PERGAMON

Vision Research 39 (1999) 2729–2737

Vision Research

Bayesian adaptive estimation of psychometric slope and threshold

Leonid L. Kontsevich *, Christopher W. Tyler

Smith-Kettlewell Eye Research Institute, 2232 Webster Street, San Francisco CA 94115, USA

Received 1 October 1997; received in revised form 10 August 1998

Abstract

We introduce a new Bayesian adaptive method for acquisition of both threshold and slope of the psychometric function. The method updates posterior probabilities in the two-dimensional parameter space of psychometric functions and makes predictions based on the expected mean threshold and slope values. On each trial it sets the stimulus intensity that maximizes the expected information to be gained by completion of that trial. The method was evaluated in computer simulations and in a psychophysical experiment using the two-alternative forced-choice (2AFC) paradigm. Threshold estimation within 2 dB (23%) precision requires less than 30 trials for a typical 2AFC detection task. To get the slope estimate with the same precision takes about 300 trials.

Keywords: Psychometric function; Two-alternative forced-choice; Bayesian adaptive method

1. Introduction

A majority of experimental psychophysics deals with measuring sensitivity thresholds that specify the intensity at which the stimulus is just detectable. The threshold is only a part of the story because the transition from non-detectability to detectability is not abrupt; it occurs over some finite intensity range. This transition is conceptualized by the psychometric function relating the probability of correct detection to the stimulus intensity. The slope of this function reflects the width of the transitional range; the threshold defines its absolute position along the intensity axis.

Most experimental studies avoid the issue of slope because its measurement is too laborious. This practice may be dangerous in studies that use the modern adaptive methods designed for a typical slope value of 2.8 when represented in log d' versus log contrast coordinates (which corresponds to an exponent of 3.5 in the Weibull approximation; Pelli, 1987a). Often the actual slopes deviate markedly from that assumed value (see, for example, Mayer & Tyler, 1986; Legge et al., 1987), which make the measurements less precise and subject to systematic errors. In avoiding the slope issue, many studies miss valuable information about the sensory processing such as transducer nonlinearities (Foley & Legge, 1981) and uncertainty effects (Pelli, 1985). There is need, therefore, for an experimental method that would efficiently measure the threshold for any slope and measure the slope value, if necessary. The present study introduces a new Bayesian adaptive method we call the Ψ method (to reflect its applicability to the psychometric function) that fits these requirements.

2. Rationale

Any adaptive method for estimating psychophysical parameters needs to address three major issues: estimation of the psychometric parameters (threshold and slope), the termination rule, and placement of the next trial.

2.1. Estimation

The most efficient way to get the threshold estimate from the results of the completed trials is to keep updating the posterior probability distribution for the sampled thresholds based on Bayes' theorem (Hall, 1968; Watson & Pelli, 1983). The best threshold esti-

* Corresponding author. Tel.: +1-415-5611793; fax: +1-415-5611610.

E-mail address: lenny@skivs.ski.org (L.L. Kontsevich)

0042-6989/99/$ - see front matter © 1999 Elsevier Science Ltd. All rights reserved.
PII: S0042-6989(98)00285-5

mate on any trial is the mean of this distribution because it minimizes the variance of the threshold estimate (Gelb, 1982, p. 103) and is proven to be more stable than the maximum rule (Emerson, 1986). To keep track of both thresholds and slopes, the posterior probability distribution has to be two-dimensional, where each *pair* of slope and threshold parameters is associated with some probability. This idea was mentioned by Watson and Pelli (1983) and elaborated by King-Smith et al. (1995). Evaluation of this two-dimensional posterior probability distribution is the core of the Ψ method.

2.2. Termination criterion

A purist approach to the termination rule would be toevaluate the expected error from the posterior probability distribution and terminate the experiment when the estimated error goes below a certain level. The confidence interval, as suggested by Treutwein (1995), can be obtained by truncation of the left and right tails of the posterior distribution as their areas reach a value of $(1-\gamma)/2$, where γ is the confidence level. This is a straightforward and computationally inexpensive method.

Watson and Pelli (1983) proposed a different approach for estimation of the confidence interval, which is based on likelihood-ratio test. A further development of this approach was provided by Laming and Marsh (1988), who proposed an approximation formula to compute the variance.

Another possible termination rule is to stop the experiment after a certain number of trials is completed. This rule may be not as efficient as the previous two but it has the advantage of certainty, which is important in the psychophysical milieu (Watson & Pelli, 1983). In our experience, observers have difficulty in distributing their effort evenly along an experimental run when the duration of experiment varies. The error of the adaptive method with termination after a fixed number of trials can be obtained by repeating measurements three to four times (which is always wise to do) or from the results of computer simulations of the method, such as in Fig. 1 below. For the sake of practicality, therefore, the Ψ method that we implement terminates the experiment based on a particular number of trials.

