Some performance tests of QUEST on measurements of vibrotactile thresholds

DONALD LAMING
University of Cambridge, Cambridge, England

and

DAVID MARSH
University of Manchester, Manchester, England

A development of Watson and Pelli's QUEST procedure is described and illustrated with some recent measurements of vibrotactile thresholds. This enhanced QUEST procedure terminates as soon as some prescribed precision is attained in the threshold estimate and offers a choice of three models of the psychometric function to accommodate different stimulus configurations. It is validated here by (1) comparing the model of the psychometric function on which the calculations are based with the frequencies of detecting each of a well-chosen set of stimuli; and (2) comparing the variability of terminal estimates from QUEST runs started at different stimulus values with the intended precision. When detecting a vibration of 25 Hz, with the contactor set within a rigid surround, this new version of QUEST performs exactly as statistical theory says it should; but at 250 Hz and at 25 Hz without a surround, there are instrumentation problems connected with the coupling between contactor and finger.

Our version of QUEST was developed to meet a specific clinical problem:

Suppose you are so unfortunate as to put your hand through a plate glass window and suffer serious cuts to the wrist and forearm. In such circumstances, severance of the median or ulnar nerves, which innervate the fingers, is not uncommon. One of us (D.M.) is an orthopedic surgeon interested in the repair of that particular injury. After suture, the nerve fibers regenerate, but that regeneration is slow and always less than perfect. To monitor the patient's recovery, it would be helpful to have some accurate assessment of tactile sensibility—much more accurate than existing methods afford. Such an assessment would provide guidance for rehabilitative therapy and also evaluation of the efficacy of different surgical techniques. The measurement of vibrotactile thresholds is a candidate for this role.

A test procedure for clinical patients needs to be brief, especially as a series of tests are often required. We therefore developed Watson and Pelli's (1983) QUEST procedure in two ways—developments which they themselves suggested but evidently did not implement:

1. We added a simple termination rule that brings the procedure to a close as soon as the estimate of the threshold value achieves some prescribed precision. Different accuracies might be needed on different tests, and the prescribed precision can be specified according to the requirements of each test.

2. We incorporated a choice of three different models of the psychometric function. For efficient and reliable operation, it is essential that the model used accord well with the state of nature, and it happens that different stimulus configurations require different models. In visual experiments, for example, the psychometric function assumes a different shape depending on whether it relates to a discrimination between two separate luminances, the detection of a single flash added to a uniform background, or the detection of a sinusoidal grating; these differences in stimulus configuration are reflected in other senses as well (Laming, 1986, 1987).

To validate QUEST, we first discovered the form of the natural psychometric function in an experiment of traditional design. We then used QUEST many times over to check that it converged to the stimulus value prescribed by that natural function with the intended precision, neither more nor less. These checks were carried out using real subjects detecting vibrotactile stimuli and therefore tested these three premises in combination: (1) The QUEST algorithm behaves as statistical theory says it should. (2) (Asymptotic) statistical theory is applicable to the small numbers of trials to which a practicable clinical test must be limited. (3) The model psychometric function used adequately represents the state of nature. Most existing evaluations of adaptive procedures (e.g., Hall, 1981; Taylor & Creelman, 1967; Watson & Pelli, 1983) have used computer simulations with greater numbers of trials than we have here and premises (2) and (3) are untried.

Our computer program incorporates routines for driving a vibrator. However, those routines can be replaced by drivers for other kinds of apparatus, and our proce-

dure can be applied to any kind of sensory discrimination in which the presentation of the stimuli is amenable to control by computer. Again, although we address a clinical application in which the testing needs to be brief but ordinarily of no great precision, our procedure is equally suited to experimental use, since the prescribed precision is at the operator's disposal. In experimental use, it delivers threshold estimates of a *known* precision, which is of particular usefulness in evaluating the fit of a model to the data.

THE QUEST PROCEDURE

We first describe our procedure in general terms. It is substantially the procedure of Watson and Pelli (1983), but we think a clearer description is desirable.

On each of a series of trials, there are two observation intervals, marked by some extraneous stimulus, in one of which a stimulus of magnitude x is presented. The subject's task is to pick the observation interval that contains the stimulus. His choice is recorded by a computer, which then selects a possibly different value of x for the next trial. We *assume* that the relation between x and the probability of a correct choice in this two-alternative forced-choice (2AFC) task is of *known form*, $p(z)$, subject only to an unknown scale constant, a; that is,

$$P_{2AFC}(x) \equiv p(ax). \qquad (1)$$

Our objective is to estimate the scale constant a with maximum efficiency, and this is accomplished by selecting for each trial that value of x that, in the light of all the accumulated responses, is estimated to yield the maximum increment in the precision (measured as the inverse of the variance) with which a can be estimated.

