

## COMPUTER TECHNOLOGY

# Microcomputer-based estimation of psychophysical thresholds: The Best PEST

HARRIS R. LIEBERMAN and ALEX P. PENTLAND  
*Massachusetts Institute of Technology, Cambridge, Massachusetts 02139*

A new, maximally efficient technique for measuring psychophysical thresholds (Pentland, 1980) has been implemented on the microcomputer. This PEST (parameter estimation by sequential testing) technique is the most efficient sequential parameter estimation technique possible, given that the form of the psychometric function is known. The technique is similar to but faster and more accurate than other staircase procedures and may be applied whenever staircase techniques are applicable. The "Best PEST" is easily implemented on the microcomputer; a BASIC program for the Apple II which does so is presented. The Best PEST is compared with other staircase procedures, including one recently implemented on a microcomputer (Corwin, Kintz, & Beaty, 1979).

The appearance of the microcomputer in the behavioral laboratory has provided opportunities for increasing the ease and efficiency with which psychophysical measurements may be made. We have implemented, on a microcomputer, the most efficient PEST (parameter estimation by sequential testing) procedure possible, given that the form of the psychometric function is known.

Staircase-like procedures, such as the one described here, are applicable to a large class of psychophysical experiments in which the experimenter wishes to determine a threshold for some psychometric function. This threshold is usually defined as the level of an independent physical variable that determines a particular response probability. The objective of the experimenter is to determine, as rapidly and accurately as possible, the value of the physical variable that yields this threshold response.

Earlier staircase procedures (Cornsweet, 1962; Dixon & Mood, 1948; Wetherill, 1963; Wetherill & Levitt, 1965), although representing an improvement over the method of constant stimuli, could not efficiently utilize information gained on each trial. Taylor and Creelman (1967) and, later, Findlay (1978) devised ad hoc PEST techniques that significantly reduce the number of measurements necessary to reach a given level of accuracy. The Best PEST (Pentland, 1980) takes this evolution to its conclusion by using a maximum-likelihood estimation technique that results in the most efficient parameter-estimation procedure possible, given that the form of the psychometric function is known.

Funding for this research was provided by NSF Grant MCS79-23110 (to H. R. Lieberman and A. P. Pentland) and by NIH Training Grant 5 T32 GM07484 (to A. P. Pentland). Additional funding was provided by NIH Grant 5 P30 EY02621. We thank Terry Allard for his careful reading of the manuscript and Carol Papineau for technical assistance. Requests for reprints should be sent to Harris R. Lieberman, Department of Psychology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.

### THE BEST PEST

The Best PEST, like the Findlay (1978) and Taylor and Creelman (1967) PEST methods, depends on information gathered in previous trials to guide further measurement. These PEST techniques, unlike earlier staircase methods, adjust staircase step size on the basis of information already gathered, thereby increasing the amount of information about the threshold gained on subsequent trials.

The Best PEST functions by producing a maximum-likelihood estimate of the setting of the independent variable which will produce the maximum amount of information about the position of the threshold at each trial on the basis of all previous responses. By measuring at this variable setting, the maximum amount of information about the position of the threshold can be gained at each step in the procedure. This results in the most efficient parameter estimation possible. A more detailed description of the Best PEST, including a mathematical formulation, is presented in the appendix.

### IMPLEMENTATION

The Best PEST is easily implemented on a microcomputer, in this case, an Apple II microcomputer, in APPLESOFT BASIC. Figure 1 is a BASIC program that implements the Best PEST algorithm. The implementation consists of one subroutine or function, "THRESHOLD," which performs the necessary computations when called by the main program, some initialization statements, and a main program to call subroutine THRESHOLD, present stimuli, and record the observer's response.

#### Initialization: Figure 1, Lines 10-130

First, it is necessary to declare and initialize the arrays. Array "PROB" holds the cumulative probability that the threshold is at each of the possible measurement

## INITIALIZATION

```

10 INPUT "STIMULUS RANGE";RANGE: REM OVER HOW MANY STIMULUS
INTERVALS ARE YOU TESTING?
20 INPUT "NO. OF TRIALS";NTRIALS
30 DIM PROB(RANGE*2),PLGIT(RANGE*2),MLGIT(RANGE*2)
40 STD = RANGE/5: REM STD SETS THE SLOPE OF THE PSYCHOMETRIC
FUNCTION
50 FOR I = 1 TO 2*RANGE
60 PROB(I) = 0
70 LGIT = 1/(1 + EXP ((RANGE - I)/STD))
80 PLGIT(I) = LOG(LGIT)
90 MLGIT(I) = LOG(1 - LGIT)
100 NEXT I
110 M = RANGE: R = 1
120 GOSUB 2000
130 M = 1: R = -1