2.3. Placement

Traditionally, adaptive methods place the next trial at the threshold intensity predicted from the completed trials (Watson & Pelli, 1983; Emerson, 1986; King-Smith, Grigsby, Vingrys, Benes & Supowit, 1994). This heuristic rule can be tuned to provide the optimal asymptotic convergence rate by a proper setting of the threshold level (Taylor, 1971). Such tuning makes sense when the number of trials is large and the goal is to get a very precise estimate. For short experiments, as will be shown below, this strategy leaves some room for improvement.

The ideal solution for the placement problem would be to scan all possible scenarios of the sequence of placement choices before all subsequent trials and choose the intensity that provides the minimum expected number of steps before a certain level of accuracy is reached (King-Smith et al., 1994). Unfortunately, this approach is computationally intractable because of its exponential complexity. A surrogate greedy algorithm, typically used in such cases, makes the estimation tractable by looking ahead a small number of steps.

This limited approach for adaptive psychophysical methods was introduced by King-Smith (1984), whose minimum variance method minimized the expected variance of the posterior probability distribution after completion of the next trial. Later, King-Smith et al. (1994) compared the one-step and two-step ahead search and found no significant advantage for the latter strategy. This result is a clear indication that, for the particular task of variance minimization, the greedy search just one step ahead is about as good as an exhaustive search in full depth. The Ψ method therefore adopts the one-step strategy.

Besides near-optimal performance, another advantage of the minimum variance method is that it has an implicit placement rule defined by the variance-based cost function. This feature makes the method highly flexible. The user does not need to bother with the ideal sweat factor (Taylor, 1971) or tabulating the numbers in the procedure for each particular experimental paradigm and parameterization of the psychometric function. For every trial the method finds the optimal (within the one-step constraint) test intensity driven by its own goal.

However, the variance of the posterior probability distribution, which sets this goal for the minimum variance method, cannot be readily expanded to two dimensions because the threshold and slope dimensions are incommensurate. The properties of two-dimensional version of the minimum variance method would depend on arbitrary weights assigned to these dimensions, which, in turn, depend on the sampling rates of the dimensions.

A good approach to overcoming the latter drawback is to define the cost function as the *entropy* of the posterior probability distribution, which specifies how much *information* is needed to get a complete knowledge about the studied system (in our case, the parameters of the psychometric function that controls the observer's responses). This cost function, first suggested by Pelli (1987b) in his ideal psychophysical procedure, is similar to variance because it measures the spread of the posterior probability distribution. Its critical feature is that linear transforms of the variables do not affect the ranking imposed by entropy among distributions (Cover & Thomas, 1991, p. 234) and, consequently, the entropy-based placement rule is insensitive to the sampling rates chosen for the dimensions. The Ψ method therefore employs an entropy-based cost function, which is minimized in the placement of each successive trial intensity.

To summarize, the Ψ method is a combination of solutions known from the literature. The method updates the posterior probability distribution across the sampled space of the psychometric functions based on Bayes' rule (Hall, 1968; Watson & Pelli, 1983). The space of the psychometric functions is two-dimensional (Watson & Pelli, 1983; King-Smith & Rose, 1997). Evaluation of the psychometric function is based on computing the mean of the posterior probability distribution (Emerson, 1986; King-Smith et al., 1994). The termination rule is based on the number of trials, as the most practical option (Watson & Pelli, 1983; King-Smith et al., 1994). The placement of each new trial is based on one-step ahead minimum search (King-Smith, 1984) of the expected entropy cost function (Pelli, 1987b).

There are currently three methods in the same domain of psychometric function estimation as the Ψ method: the method of constant stimuli (Fechner, 1860; McKee, Klein & Teller, 1985), APE (Watt & Andrews, 1981) and the method recently proposed by King-Smith and Rose (1997). All these methods have obvious drawbacks relative to the Ψ method. The method of constant stimuli is non-adaptive and inefficient. APE, although adaptive, uses a non-optimal heuristic placement rule based on a sequence of blocked estimates. The King-Smith and Rose (1995) method places trials based on asymptotically optimal solution for slope estimation. There is no evidence, however, that this method is efficient at the beginning of the experiment, when the threshold is unconstrained.

3. Method

The Ψ method is based on the following logic. We define a sample space of possible psychometric functions $\Psi_\lambda(x)$ where $\lambda = (a, b)$ is a vector with threshold and slope as the coordinates. We assume that the actual psychometric function determining performance falls within this parameter space. At the t-th trial, presentation of the stimulus with intensity x_t produces a binary response r_t which may be *success* or *failure*. The probabilities $p_t(\lambda)$ that reflect the chance for the λ-th psychometric function to match the actual psychometric function are updated after each trial.