Statistical Theory

Suppose that at some interim stage there have been n trials for which values of x equal to x_i, $i = 1, 2, \ldots n$, respectively, were selected. We denote the responses on those trials by R_i, $i = 1, 2, \ldots n$, putting $R_i = 1$ if the response on trial i was correct and 0 if it was an error. The likelihood of observing that sequence of responses, given the particular sequence of stimulus values and conditional on a particular value of a, can be expressed in the form

$$L(a) = \prod_{i=1}^{n} [p(ax_i)]^{R_i}[1 - p(ax_i)]^{(1-R_i)}. \qquad (2)$$

The most plausible value of a at this interim stage is the value that makes $L(a)$ a maximum (the *maximum-likelihood* estimate). It is conventionally calculated by setting the differential of the logarithm of the right-hand side of Equation 2 to zero and solving the resultant equation. The logarithm of $L(a)$ takes the form

$$\ln L(a) = \sum_{i=1}^{n} \{R_i \ln p(ax_i) + (1-R_i)\ln[1-p(ax_i)]\}, \qquad (3)$$

and we remark that it is the *sum of identifiable contributions from individual trials.* We denote the maximum-likelihood estimate by $\hat{a}$.

Statistical theory (e.g., Wilks, 1962, chap. 12) tells us that as the number of trials increases, $\hat{a}$ converges to the true value of a, that is, $\hat{a}$ is a *consistent* estimator (provided that the assumed form of the function $p(z)$ accurately represents the state of nature). In addition, the variance of $\hat{a}$ with respect to its true value is approximately the inverse of minus the expectation of the second differential of $\ln L(a)$, evaluated at $a = \hat{a}$. Differentiating Equation 3 twice with respect to $\ln a$ (ln for a reason which will appear) and taking expectations with respect to the R_i gives

$$[\text{Var}(\ln \hat{a})]^{-1}$$

$$= \sum_{i=1}^{n} [\hat{a}x_i p'(\hat{a}x_i)]^2/\{p(\hat{a}x_i)[1-p(\hat{a}x_i)]\}^2, \qquad (4)$$

so that the *precision* of the QUEST estimate is also the sum of identifiable contributions from individual trials. This inverse variance was proposed by Fisher (1922) as a measure of the information in the data (the $\{R_i\}$) about a. It provides the means of monitoring the increasing precision of the QUEST estimate and of terminating the procedure as soon as some desired accuracy has been attained.

It is obviously desirable to set each stimulus at the value that will result in the greatest increment to the inverse variance. This means selecting, as the stimulus on trial $n+1$,

$$x_{n+1} = z_*/\hat{a}_n, \qquad (5)$$

where $\hat{a}_n$ is the interim maximum likelihood estimate after the nth trial and z_* is the argument that makes the contribution per trial to the inverse variance, that is,

$$[zp'(z)]^2/\{p(z)[1-p(z)]\}, \qquad (6)$$

a maximum. This contribution is the inverse of Taylor and Creelman's (1967) "ideal sweat factor," and z_* is the argument of the psychometric function at which the ideal sweat factor reaches its minimum value. We call z_* the *sweet point* of the psychometric function, and it is illustrated for each of our models in Figure 1. In this way, each trial is made to yield the maximum information about the value of a.

Carrying Out the Calculations

The value of a that maximizes Equation 2 has to be calculated afresh after each trial. Except in the simplest cases, the solution of maximum-likelihood equations is a complicated business taking a far from negligible time. Its implementation on a microcomputer within an experiment conducted in real time would not be practicable *except* for the simplifications that follow from an ingenious trick introduced by Watson and Pelli (1983).

Following their lead, we restrict our inquiry to a finite set of discrete stimulus values. So $\ln L(a)$ need be recal-
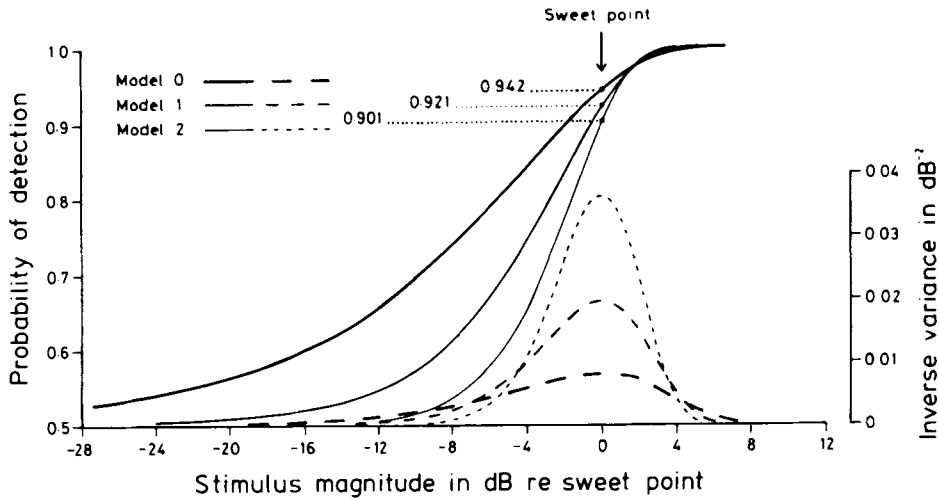
Figure 1. The detectability functions (continuous curves) and the increments in inverse variance (broken curves) as functions of stimulus magnitude in relation to the sweet point.