```

## MAIN PROGRAM

```

1000 REM MAIN PROGRAM PRESENTS STIMULUS AT VARIABLE SETTING OF "M";
RECORDS OBSERVER RESPONSE AS VARIABLE "R"
1010 FOR TRIAL = 1 TO NTRIALS
1020 GOSUB 2000
1030 Print "STIMULUS PRESENTED = ",M: REM YOUR PROGRAM PRESENTS THE
STIMULUS HERE, AT PARAMETER SETTING 'M'.PrintM.
1040 INPUT "RESP=";R:REM INPUT SUBJECT'S RESPONSE(1 if correct; -1 if incorrect)
1050 NEXT TRIAL
1060 END:REM TESTING COMPLETED

```

## SUBROUTINE THRESHOLD

```

2000 REM SUBROUTINE THRESHOLD
2010 MAX = -10000
2020 FOR I = 1 TO RANGE
2030 IF R = 1 THEN PROB(I) = PROB(I) + PLGIT(RANGE + M - I)
2040 IF R = -1 THEN PROB(I) = PROB(I) + MLGIT(RANGE + M - I)
2050 IF PROB(I) > MAX THEN MAX = PROB(I):P1 = I
2060 IF PROB(I) = MAX THEN P2 = I
2070 NEXT I
2080 M = INT((P1 + P2)/2)
2090 RETURN

```

Figure 1. The Best PEST implemented in APPLESOFT BASIC.

values of the independent variable, based on the observers' responses. Variable "RANGE" is the range of possible stimulus values (i.e., the number of possible independent variable testing values). The maximum, at stimulus value "RANGE," should be the easiest stimulus to detect.

Arrays "PLGIT" and "MLGIT" hold the psycho-

metric function. PLGIT is the probability of a positive response and MLGIT is the probability of a negative response. The standard ogive for detection tasks is assumed in this sample program. For two-alternative forced-choice tasks, in which lowest probability of a positive response is .5, Lines 70, 110, and 130 in Figure 1 would be changed to:

$$70 \text{ LGIT} = 0.5 + 0.5/(1 + \text{EXP}((\text{RANGE} - 1)/\text{STD}))$$

$$110 \text{ M} = 1: \text{R} = -1$$

$$130 \text{ M} = 1: \text{R} = 1$$

In arrays PLGIT and MLGIT, the logarithms of the psychometric function are stored, rather than the actual values, to protect against underflows or overflows in the machine. If it is undesirable to use this much memory, one array containing the positive psychometric function can be used, and the values of PLGIT and MLGIT can be computed from that array. This requires calling the log function RANGE times per stimulus presentation.

Variable "STD" is an estimate of the steepness of the psychometric function; a typical value is one-fifth on the range. The performance of the estimation procedure is fairly independent of the value of STD; however, if the true value is unknown (or if the observer is likely to make errors), it is better to use a larger value of STD than to use a smaller one. To complete the initialization procedure, we set the maximum of the independent variable's (M) range as detected by the observer (Lines 110-120), and the minimum as not detected by the observer (Line 130).

#### The Main Program: Figure 1, Lines 1000-1060

Lines 1000-1060 show a program that uses the threshold subroutine to estimate a psychophysical parameter. A stimulus is presented repeatedly (NTRIALS sets the number of presentations), and a response is recorded. Subroutine THRESHOLD is called before each stimulus presentation. In a detection task, the first stimulus presented will be at the midpoint of the stimulus range; in a two-alternative forced-choice task, the first stimulus will be at the maximum stimulus value.

Rather than a set number of trials, the PEST can also be terminated when the threshold is determined to within a previously specified confidence interval, by using the PROB array. The values of the PROB array are the unnormalized probabilities that the threshold is at the corresponding value of the independent variable. Thus, to obtain a 95% confidence interval that is plus or minus two settings of the independent variable in width, one simply continues the PEST until the five elements of the PROB array surrounding the estimated threshold (the element at the position of the estimated threshold plus two elements on either side) contain 95% of the sum of the values in the whole array.

#### Subroutine Threshold: Figure 1, Lines 2000-2090

This is the subroutine that carries out the computations needed for the Best PEST, as given in the last equation in the appendix. This subroutine determines the value of the independent variable to be tested next.

Variable "M" is the value of the independent variable at which the last test was made (corresponding to  $m_{n-1}$

in the appendix), and Variable "R" is the response given to the presentation of the stimulus at the value "M"; "R" is +1 if the answer was positive (correct detection) and -1 otherwise. When subroutine Threshold is called, "M" is set to the next test value of the independent variable, corresponding to  $m_n$  in the appendix.