The psychometric functions $\Psi_\lambda(x)$ with the assigned probabilities $p_t(\lambda)$ comprise a probability space that we can evaluate by means of its entropy

$$H_t = -\sum_\lambda p_t(\lambda)\log(p_t(\lambda)).$$

Smaller entropies correspond to higher confidence as to which particular $\Psi_\lambda(x)$ is the actual psychometric function controlling the observer's performance. In the extreme case where one $\Psi_\lambda(x)$ has a probability of one and all others are zero, the entropy has the smallest value of zero. If we compare the entropy on two consecutive trials, t and $t+1$, the difference $H_t - H_{t+1}$, as defined in information theory, is the amount of information about the psychometric function gained in trial $t+1$. Before starting trial $t+1$, one can estimate the expected entropy $E[H_t(x)]$ for the two possible outcomes of the experiment (success or failure) for each test intensity x. The strategy for placing the new trial is to choose the one that leads to the smallest expected entropy one step ahead. Testing this intensity maximizes the expected gain of information about the psychometric function after completion of the next trial.

Before each experiment, two functions should be initialized. First, a prior probability distribution $p_0(\lambda)$ for the psychometric functions must be set up. Second, to speed up the method, a look-up table of conditional probabilities $p(r|\lambda, x)$ should be computed for each combination of the parameters r, x and the λ according to the following formulae:

for all intensity

$$p(success|\lambda, x) = \Psi_\lambda(x) \text{ and } p(failure|\lambda, x) = 1 - \Psi_\lambda(x).$$

Each trial consists of the following steps:

1. Calculate probability of getting response r after presenting test x at the next trial.

$$p_t(r|x) = \sum_\lambda p(r|\lambda, x)p_t(\lambda)$$

LUT

2. Estimate by Bayes' rule the posterior probability of each psychometric function given that the next trial will produce the response r to the test of the intensity x.

$$p_t(\lambda|x, r) = \frac{p_t(\lambda)p(r|\lambda, x)}{\sum_\lambda P_t(\lambda)p(r|\lambda, x)}$$

3. Estimate the entropy of the probability density function over the space of psychometric functions, given that at the next trial a test of intensity x will produce the response r. PDF

$$H_t(x, r) = -\sum_\lambda p_t(\lambda|r, x)\log(p_t(\lambda|x, r))$$

4. Estimate the expected entropy for each test intensity x.

$$E[H_t(x)] = H_t(x, success)p_t(success|x) + H_t(x, failure)p_t(failure|x)$$

5. Find the test intensity that has the minimum expected entropy

$$x_{t+1} = \arg\min_x E[H_t(x)],$$

where arg min returns the argument value at which the function has its minimum.
6. Run a trial with intensity x_{t+1} to obtain the response r_{t+1}.
7. Keep the posterior probability distribution from step 2 that corresponds to the completed trial.

$$p_{t+1}(\lambda) = p_t(\lambda|x_{t+1}, r_{t+1})$$

8. Find a new estimate of the psychometric function based on the new posterior probability distribution $p_{t+1}(\lambda)$. The expected value of λ provides the answer:

$$\lambda_{t+1} = \sum_\lambda \lambda p_{t+1}(\lambda),$$

which is expressed solely in terms of the vector λ, whose coordinates are threshold a and slope b of the psychometric function.
9. Return to step 1 unless a specified number of trials is completed.

4. Implementation

The Ψ method was implemented in C on a Power Macintosh with a 100 MHz clock. On this computer for the specified ranges of intensity, threshold and slope, preparation for each trial took about 400 ms.

The initial ranges for the three variables of intensity x, threshold a and slope b were 3, 3 and 1 decimal log units, respectively, which are the full ranges available for typical contrast detection and discrimination tasks. The psychometric functions were sampled in 1 dB steps (1/20-th of the log unit) in each dimension. The prior distribution was chosen to be uniform across the full sampling range of the threshold and slope. (This choice of the prior is, of course, arbitrary. The prior for a psychophysical task estimated from previous work should be used, if available, to maximize the speed of convergence).

The psychometric functions were implemented based on a power-law approximation to the d' function, which originates from signal detection theory. One natural interpretation of this approximation is that the noise in the visual system is constant for small signals and that the internal signal is a power function of the intensity (Nachmias & Sansbury, 1974; Foley & Legge, 1981; Pelli, 1987a). The consequent signal-to-noise ratio $d'(x)$ for stimulus with intensity x is

$$d'(x) = (x/a)^b$$

where b is the slope parameter; if the threshold is defined as $d' = 1$, then a is the threshold. In log–log coordinates, d' as a function of intensity is represented by a straight line with slope b. The threshold parameter a defines the horizontal shift of the line.