culated for a finite set of a only and its maximum quickly found by simple exploration. We also work in a *logarithmic* metric. If the psychometric function is expressed with respect to the logarithm of its argument,

$$p(ax) \equiv \psi(\ln a + \ln x), \tag{7}$$

dissociating the scale constant and the physical stimulus magnitude. It is for this reason that $L(a)$ is differentiated with respect to $\ln a$ in Equation 4, and all our calculations of posterior probability and variance will be expressed in terms of $\ln a$.

Suppose now that a geometric ladder of stimulus values is selected, a discrete set of the form

$$x_k = x_0 e^{k\Delta}, \tag{8}$$

where k is an integer which may take negative as well as positive values. This is a mild restriction in practice, especially as computer control of stimulus magnitude implies a restriction to a finite digital set to which it is often convenient to give a geometric structure. Such a set is equally spaced in our logarithmic metric. We seek the member of this stimulus set that is closest to the sweet point of the psychometric function; when that sweet point is known in physical stimulus units, all other statistics can be calculated.

Searching for the sweet point in this way, we implicitly restrict our inquiry to a similar geometric ladder of values of the scale constant

$$a_k = z_s/x_k = (z_s/x_0)e^{-k\Delta}. \tag{9}$$

The calculation of likelihood ratio therefore takes this especially simple form:

For each stimulus x_k and each value of the scale constant $a_{k'}$,

$$p(a_{k'}x_k) \equiv \psi[\ln z_s + (k-k')\Delta], \tag{10}$$

which depends on the stimulus presented and on the candidate value of a only via the scalar variable $(k-k')$. Substituting this expression into Equation 3 gives

$$\ln L_n(a_{k'}) = \sum_{i=1}^{n} \{R_i \ln[\psi(\ln z_s + (k_i-k')\Delta)] + (1-R_i)\ln[1 - \psi(\ln z_s + (k_i-k')\Delta)]\}. \tag{11}$$

In this equation, $a_{k'}$ is a candidate value of a, and k_i indexes the stimulus presented on trial i.

The corresponding value on trial $n+1$, $\ln L_{n+1}(a_{k'})$ differs from Equation 11 simply by the addition of one further term. That term is

$$\ln[\psi(\ln z_s + (k_{n+1}-k')\Delta)] \tag{12a}$$

if R_{n+1} is correct and

$$\ln[1 - \psi(\ln z_s + (k_{n+1}-k')\Delta)] \tag{12b}$$

if R_{n+1} is wrong. These decrements are *calculated in advance of the experiment* and $\ln L_n$ is calculated cumulatively for all candidate values of a. The part of the calculation of $\ln L_n$ that has to be done in real time involves only additions and subtractions with respect to a finite vector of values—easily within the capability of a microcomputer. This is the very ingenious trick introduced by Watson and Pelli (1983).

The same trick works for the calculation of the inverse variance. Equation 4 shows that that quantity also decomposes into a sum of contributions from individual trials; and after substitution of Equation 10, those contributions can be seen likewise to depend only on the scalar variable (k_i-k'). This vector of increments to the inverse variance is also calculated in advance of the experiment.

Implementation

We now give some details of our particular computer program:

Prior to the experiment, we calculate three vectors covering a sufficient range (equivalent to about ±34 dB) of integer values $(k-k')$. These vectors list the decrements to $\ln L_n$ when the response is correct (Expression 12a) and when the response is wrong (Expression 12b), and the increment to the inverse variance (Expression 6, which is independent of the response). These vectors are stored in a disk file to be read later by the QUEST program. Since the step size Δ is often dictated by the apparatus employed (Δ is 1.54 dB for our vibrator), one set of prior calculations will serve for many experiments. But note that these prior calculations depend on the model of the psychometric function employed.

In the experiment itself, we first select an initial stimulus magnitude, x_0 in Equation 9, and a geometric range of stimuli extending equally above and below it. Starting from x_0, we present on each trial the member of this finite set that has the greatest calculated likelihood of being the stimulus that corresponds to the sweet point of $p(z)$. This is the x_k that, when multiplied by $\hat{a}_n$, most closely approximates z_s. We call the magnitude of this maximum-likelihood member the *QUEST value*. Following each trial, we update the values of $\ln L_n(a_{k'})$ for all the candidate $a_{k'}$, according to the outcome of the trial. We call this set of values the *QUEST vector*; it catalogs the likelihood attaching to each member of the stimulus set. We also increment the inverse variance vector, which catalogs the inverse variance attaching to each $a_{k'}$. Finally, before initiating the next trial, we test whether the inverse variance attaching to the new QUEST value exceeds the prescribed limit, in which case the QUEST procedure terminates.