Determination of the next value of the independent variable is done by updating the PROB array with the information gained in the last stimulus presentation. A pointer, "P1," is kept that points at the maximum value in the updated PROB array. Normally, the next value for the independent variable is the value of P1 (i.e., M is set to P1). A second pointer, "P2," is kept in case the probability function contained in PROB is flat in the center; then, the next measurement value (M) is the average of the endpoints of the flat region, which are "P1" and "P2."

#### COMPARISON WITH MICROCOMPUTER-BASED WETHERILL TRACKING

A comparison of the efficiency of the Best PEST procedure relative to Wetherill tracking as implemented on the microcomputer by Corwin, Kintz, and Beaty (1979) was made through a series of simulations. The Corwin et al. Wetherill tracking paradigm is as follows. First, an initial step size and starting point are selected and eight peak-valley reversals are recorded. At this point, a preliminary estimate of threshold is determined by averaging the eight peak-valley pairs. Testing then resumes, starting at that point, using a step size reduced by one-half and continuing until eight more peak-valley pairs are accumulated. The threshold stimulus level is then calculated as the mean of all peaks and valleys.

A standardized simulation procedure<sup>1</sup> is employed. Simulations, as opposed to experiments, are performed so the Best PEST can be compared with Wetherill tracking while critical parameters are varied systematically. The simulation paradigm assumes the psychometric function measured is a standard symmetrical ogive. It can be described by the logit function,  $P(+)=1.0/(1.0+e^{-L})$ , where P(+) is the probability of a correct response as a function of L, the stimulus level. When the stimulus level (L) (for example, brightness, loudness, etc.) varies by one step, this is termed a logit unit, an arbitrary metric for describing changes in stimulus level. For these simulations, threshold is chosen as the 50% response point. The threshold value is placed 5-10 logit units from the center of the independent variable's range. Since the Wetherill tracking technique also requires the experimenter to select step size, Wetherill tracking simulations are run at step sizes ranging from .2 to 6.4 logit units.

The results of the simulation are presented in Figure 2. The accuracy of each technique in approaching the actual threshold is given as a standard deviation in logit units as a function of number of trials. Each point

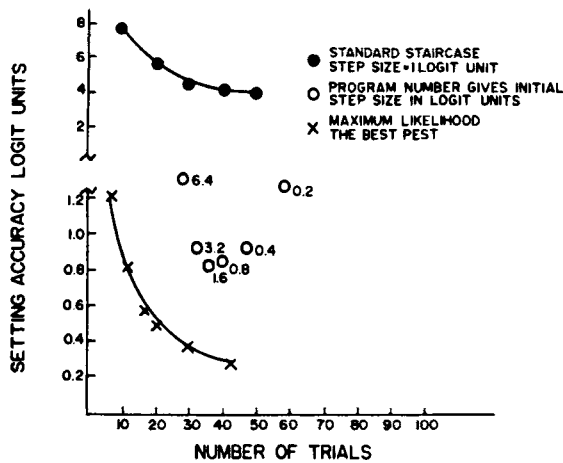


Figure 2. The accuracy of each technique in approaching the actual threshold is given as a standard deviation in logit units as a function of number of trials. Filled circles are for the standard staircase, open circles for Wetherill tracking, and the crosses for the Best PEST. Each point is based on at least 50 simulation trials. The accuracy measure used here is the standard deviation of the estimator (in logit units) after  $n$  steps of the algorithm have been completed (i.e.,  $n$  measurements have been taken).

is based on at least 50 simulation trials. For the Wetherill tracking program, the initial step size (in logit units) is shown next to the data point. This figure also shows a conventional up-down staircase procedure, for purposes of comparison with the Corwin et al. (1979) Wetherill tracking and the Best PEST. A comparison of the Best PEST with other PEST procedures is presented by Pentland (1980).

Both the Wetherill tracking and the Best PEST perform much better than the conventional staircase procedure. It is not possible to do much better than the performance shown in Figure 2 using a conventional staircase, as changing the step size trades off speed against accuracy. Even with the most advantageous choice of initial step size (1.6 logit units), the estimation of threshold provided by Wetherill tracking is about 20% less accurate than the Best PEST estimate.<sup>2</sup> The number of trials required to reach this accuracy even in the best case is one-third more than that required for the Best PEST.

As the initial choice of step size in Wetherill tracking varies from the optimal estimate, accuracy decreases and a greater number of trials may be required. Thus, the technique of Wetherill tracking is highly dependent on an initial choice of step size and starting point unlike the Best PEST. These results indicate that in a variety of psychophysical situations, especially when no accurate estimate of threshold is available preliminary to testing, such as is typical of the nonhomogeneous populations found in clinical situations, the Best PEST is preferable to Wetherill tracking.