The Ψ method was implemented and tested for the 2AFC task (the yes/no paradigm is omitted in this study as insufficiently controlled). The signal-to-noise ratio d' in this task defines the psychometric function

$$\Psi(d') = N(d'/\sqrt{2}, 1)$$

where

$$N(\mu, \sigma) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\mu} e^{-(t-\mu)^2/2\sigma^2} dt$$

is the cumulative function for the Gaussian distribution.

In practice, observers miss some small portion of trials and have to guess, no matter how intense the stimulus is. Incorporating the probability of misses δ in the definition of the parameterization of the psychometric function, we arrive at the following expression:

$$\Psi_{a,b}(x) = \delta/2 + (1-\delta)N((x/a)^b/\sqrt{2}, 1)$$

As a function of x, this psychometric function rises from the baseline level of 0.5 and saturates at $1 - \delta/2$.

In our evaluations, the miss rate was set at 0.04 and the psychometric function saturated at 0.98 proportion correct. The threshold was set at $\Psi = 0.75$, which corresponds to $d' = 1$. This definition of threshold is used in all computational and psychophysical experiments unless otherwise specified.

5. Computer simulations

5.1. Evaluation of convergence rate and estimation bias

First, we evaluated the convergence of the Ψ method for threshold and slope estimates. The evaluation was conducted for assumed observer slope values of 1, 2 and 4; the slope range in the Ψ method was set from 0.7 to 7, i.e. 1 decimal log unit. For each observer slope value we ran at least 1000 Monte-Carlo simulations of a 2AFC psychophysical experiment.

The simulated observer was set to have a 0.04 miss rate, matching the miss rate assumed for the Ψ method. The threshold was defined at $\Psi = 0.75$ to be consistent with the method. To avoid edge effects, the threshold range for the simulated observer was set from 0.5 to 2.5

log units, somewhat narrower than the range of intensities that the method potentially can evaluate.

For each trial number, the bias and standard deviation statistics were accumulated for slope and threshold across the simulated experiments. Since the statistical analyses for threshold and slope were the same, they will be described for the threshold case. Let $i = 1,\ldots, I$ be the experiment number, $j = 1,\ldots, J$ the trial number, a_i the threshold value randomly set at the i-th experiment, and a_{ij} the threshold estimate obtained at j-th trial in i-th experiment. All computations were carried out on a log scale to match the scale adopted in the Ψ method. The discrepancy between two positive values x and y was estimated in dB units: $20\ \log_{10}(x/y)$. No discrepancy corresponds to 0 dB and near this value the behavior of the residual expressed in dB units is essentially linear.

The bias of the threshold estimate after j-th trial was calculated with the following formula:

$$\text{bias}(a)_j = \frac{\sum_{i=1}^{i=I} (\log a_{ij} - \log a_i)}{I} \cdot 20\ \text{dB}.$$

The standard deviation was calculated as

$$\text{S.D.}(a)_j = \sqrt{\frac{\sum_{i=1}^{i=I} (\log a_{ij} - \log a_i)^2}{I-1}} \cdot 20\ \text{dB}.$$

Fig. 1a shows the standard deviation of the threshold estimate as a function of the number of trials. The convergence rate depends on the slope value: for steeper slopes the estimates are more precise. Asymptotically, the error of the threshold estimate is reciprocal to the square root of the number of trials, which corresponds to a slope of -0.5 on the log–log plot (the dB units of the error on the logarithmic scale should not be mistaken for a doubly-logarithmic scale, since the residual near zero is essentially linear). For a large number of trials, the standard deviation is proportional to the slope value.

Fig. 1b shows the convergence rate for the slopes. Initially, the minimum entropy criterion operates to orient the Ψ method toward estimating the threshold parameter; the slope estimate stays at its starting value close to the middle of the slope range in the log scale; i.e. at 2.21. For this reason, the slope estimate for the actual slope of 2 happened to be relatively precise from the beginning; it then became less precise as the Ψ method started to evaluate the slope before finally reconverging toward its actual value. Asymptotically, the standard deviation of the slope estimate is reciprocal to the square root of the number of trials. It is important to note that the accuracy of the slope estimate is essentially independent of the actual slope value within the range evaluated.

Fig. 1c, d show that beyond 40 trials the bias values for both threshold and slope estimates rapidly converge to zero. This feature indicates that asymptotically the Ψ method produces unbiased estimates in the 2AFC implementation.

5.2. How the Ψ method works

The placement strategy of the Ψ method is to gain maximum information at each trial, which results in different placement rules as the staircase progresses. As depicted in Fig. 1, at the beginning of the experiment the method attempts to localize the threshold; its placement strategy is reminiscent of the bisection method (Press, Teukolsky, Vetterling & Flannery, 1992, p. 353).