The initial stimulus value is discovered by serial exploration (method of limits), which delivers a value within reasonable range of our objective. We then erect a prior distribution over our set of stimuli which has this function: Late in the QUEST procedure a single trial can have at most a small effect because of the accumulated weight of preceding trials. But at the outset, that weight of preceding trials is absent and a single aberrant response might displace the QUEST value a long way in a contrary direction, from which recovery will be slow. In fact, in the absence of any prior distribution, the QUEST value would be undefined until at least one error and also one correct response had been recorded. Our prior distribution constitutes an initial QUEST vector and limits the possible movement of the QUEST value in the first few trials, the more so the smaller its standard deviation. Although default values are supplied for both the standard deviation and the location of the prior distribution relative to the outcome of the serial exploration process, these parameters can be varied according to the application of the QUEST procedure. The mean of the prior distribu-
tion, which is calculated to be normal with respect to log stimulus magnitude, is taken as the initial QUEST value.

It is our present practice to record the stimuli presented and the responses uttered during each QUEST procedure in a file for further examination if required. At the termination of the procedure, our program displays the initial and terminal QUEST values and converts the terminal value into an estimate of the 75% threshold.

MODELS OF THE PSYCHOMETRIC FUNCTION

We provide a choice of three different models of the psychometric function derived from the theory of sensory discrimination by Laming (1986). These are by no means the only models one might use; the logistic and Weibull functions are common alternatives. But one or another of these three will provide an adequate approximation to each of those psychometric relations known, at present, to occur in nature.

Model 0. For two separate suprathreshold stimuli (e.g., two flashes of light of luminances L and $L+\Delta L$), the discriminability function is approximately a normal integral with respect to ΔL (Laming, 1986, p. 75). In Model 0, we take

$$p_0(z) = \Phi(z/\sqrt{2}), \quad z \geq 0, \qquad (13)$$

where $\Phi(\ )$ is the normal integral function. Its argument is here written as $z/\sqrt{2}$ in recognition of the fact that p_0 relates to a 2AFC task with the stimulus z presented in one observation interval only. The logistic function is an alternative to our Model 0.

Model 1. The detectability of an increment added to a uniform stimulus (e.g., a flash of light of size ΔL superimposed on a background luminance, L) typically varies as a normal integral with respect to the *square* of the increment (i.e., $(\Delta L)^2$; Laming, 1986, pp. 44-52). In this case, we take

$$p_1(z) = \Phi[2h(\tfrac{1}{2}z)/\sqrt{2}], \quad z \geq 0, \qquad (14)$$

where

$$h(z) = [\tfrac{1}{2}(1-1/\pi)]^{-\tfrac{1}{2}}[z(\Phi(z)-\tfrac{1}{2}) + \Phi'(z)-\Phi'(0)]$$
$$(15)$$

is a transfer function (Laming, 1986, p. 150). For z large,

$$h(z) \rightarrow [\tfrac{1}{2}(1-1/\pi)]^{-\tfrac{1}{2}}[\tfrac{1}{2}z-(2\pi)^{-\tfrac{1}{2}}], \qquad (16)$$

giving linear transmission, while for z small,

$$h(z) \sim [\tfrac{1}{2}(1-1/\pi)]^{-\tfrac{1}{2}} \cdot \tfrac{1}{2}(2\pi)^{-\tfrac{1}{2}}z^2. \qquad (17)$$

Model 2. The detectability of a sinusoidal grating increases even more steeply, approximately as a normal integral with respect to the fourth power of its contrast (Laming, 1986, pp. 52-57). In this case, we take

$$p_2(z) = \Phi\{2h[h(\tfrac{1}{2}z)]/\sqrt{2}\}, \quad z \geq 0, \qquad (18)$$

Table 1
Parameters of the Model Psychometric Functions

	Model 0	Model 1	Model 2
Sweet point*	2.228	3.886	5.878
Probability correct at sweet point	0.942	0.921	0.901
Inverse variance at sweet point	0.008	0.020	0.036
75% threshold*	0.954	2.509	4.509
Threshold factor	0.428	0.646	0.767

*In normalized stimulus units.

applying the transfer function $h(z)$ twice in succession. A Weibull function with an exponent somewhat less than 4 provides an alternative to our Model 2.

These three model functions, p_0, p_1, and p_2 are exhibited in Figure 1.

