, estimate confidence intervals, apply termination rules, and so on.

Some Extensions

Here we note some possible extensions, generalizations, or amendments to the QUEST procedure that may be useful in some circumstances.

Final estimates. In the example given above, the log intensity axis was specified with a grain (Δx) of 1 dB. Thus, each estimate of T will be specified only to the nearest decibel. This may be accurate enough for the placement of trials, but not for the final estimate. Two remedies are available. One is to reduce the grain at certain times during the experiment, for example, between blocks. However, this requires that the quest function be evaluated for all newly introduced points. A second and simpler alternative is to use the data, after the completion of the experiment, to derive a maximum likelihood estimate of T that is not constrained to any particular grain. Watson (1979) describes a simple estimation procedure for use in this situation.

Relative placement rules. Often we wish to place trials not *at* threshold, but somewhere *relative* to threshold. In these cases, QUEST can provide a best estimate of threshold relative to which the trial may be placed. For example, Levitt (1971) notes that to estimate the "slope" of the psychometric function, trials are best placed somewhat above and below T. As a second example, we note that even when estimating T, observers may become unhappy if every trial is placed at threshold. For this reason, we often place each trial at T_{post} plus a random value uniformly distributed over an interval of plus and minus 2 dB. This jitter reduces efficiency a little, but makes the subject happier.

Rating scales. In principle, allowing more response types should reduce the number of trials required to obtain a given variance in the estimate of T. For example, QUEST might be applied to a rating scale experiment with n rating categories. Each type of response (r_i , $i = 1, \dots, n$) has a different psychometric function, $p_{r_i, T}(x)$, all of which must sum to one at each intensity.

$$\sum_{i=1}^n p_{r_i, T}(x) = 1 \quad \text{for all } x,$$

Shape invariance is assumed for each psychometric function. Equations 2-6 are valid for any number of response types, and may be applied to this case in a straightforward way.

Variable β . If considerable uncertainty exists as to the value of β in a particular condition, this parameter may be allowed to vary over a modest number of fixed values. This is accomplished by concurrently maintaining several quest functions, each associated with a particular value of β and its own success and failure functions. After each trial, the quest function with the largest value identifies the most probable of the several possible values of β .

Discussion

Taylor and Creelman (1967; Taylor, 1971) have developed a measure which allows the performance of different psychometric procedures to be compared. They define asymptotic efficiency as

$$\frac{1}{n\sigma_n^2} \left\{ \frac{p(x)[1-p(x)]}{[dp(x)/dx]^2} \right\}_{x=T}, \quad (15)$$

where σ_n is the standard deviation of the estimate after n trials. However, this is an absolute measure of efficiency for which a value of 1 would be achievable and unsurpassable only in the limiting case, where trials are placed so close to T that the psychometric function may be treated as a straight line through $p(T)$. In practice, most experiments begin (and may finish) well before that point. Consequently, application of this measure to a practical adaptive procedure yields an "efficiency" which depends upon the number of trials collected and the accuracy of the prior estimate. It will also depend upon how well the assumptions underlying the procedure match the true state of affairs. With these reservations, we note that when the simulated psychometric function is identical in shape to the assumed psychometric function (Table 1), and the initial standard deviation of T is 6 dB, the QUEST procedure yields an efficiency of 84% after 128 trials. This compares favorably with an efficiency of around 45% for the PEST procedure after 128 trials with unspecified method and initial variance (Taylor & Creelman, 1967) and around 71% after 128 trials for Hall's (1981) hybrid procedure simulated with four-alternative forced-choice and 0 initial variance.

Ultimately, the best test of a procedure is actual use. We have used the QUEST procedure in a wide variety of psychophysical experiments for over 3 years. We have found it to be flexible, reliable, efficient, and easy to use.

REFERENCE NOTE

1. Klein, S. *Rapid determination of the psychometric function.* Paper presented at the annual meeting of the American Academy of Optometry, Orlando, Florida, December 1981.

REFERENCES

- CORNSWEET, T. N. The staircase-method in psychophysics. *American Journal of Psychology*, 1962, 75, 485-491.
- GREEN, D. M., & LUCE, R. D. Parallel psychometric functions from a set of independent detectors. *Psychological Review*, 1975, 82, 483-486.
- HALL, J. L. Maximum-likelihood sequential procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America*, 1968, 44, 370. (Abstract)
- HALL, J. L. Hybrid adaptive procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America*, 1981, 69, 1763-1769.
- HOEL, P. G., PORT, S. C., & STONE, C. J. *Introduction to statistical theory*. Boston: Houghton Mifflin, 1971.
- LEVITT, H. Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 1971, 49, 467-477.
- MASSOF, R. W. Wavelength dependence of the shape of the foveal absolute threshold probability of detection functions. *Vision Research*, 1981, 21, 995-1004.
- NACHMIAS, J. On the psychometric function for contrast detection. *Vision Research*, 1981, 21, 215-223.
- PAVEL, M. A new adaptive method for forced-choice experiments. *Journal of the Optical Society of America*, 1981, 71, S48. (Abstract)
- PENTLAND, A. Maximum likelihood estimation: The best PEST. *Perception & Psychophysics*, 1980, 28, 377-379.
- QUICK, R. F. A vector magnitude model of contrast detection. *Kybernetik*, 1974, 16, 65-67.
- ROUFS, J. A. J. Dynamic properties of vision—VI. Stochastic threshold fluctuations and their effect on flash-to-flicker sensitivity ratio. *Vision Research*, 1974, 14, 871-888.
- TAYLOR, M. M. On the efficiency of psychophysical measurement. *Journal of the Acoustical Society of America*, 1971, 49, 505-508.
- TAYLOR, M. M., & CREELMAN, C. D. PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 1967, 41, 782-787.
- WATSON, A. B. Probability summation over time. *Vision Research*, 1979, 19, 515-522.
- WATSON, A. B., & PELLI, D. G. The QUEST staircase procedure. *Applied Vision Association Newsletter*, 1979, 14, 6-7.
- WEIBULL, W. A. A statistical distribution function of wide applicability. *Journal of Applied Mechanics*, 1951, 18, 292-297.
- WETHERILL, G. B., & LEVITT, H. Sequential estimation of points on a psychometric function. *British Journal of Mathematical and Statistical Psychology*, 1965, 18, 1-10.
- WILKS, S. S. *Mathematical statistics*. New York: Wiley, 1962.

(Manuscript received April 22, 1982;
accepted for publication September 30, 1982.)