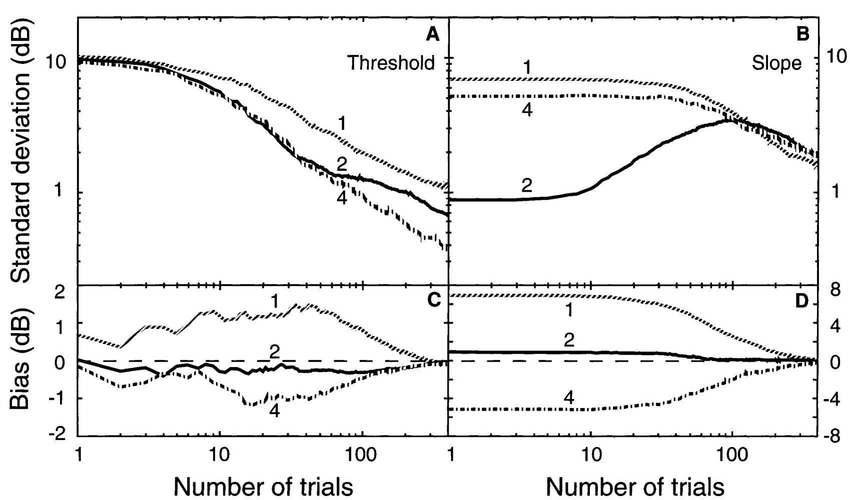

Fig. 1. Convergence properties of the Ψ method in the 2AFC task for observer slopes of 1, 2 and 4. The horizontal axis in each panel represents the trial number on a logarithmic scale. The vertical axes in the top panels (a, b) represent the standard deviation for the threshold and slope estimates at each trial; the standard deviation is computed as a ratio in the log space and expressed in dB units, which is essentially linear for the small error values. The vertical axes in the bottom panels (c, d) show the bias of the threshold and slope estimates on a linear scale.

After the psychometric function is positioned accurately on the intensity axis, the slope becomes the object of major concern. At this stage the entropy profile takes on a shape with two local minima and the global minimum alternates between them from trial to trial, as shown in Fig. 2. For a large number of trials, these minima are located at the intensities that correspond to the 0.69 and 0.92 probability levels. Focusing on the slope measurement does not mean that the method abandons the threshold evaluation: it continues to improve the threshold estimate as a by-product of the slope acquisition. Apparently, this is the best strategy when the threshold and the slope are already known with some accuracy.

It should be noted that we have no proof that the alternation between the two intensities, to which the method converges while estimating the slope, is random. There is a possibility, therefore, that observers, after seeing the intensity of the stimulus in a current trial, may make a correct guess regarding the intensity in the next trial. Nevertheless, we argue that the knowledge of the particular intensity to be presented at any trial does not affect the outcome of the 2AFC task since it does not help discrimination between the test and blank intervals presented in random order.

5.3. Effect of the miss rate

In our implementation of the Ψ method, the assumed

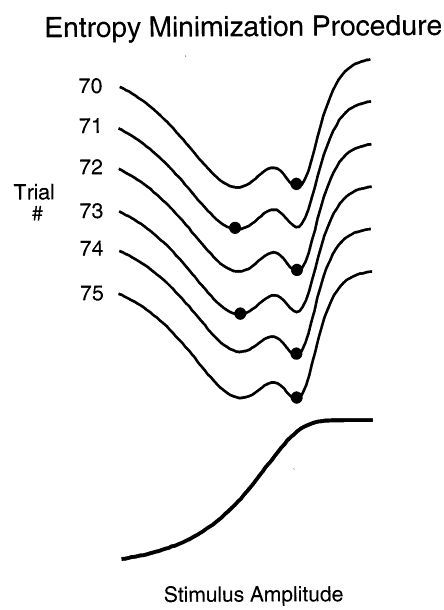

Fig. 2. Entropy profiles (thin curves) in six consecutive trials after the method converged to threshold. The dots on the curves depict the minimum entropy points where the test intensity will be placed in the next trial. The thick line at the bottom shows the actual psychometric function assumed in this computational experiment. The Ψ method places test points approximately at the ends of the linear region in the rising part of the psychometric function. This strategy is evidently optimal for slope estimation.