Table 1 lists certain parameters of our three models. The sweet point and the 75% threshold are given as normalized stimulus magnitudes, expressed in units of root-mean square noise level. The fact that these normalized values vary greatly from one model to another is of no consequence, since the internal noise level cannot be known independently. The probability correct at the sweet point, on the other hand, varies little from one model to another, being uniformly high. The reason is explained diagrammatically in Figure 1, which shows how the increment in inverse variance per trial (the inverse of Taylor's ideal sweat factor) varies with stimulus magnitude for each model function. That increment reaches a peak at a detection probability equal to 0.942, to 0.921, and to 0.901, respectively, because it varies chiefly as the inverse binomial variance of the detection response—an inverse variance which becomes infinite at a probability of one.

It is traditional to take "threshold" to be the stimulus value that affords 75% correct responses, and this 75% threshold is easily obtained from the estimate of the sweet point by multiplying by the *threshold factor* (the ratio of the 75% threshold to the sweet point). This factor varies greatly from one model to another. It applies equally to physical stimulus magnitude, and demonstrates the importance of using a model that accords with the state of nature. However, the criterion of 75% correct responses is an arbitrary choice, and one could equally well talk about a 70.7% threshold or an 84% threshold, and so on. In that terminology, our sweet point is a 94.2%, a 92.1%, or a 90.1% threshold, depending on the model in question.

To validate our QUEST procedure, there are two matters that need to be checked: (1) that the psychometric function in nature is adequately represented by the model used—and we also need to know which model to use in which circumstances—and (2) that when the correct choice of model is made, the QUEST value converges on the sweet point with the prescribed precision, neither more nor less.

PERFORMANCE TESTS

In the experiments that follow, we used a vibrator (Unilab vibration generator) driven by a BBC microcomputer, with frequency and amplitude selected under program control. The vibrator was externally damped by a stiff steel strap to the point that any additional damping contributed by the subject's finger could safely be ignored. The steel strap artificially increased the driving current, permitting current control of the amplitude of vibration. This method proved reliable outside the range of resonance (80-200 Hz), and all our experiments were conducted outside that range.

The subject placed his left index finger on a contactor 4 mm in diameter; when the contactor was fitted with a surround, it protruded about 0.3 mm through a hole of 5-mm diameter, leaving a gap of 0.5 mm all around. To stabilize the pressure between finger and contactor, the finger was covered with a small sandbag measuring 23 × 13 cm and weighing 300 g. The stimuli were switched on and off at a zero-crossing of the driving waveform, and their duration was 500 msec throughout. There was a similar interval between the two stimuli on each trial and a further 1,500 msec between each response and the onset of the following trial. Each trial cycle therefore took about 4 sec (allowing 1 sec for uttering the response).

We investigated three different stimulus configurations, chosen in the hope of stimulating different receptor populations:

(1) *250 Hz with a surround.* This stimulus is known from the comparison of human sensitivity with the responsiveness of different classes of nerve fiber to preferentially stimulate the Pacinian corpuscles, which are at their most sensitive at approximately this frequency (Sato, 1961; Talbot, Darian-Smith, Kornhuber, & Mountcastle, 1968).

(2) *25 Hz with a surround.* Verrillo (1963) has shown that this configuration preferentially stimulates some population other than the Pacinian.

(3) *25 Hz without a surround.* Gescheider, Sklar, Van Doren, & Verrillo (1985) argue that, other than the Pacinian, there are two distinct populations active at 25 Hz, of which one is specifically sensitive to the *gradient* of deformation between the moving contactor and the static surround (Verrillo, 1979). This configuration reduces that gradient to a minimum in an attempt to engage selectively yet another population of receptors.

It is important to both the accuracy and efficiency of the QUEST procedure that the particular model of the psychometric function used should accurately represent the increase in detectability with stimulus magnitude that actually occurs in nature. To test this first point, we proceeded as follows:

The QUEST procedure was first used to estimate the sweet point of the model function in physical stimulus units. We then selected six stimulus magnitudes from the

range available with our vibrator, calculated, in the light of our knowledge of the sweet point, to cover the greater part of the variation of detectability in our candidate psychometric function. Using these stimuli, we carried out a straightforward assessment of the model, comprising 200 2AFC trials at each magnitude. The stimulus was presented in the first observation interval in exactly half of these trials, equally divided between two sessions on successive days. Within each session, there was a total of 600 trials, randomly assigned to the different stimulus magnitudes and to the two orders of presentation.

The second point to be checked was that QUEST converges on the sweet point with a precision equal to that prescribed. We conducted seven QUEST runs with each of the three stimulus configurations on each of 5 different days. Each QUEST run in a set of seven started at a different initial value, either the (already known) sweet point or offset from the sweet point by ±6, ±12, or ±18 dB; these seven different start points were selected in random order. We used these values of the QUEST parameters: The prior distribution was erected about the chosen start point as mean, with a standard deviation of 10 dB, and the procedure was terminated as soon as the standard deviation of the estimate (not counting the contribution from the prior distribution) fell below 3 dB. The critical question was whether the 35 terminal QUEST values obtained for each stimulus configuration had a standard deviation equal to (that is, neither significantly greater nor less than) 3 dB. The two authors served as subjects.