#### REFERENCES

- CORNSWEET, T. N. The staircase-method in psychophysics. *American Journal of Psychology*, 1962, **75**, 485-491.

- CORWIN, T. R., KINTZ, R. T., & BEATY, W. J. Computer-aided estimation of psychophysical thresholds by Wetherill tracking. *Behavior Research Methods & Instrumentation*, 1979, **11**, 526-528.
- DIXON, W. J., & MOOD, A. M. A method for obtaining and analyzing sensitivity data. *Journal of the American Statistical Association*, 1948, **43**, 109-126.
- FINDLAY, J. M. Estimates on probability functions: A more virulent PEST. *Perception & Psychophysics*, 1978, **23**, 181-185.
- PENTLAND, A. P. The Best PEST, a maximum-likelihood parameter estimation procedure. *Perception & Psychophysics*, 1980, **28**, 377-379.
- TAYLOR, M. M., & CREELMAN, C. D. PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America*, 1967, **41**, 782-787.
- WETHERILL, G. B. Sequential estimation of quantal response curves. *Journal of the Royal Statistical Society*, 1963, **25** (Series B), 1-48.
- WETHERILL, G. B., & LEVITT, H. Sequential estimation of points of a psychometric function. *British Journal of Mathematical Statistical Psychology*, 1965, **18**, 1-10.

#### NOTES

- Used first by Taylor and Creelman (1967), and later, by Findlay (1978) and Pentland (1980).
- A somewhat better estimate of the threshold, and one that is unbiased, can be made by modifying this Wetherill tracking program to use the average of only the last eight peaks and valleys.

#### APPENDIX

The approach taken to the problem of determining a threshold is to maximize the information gained with each measurement. If the amount of information gathered with each measurement is maximized, then the smallest possible number of measurements will be required.

For any value  $x$  of the independent variable, there is a probability  $P_+$  of a positive answer. Given  $n$  samples taken at  $x$ , of which  $p$  were positive, our estimate of  $P_+$  is  $p/n$ , the variance is  $P_+(1 - P_+)/n$ , and the width of the confidence intervals,  $CI_{P_+}$ , about the estimate of  $P_+$  is

$$CI_{P_+} = k \sqrt{\frac{p/n(1 - p/n)}{n}}$$

where  $k$  depends on the level of confidence desired (e.g., 90%, 95%, etc.). The range of the independent variable  $I$  corresponding to this confidence interval,  $CI_I$ , is the slope of the response curve,  $dI/dP_+$ , times  $CI_{P_+}$ , the width of the confidence interval in  $P_+$ :

$$CI_I = (dI/dP_+)(CI_{P_+}). \quad (1)$$

Thus, to minimize the range of the independent variable  $I$  for a given number of samples, sample the point at which Equation 1 is minimized, that is, the point at which the variance of the response variable times the slope of the response curve is at a minimum. In order to estimate the position of this point, use the maximum-likelihood estimator, because it is known to be the most efficient unbiased estimator.

In the normal sigmoid-shaped psychophysical function, Equation 1 is minimized at the "center" of the function, which, for the discrimination tasks, is the 50% positive response point. Thus, the maximum amount of information about the threshold is gained when sampling is done at this point. Our strategy, then, is to obtain the best possible estimate of the 50% point and to sample there. This we do by calculating the likelihood of the 50% point's being at each point within the independent vari-

able's range and taking as our estimate of the 50% point the location that is the most likely point in that range. After  $n - 1$  measurements, we find the  $n$ th measurement,  $m_n$ , by solving

$m_n =$

$$\max_{x \in (a,b)} P[x \text{ is 50\% point} | (m_1, r_1), \dots, (m_{n-1}, r_{n-1})], \quad (2)$$

where  $(a,b)$  is the range of the independent variable  $x$  and  $(m_j, r_j)$  denotes the results of the  $i$ th measurement that was taken at value  $m_i$  of the independent variable. The value of  $r_i$  is +1 if the observer gave a positive response and -1 if a negative response was obtained.

For the case of the sigmoid-shaped logit function, Equation 2 becomes

$$m_n = \max_{x \in (a,b)} \prod_{j=1}^{n-1} (1.0 + e^{-r_j(m_j - x)})^{-1}.$$

Before each measurement we compute  $m_n$ , which is the maximum-likelihood estimate of the position of the 50% point on the response curve, and then take the  $n$ th measurement at Position  $m_n$ . Note that at the start we know that a measurement taken at Point  $a$  of the independent variable's range will give a negative response and a measurement at Point  $b$  will give a positive response. Thus, in this case, the first measurement is taken at the center of the independent variable's range.

(Received for publication January 15, 1982;  
accepted January 20, 1982.)