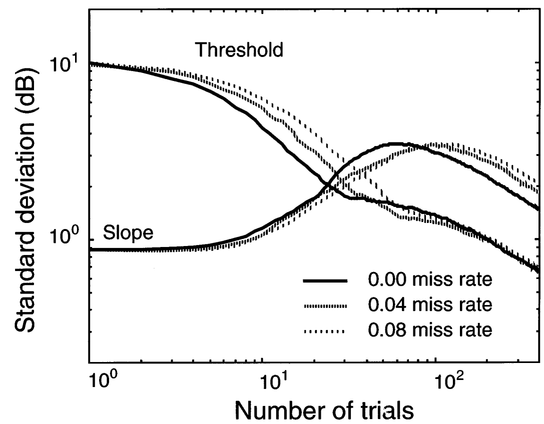

Fig. 3. Performance of the Ψ method with assumed miss rates of 0, 0.04 and 0.08. The miss rate for the simulated observer matched the assumed miss rate of the method. The assumed value of the miss rate greatly affects the method efficiency.

miss rate is arbitrarily set at a conservative level of 0.04; for trained observers it may be smaller. To evaluate the effect of the assumed miss rate on the convergence of the method, we carried out simulations at three different miss rates: 0, 0.04 and 0.08. The slope of the psychometric function of the simulated observer was set at 2, while its miss rate matched that assumed by the method. The results presented in Fig. 3 indicate that at the beginning of experiment the convergence of the threshold estimate greatly depends on the assumed miss rate; at larger numbers of trials this dependence gradually vanishes. The slope estimate does retain a residual dependence on the miss rate for the number of trials studied.

This result suggests that a special effort should be made to reduce the occurrence of misses. First, if possible, observers with high miss rates should be avoided. Second, the first few trials may be discarded since they may have a higher miss rate as the observer settles into the task.

5.4. Threshold estimate: comparison with ZEST

In many experiments the threshold estimate is the only value needed, and an experimenter may not wish to spend an effort to measure slopes. The traditional Bayesian adaptive methods (QUEST, ZEST) address this concern: they estimate threshold based on a realistic assumption about the slope value. This assumption is easy to incorporate in the Ψ method by setting the range of slopes at a single value. We call this version the slope-constrained Ψ method as distinct from the unconstrained version where the method estimates a range of slopes (1 decimal log unit in our simulations). Indeed, the prior information carried by the assumption, given that it is valid, should be beneficial for the method performance. Moreover, the next computational experiment shows that ZEST, a modification of

QUEST that apparently is the best among popular adaptive methods, and the slope-constrained Ψ method have no advantage in practice against the more general unconstrained Ψ method, even when the assumed and actual slopes match.

The slope of the psychometric function in the simulated observer was set at a value of 2; the assumed slopes in ZEST and the slope-constrained Ψ method matched this value. The other parameters were the same as in the first simulation, which resulted in Fig. 1. In the implementation of ZEST, the test intensity was placed at 90% point of the estimated psychometric function, which provided the maximum efficiency to this method. (According to our estimates, this point corresponds to the minimum of the *ideal sweat factor*. Taylor, 1971.) The slope-constrained Ψ method did not require any preliminary settings since the method itself found the optimal intensity to be tested. The results of the simulations are presented in Fig. 4.

The simulated convergence curves for ZEST and the slope-constrained Ψ method completely overlapped. This result shows that, if the experimenter knows the slope of the psychometric function, both methods are equally good. The unconstrained Ψ method has practically the same performance as the other two methods up to 30 trials, after which its performance degrades somewhat. As mentioned above, after 30–40 trials the Ψ method starts measuring the slope and the threshold estimate becomes slope-tolerant. ZEST and the slope-constrained Ψ method continue to take advantage of the prior knowledge of the slope to set the stimulus optimally at each trial. However, this advantage is based on the shaky ground of the slope assumption. If this assumed slope deviates from the real one, the advantage would disappear since the slope-constraining methods would start making systematic errors and increasing the number of trials would not improve the estimate. There is no such problem with the unconstrained Ψ method.

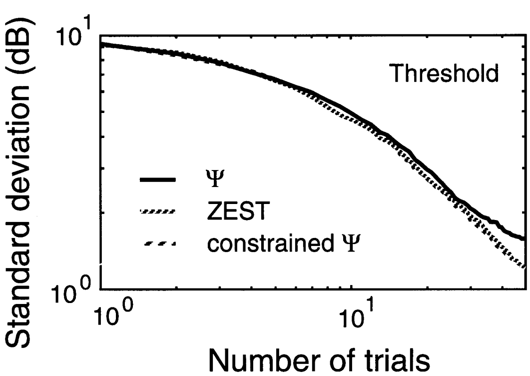

Fig. 4. Convergence curves for the full Ψ method (the slope range is 1 decimal log unit), ZEST, and the slope-constrained Ψ method. The assumed slope in the last two methods matched that in the simulated observer. The placement in the ZEST method was at the 90% point on the estimated psychometric function, which provided the maximum efficiency.