Results

The tests at 25 Hz with a surround were satisfactory in every respect, with outcomes that accord nicely with statistical theory; the tests at 250 Hz and at 25 Hz without a surround were not. We present the results with the first stimulus in detail, but merely indicate how the results with the other two differed from the first.

25 Hz With a Surround

Psychometric function. The detection data are displayed in Figure 2. Each point shows the proportion of correct responses in 200 AFC trials; the curves are Equation 14 (Model 1) shifted laterally to best fit the data. Estimates of the 75% threshold, its standard error, and the goodness-of-fit statistic (2 ln likelihood-ratio) are set out in Table 2; both sets of data accord adequately with the model. We compared these data with the other two models as well. D.L.'s data accorded well with Model 2 (2lnλ = 4.409 with 5 df) but not with Model 0; D.M.'s data accorded with Model 0 (2lnλ = 9.764) but not with Model 2. Differences between threshold estimates from the two different sessions were small in relation to their respective standard errors.

Convergence. The mean of the 35 QUEST runs was 6.32 μm for D.L. and 1.81 μm for D.M. These are estimates of the sweet point (equal to the 75% threshold divided by the scaling factor in Table 1) and should be compared with estimates from the psychometric function
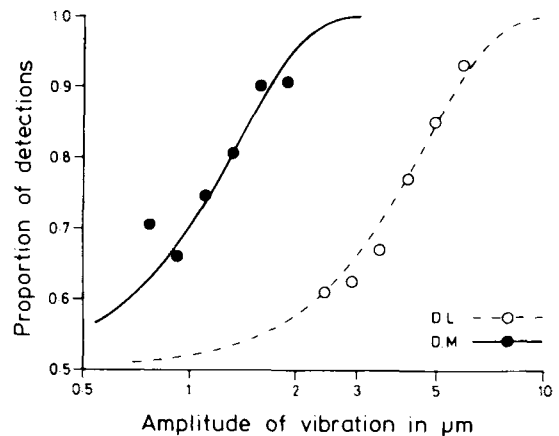

Figure 2. Detection data for 25 Hz with a surround. Each point shows the proportion of correct responses in 200 2AFC trials. The detectability functions are the best-fitting curves from Model 1.

data of 6.03 and 1.77 μm, respectively. The differences amount to 0.40 and 0.16 dB, whereas the theoretical standard error is 0.51 dB (3 dB divided by √35).

The standard deviations of the terminal QUEST values were 2.34 and 3.29 dB, respectively. Neither of these values differs significantly from 3 dB (at .05 on a two-tailed test).

Duration and bias in terminal value. Table 3 sets out the average number of trials to termination and the mean displacement of the terminal QUEST value as a function of the start point.

The average number of trials is increased when the start point is displaced. The additional trials are taken up with traveling to the neighborhood of the sweet point; and, because those additional trials are spent at uninformative stimulus values, they contribute little to the ultimate inverse variance. The additional number of trials is greater for a start point displaced above threshold than for one displaced below. The sweet point to which QUEST converges corresponds to about 11 responses correct out of 12. So the upward adjustment (from below) of the QUEST value produced by one error is substantially greater than the downward adjustment (from above) produced by a correct response.

The choice of start point does not create any appreciable bias in the terminal value. Such a bias is, in principle, to be expected because, although the prior distribution does not enter into the calculation of the inverse variance used to terminate the procedure, it does affect the QUEST value. Since the prior distribution had a standard deviation (in our tests) of 10 dB and those tests were

Table 2
Goodness-of-Fit Statistics for the Psychometric Function Data in Figure 2

Subject	75% Threshold	SE	2 ln λ	df
D.L.	6.631*	0.221*	2.806	5
D.M.	1.163	0.035	9.874	5

*Thresholds and standard errors in microns.

Table 3
Average Trials to Termination and Mean Displacement of
Terminal Value in Detection of 25 Hz With a Surround

	Displacement of Start Point (dB)						
	−18	−12	−6	0	+6	+12	+18
Average trials	24.8	21.7	18.9	17.2	17.4	24.8	38.9
Displacement of terminal value	−0.20	0.72	−1.12	−0.82	1.19	0.72	−0.51

terminated at an accuracy of 3 dB, the weight attaching to the prior distribution in the determination of the terminal value ought to be 0.08 [$= 10^{-2}/(10^{-2} + 3^{-2})$], giving a terminal displacement of 1.5 dB following a starting displacement of 18 dB. The terminal displacements in Table 3 seem to be no more than random error.

250 Hz With a Surround

The differences in performance with this stimulus vis-à-vis the 25-Hz stimulus are easily summarized: The detectability function appears a little steeper, and there are significant variations in sensitivity from session to session. These variations vitiate our test of QUEST.

Psychometric function. The data from one subject (D.L.) accord with Model 2 but not with Model 1; those from the other subject (D.M.) accord with Model 1 but not with Model 2 or Model 0. Moreover, there are highly significant differences in estimated threshold between the two sessions, which are large in relation to the standard errors.

Convergence. The mean of the 35 QUEST runs was 0.52 μm for D.L. and 0.49 μm for D.M. Calculation shows that the first mean is too small (by −10.23 dB) to match the threshold estimate from the psychometric function data, and that the second is too large (by 2.29 dB). In addition, the standard deviations of the terminal values were both greater than intended (4.37 and 6.19 dB, respectively); these differences are both highly significant.

Since, with this stimulus, QUEST does not converge satisfactorily, data on the number of trials taken and on residual bias in the terminal values are irrelevant. Although the failure to converge may have several contributory causes, the significant variation between sessions, and presumably within sessions as well, is obvious. For the present, we comment on that problem alone.

Vibration at 250 Hz is detected by means of the Pacinian corpuscles. These lie in the deeper tissues of the dermis and subcutaneous fat, which is fluid at body temperature. It is to be expected, therefore, that the coupling between the vibrating contactor and the receptors will depend on the force with which the pronated finger bears on the contactor. It is also to be expected that the subject will adjust his finger from time to time to obtain a better feel of the stimulus. Indeed, at times, the 250-Hz vibrations seemed to be most noticeable at a distance from the point of stimulation, at the base of the finger, for example. We think such changes in the coupling of between contactor and receptors are the chief source of variability in our measurements.

Craig and Sherrick (1969) have examined the variation of detection threshold with contactor area under three different loading conditions, constant penetration of the contactor, constant force, and constant pressure (force increasing in proportion to area). At 250 Hz, the empirical relation between area and threshold is quite different for these three conditions. We used constant penetration, as Verrillo has done, and that condition now appears unstable. That instability might be of small order on the forearm or even on the ball of the thumb; but we had a specific reason for using the fingertip. This is a problem that is specific to 250 Hz. Craig and Sherrick repeated their observations at 20 Hz and found no variation in threshold with area of contactor and no difference between the different loading conditions. That explains why our tests at 25 Hz were nevertheless satisfactory.

25 Hz Without a Surround

We think that with this stimulus we are simultaneously engaging two different populations of receptors with different psychometric functions. This introduces, or perhaps augments, a similar variability from session to session to that experienced with 250 Hz.

Psychometric function. The data from one subject (D.L.) accord with Model 0 (but not with Model 1 or Model 2), while the data from the other subject (D.M.) do not accord with *any* of our models. With both subjects, there are significant variations in threshold from session to session, which are large in relation to their standard errors.

Convergence. Since D.L.'s data accorded with Model 0, that model was used as the basis for the QUEST runs. For that subject, the standard deviation was 1.54 dB, which is only *half* of what was intended, significantly smaller at the low end of the chi-square distribution. For D.M., the standard deviation was 4.18 dB, which is significantly large; but if the day-to-day variation (which was itself significant) is excluded, that standard deviation falls to 2.23 dB, which is again significantly small. So the special feature of the present stimulus is a precision in excess of that built into the QUEST program.

We think we are simultaneously stimulating two receptor populations with different detectability functions. The population with the shallower function will be the more sensitive to weak stimuli and is preferentially engaged in a traditional psychometric function experiment. But QUEST focuses on stronger stimuli, around the sweet point which gives upwards of 90% detections, where the other population with the steeper detectability function

predominates. The number of trials that QUEST *takes* decreases as the model function becomes steeper, but the number of trials it actually *needs* depends on the steepness of the detectability relation obtaining in nature. If a QUEST procedure designed for a receptor population with a shallow detectability function actually engages some other population with a steeper function, it will run on for longer than is necessary with a consequent excess in precision.

Engaging two receptor populations simultaneously might well be a consequence of our experimental conditions since, without a surround, there is no control at all over the coupling between contactor and finger. In his efforts to obtain a yet better feel of the stimulus, the subject may, at times, have been pressing very hard on the contactor, hard enough to squeeze some of the Pacinian corpuscles in their fluid medium against the bone. The relative insensitivity of the Pacinian corpuscle to low-frequency stimulation has been shown to be largely due to its rigid lamellar capsule (Loewenstein & Mendelson, 1965). This may not be so marked for forcible *deformation* of the receptor as for simple *displacements*, which do not penetrate to the central axon terminal. Thus, the subjects may have discovered that, by pressing hard, they could get detectable activity in Pacinian afferents at stimulus levels as low as those needed to activate the non-Pacinian channel.