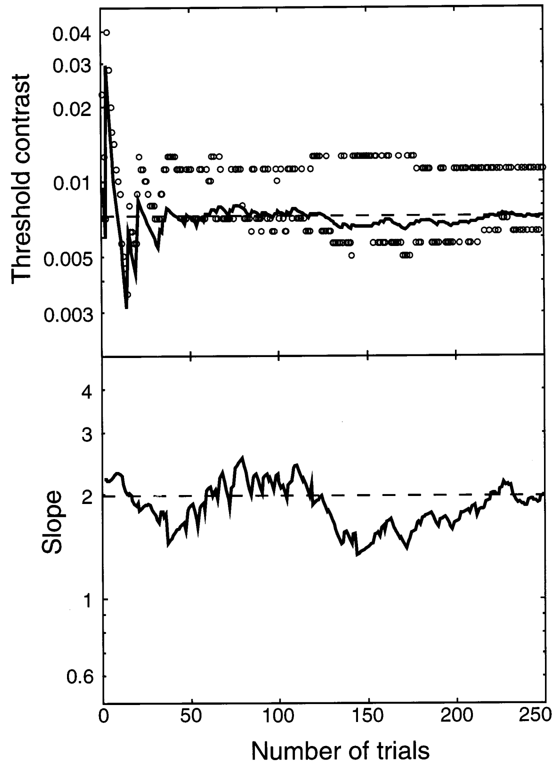

Fig. 5. The results of a real psychophysical experiment. The top panel shows the convergence of the threshold estimate. The solid line depicts the threshold estimate at each trial, the circles show the intensities tested at each trial, and the horizontal dashed line shows the threshold level as measured from all 250 trials. The bottom panel shows the convergence of the slope estimate.

Each of the methods compared provides the same 2 dB accuracy after 30 trials, which is a typical target level in psychophysical experiments. The slope-specific adaptive methods have no advantage over the unconstrained Ψ method within this run length.

6. Psychophysical experiment

Two psychophysical experiments were carried out to evaluate the behavior of the Ψ method in a real task. The experienced observer (who was nevertheless naive regarding the goal of the experiment) performed a 2AFC contrast detection task for a Gabor stimulus presented binocularly in a raised cosine temporal envelope with a 2 s period. The carrier of the Gabor profile had a spatial frequency of 2 cpd; the circular Gaussian envelope had $\sigma = 1°$. The stimuli on the monitor were controlled by a video attenuator (Institute for Sensory Research, Syracuse University) and the Video Toolbox software (Pelli & Zhang, 1991). Each experiment consisted of 250 trials; the subject took a break of a few minutes after doing 50–100 trials, performing two experiments per day over 3 days.

Table 1

Experiment	1	2	3	4	5	6	Mean value	Measured S.D. (dB)	Predicted S.D. (dB)
Threshold contrast after 30 trials (%)	0.48	0.56	1.14	0.76	0.72	0.84	0.72	2.7	2.1
Same with mistake on first trial (%)	0.75	0 59	1.08	0.90	0.65	0.42	0.70	2.9	n/a
Threshold after 250 trials (%)	0.67	0.72	0.87	0.75	0.70	0.74	0.74	0.8	0.9
Slope after 250 trials	1.52	1.98	1.88	1.59	1.76	1.99	1.78	1.0	2.5

The results of one typical experiment are presented in Fig. 5. The threshold estimate rapidly converges during the first 30–40 trials. After the threshold is localized, the method starts testing two intensity levels: one below and the other above the threshold level (see Fig. 2 and the related discussion). The observer commented that he was able to see the test in most of the trials, which made him comfortable with the task.

The intermediate (after 30 trials) and final (after 250 trials) results for each experiment are shown in Table 1. In a logarithmic scale, the means and standard deviations for the threshold (after 30 and 250 trials) and slope (after 250 trials) estimates are presented, respectively in the third and the second column from the right. The mean slope estimate after 250 trials was 1.78, which is consistent with other studies (Stromeyer & Klein, 1974; Foley & Legge, 1981). The rightmost column presents the standard deviations predicted in the computer simulations for the slope value of two. The similarity of the measured and predicted values suggests that the simulation results adequately represent the behavior of a real observer.

A major complaint of Bayesian adaptive methods is that they become unstable if the observer makes a mistake at the beginning of the experiment (S.P. McKee, personal communication). To evaluate the robustness of the Ψ method to mistakes made at the beginning of the experiment, the observer was instructed to make a mistake on the first trial where the stimulus was clearly visible (2.8% contrast, three to four times above the threshold). The results of six 30-trial experiments (see Table 1, third row) are practically identical to those with no forced mistake. A mistake at the beginning of the experiment, therefore, does not have a significant effect on the convergence of the method.

This consistency of the experimental results, together with our experience with the Ψ method applied to other tasks, allows us to conclude that the Ψ method improves the instability problems that haunt earlier adaptive methods.

7. Hardware requirements

It would not be an exaggeration to say that modern computer technology has outgrown the extant psychological methods, which use only a small fraction of available computational power and memory. It was not clear, however, how these resources might be used to improve the duration and accuracy of psychophysical measurements. The Ψ method resolves this discrepancy. Currently, to implement the Ψ method, requires the most powerful personal computers: among the Macintosh computers only those with PowerPC processor can run the method with tolerable delays. The memory requirements of the Ψ method are also quite demanding: for the ranges and the sampling rates used in our simulations the program needs about 1 Mbyte of memory. Nonetheless, these requirements are rapidly becoming routine as computing power continues to increase.

8. Conclusion

A new Bayesian adaptive method for both threshold and slope evaluation in psychophysical experiments is proposed. The method is based on maximizing the expected information gain (minimizing entropy) on each subsequent trial. The entropy cost function is proven to be insensitive to a particular choice of sampling densities for the threshold and slope parameters, given that they are sufficiently high. The method has been tested in real experiments and shown to exhibit the predicted convergence rate and to be robust to mistakes made at the beginning of the experiment.

Acknowledgements

The authors thank C.C. Chen, P. Verghese and an anonymous reviewer for numerous suggestions that substantially improved the manuscript. Supported by NIH grant 7890.

References

Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York: Wiley.

Emerson, P. L. (1986). Observations on maximum likelihood and Bayesian methods of forced choice sequential threshold estimation. *Perception and Psychophysics*, *39*, 151–153.

Fechner, G. T. (1860). In: D. H. Howes, E. C. Boring, *Elemente der Psychophysik*. New York: Holt, Rinehart & Winston (H. E. Adler, Trans.: original work published 1966).

Foley, J. M., & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research*, *21*, 1041–1053.

Gelb, A. (1982). *Applied optimal estimation*. Cambridge MA: MIT.

Hall, J. L. (1968). Maximum-likelihood sequential procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America*, *44*, 370.

King-Smith, P. E., & Rose, D. (1997). Principles of an adaptive method for measuring the slope of the psychometric function. *Vision Research*, *37*, 1595–1604.

King-Smith, P. E., Earley, M. J., Pierce, G. E., & Preston, J. L. (1995). An adaptive method for measuring both the threshold and slope of the psychometric function. *Investigative Ophthalmology and Visual Science* (*Supplement*), *36*, 906.

King-Smith, P. E., Grigsby, S. S., Vingrys, A. J., Benes, S. C., & Supowit, A. (1994). Efficient and unbiased modifications of the QUEST threshold method: theory, simulations, experimental evaluation and practical implementation. *Vision Research*, *34*, 885–912.

King-Smith, P. E. (1984). Efficient threshold estimates from yes/no procedures using few (about 10) trials. *American Journal of Optometry and Physiological Optics*, *61*, 119.

Laming, D., & Marsh, D. (1988). Some performance tests of QUEST on measurements of vibrotactile thresholds. *Perception and Psychophysics*, *44*, 99–107.

Legge, G. E., Kersten, D., & Burgess, A. E. (1987). Contrast discrimination in noise. *Journal of the Optical Society of America*, *A4*, 391–404.

Mayer, M. J., & Tyler, C. W. (1986). Invariance of the slope of the psychometric function with spatial summation. *Journal of the Optical Society of America*, *A3*, 1166–1172.

McKee, S. P., Klein, S. A., & Teller, D. Y. (1985). Statistical properties of forced-choice psychometric functions: implications of probit analysis. *Perception and Psychophysics*, *37*, 286–298.

Nachmias, J., & Sansbury, R. V. (1974). Grating contrast discrimination may be better than detection. *Vision Research*, *14*, 1039–1042.

Pelli, D. G. (1985). Uncertainty explains many aspects of visual contrast detection and discrimination. *Journal of the Optical Society of America*, *A2*, 1508–1532.

Pelli, D. G. (1987). On the relation between summation and facilitation. *Vision Research*, *27*, 119–123.

Pelli, D. G. (1987). The ideal psychometric procedure. *Investigative Ophthalmology and Visual Science* (*Supplement*), *28*, 366.

Pelli, D. G., & Zhang, L. (1991). Accurate control of contrast on microcomputer displays. *Vision Research*, *31*, 1337–1350.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in C: the art of scientific computing* (2nd). Cambridge: Cambridge University.

Stromeyer, C. F., & Klein, S. A. (1974). Spatial frequency channels in human vision as asymmetric (edge) mechanisms. *Vision Research*, *14*, 1409–1420.

Taylor, M. M. (1971). On the efficiency of psychophysical measurement. *Journal of the Acoustical Society of America*, *49*, 505–508.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, *35*, 2503–2522.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychophysical method. *Perception and Psychophysics*, *33*, 113–120.

Watt, R. J., & Andrews, D. P. (1981). APE: adaptive probit estimation of psychometric functions. *Current Psychological Reviews*, *1*, 205–214.