DISCUSSION

Using Model 1 to determine the threshold for 25-Hz vibration with a surround, our version of QUEST works just as we planned. This implies that the QUEST algorithm works as statistical theory says it should; that the asymptotic variance calculations are accurate to within the limits of our experimental error, even with the small numbers of trials we are looking at here; and that Model 1 adequately represents the detectability of our 25-Hz stimulus. When QUEST is started close to the sweet point (within ±6 dB), it takes about 18 trials. This means that a detection threshold can be measured to an accuracy of 3 dB (−30% to +40%) in about 72 sec-2 min at the most, allowing a reasonable time for the initial search procedure. A sequence of such tests would constitute, in our judgment, a practicable clinical diagnostic procedure, although greater accuracy, which will simply take longer, will be required for experimental purposes.

The purpose in developing QUEST for application to our clinical problem was to improve efficiency, but in this respect the outcome is disturbing. If every stimulus had been located exactly at the sweet point, the increment in inverse variance would have been 0.0197 dB^{-2} per trial, which is the peak value in Figure 1 of the inverse variance for Model 1. The number of trials then required to achieve an accuracy of 3 dB (total inverse variance 0.11 dB^{-2}) would have been 5.6. But, even when QUEST

was started at that sweet point, the average number of trials taken was 17.2, giving an efficiency of 32.5%. Compare the efficiencies of the psychometric function experiments: these were 50.5% and 62.7%, respectively, for D.L. and D.M.! Some improvement must surely be possible.

Since QUEST converges on the sweet point, its asymptotic efficiency is indeed 1, setting aside the loss due to the discrete ladder of stimulus values. Practical efficiencies are lower because the QUEST value wanders to less informative stimulus values, and examination of the data when QUEST is started at the sweet point shows that it wanders equally in either direction. Efficiency might be improved by restricting that wandering; this can be accomplished by reducing the standard deviation of the prior distribution, set, in our tests, at 10 dB. In short, QUEST needs to be tuned.

However, a more pressing problem is the day-to-day variability we observed with our other two stimulus configurations. In the first place, better instrumentation is needed to permit the control of applied force. That is a lesser problem in experimental research using the ball of the thumb, but our surgical application specifically requires stimulation of the fingertip. Second, fundamental research is needed as well to determine what parameter of the coupling between contactor and fingertip needs to be controlled—indentation, force, pressure, or what? We hope that when these problems are understood and mastered, the apparent stimulation of two receptor populations simultaneously at 25 Hz without a surround will disappear.

One final problem common to any procedure for measuring a threshold is that, given so many faint stimuli of which a certain proportion, of necessity, fail to be detected, the subject may temporarily forget what the stimulus feels like and thereby lose sensitivity. Our QUEST procedure converges on a stimulus value at which 90%–94% of the stimuli are correctly detected, so that for most of the time the subject is given stimuli that are relatively easy to sense. The problem that the subject may forget what the stimulus feels like is happily reduced to small proportions.

REFERENCES

CRAIG, J. C., & SHERRICK, C. E. (1969). The role of skin coupling in the determination of vibrotactile spatial summation. *Perception & Psychophysics*, **6**, 97-101.

FISHER, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London, Series A*, **222**, 309-368.

GESCHEIDER, G. A., SKLAR, B. F., VAN DOREN, C. L., & VERRILLO, R. T. (1985). Vibrotactile forward masking: Psychophysical evidence for a triplex theory of cutaneous mechanoreception. *Journal of the Acoustical Society of America*, **78**, 534-543.

HALL, J. L. (1981). Hybrid adaptive procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America*, **69**, 1763-1769.

LAMING, D. (1986). *Sensory analysis*. London: Academic Press.

LAMING, D. (1987). The discrimination of smell and taste compared with other senses. *Chemistry & Industry*, No. 1, 12-18.

LOEWENSTEIN, W. R., & MENDELSON, M. (1985). Components of receptor adaptation in a Pacinian corpuscle. *Journal of Physiology*, 177, 377-397.

SATO, M. (1961). Response of Pacinian corpuscles to sinusoidal vibration. *Journal of Physiology*, 159, 391-409.

TALBOT, W. H., DARIAN-SMITH, I., KORNHUBER, H. H., & MOUNT-CASTLE, V. B. (1968). The sense of flutter-vibration: Comparison of the human capacity with response patterns of mechanoreceptive afferents from the monkey hand. *Journal of Neurophysiology*, 31, 301-334.

TAYLOR, M. M., & CREELMAN, C. D. (1967). PEST: Efficient esti-

mates on probability functions. *Journal of the Acoustical Society of America*, 41, 782-787.

VERRILLO, R. T. (1963). Effect of contactor area on the vibrotactile threshold. *Journal of the Acoustical Society of America*, 35, 1962-1966.

VERRILLO, R. T. (1979). The effect of surface gradients on vibrotactile thresholds. *Sensory Processes*, 3, 27-36.

WATSON, A. B., & PELLI, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33, 113-120.

WILKS, S. S. (1962). *Mathematical statistics*. New York: Wiley.