Methodological examination of the PEST (parametric estimation by sequential testing) procedure: II

IRWIN POLLACK†
University of Michigan, Ann Arbor, Michigan 48104

Finite-state Markov sequences were constructed by a small digital computer, translated to interval-coded electrical pulse trains, and converted to sound. An adaptive stimulus programming procedure of variable step size, PEST, was employed to obtain interstate interval thresholds as a function of restrictions upon the internal structure of the sequences. These thresholds are not independent of the exit criterion for the adaptive procedure. An interactive approach is suggested for determining the exit criterion in order to protect the data against arbitrary decisions made by the E.

One of the recent giant steps in psychophysical methodology is the development of adaptive stimulus programming procedures (Smith, 1961). In my laboratory, I employ an adaptive stimulus programming procedure of variable step size, called PEST (Point Estimation by Sequential Testing, Taylor & Creelman, 1967). I am especially fond of this particular procedure because, when implemented upon a small digital computer, the procedure permits the operation of a truly automated psychophysical laboratory. After our listener tells the computer what test he wishes to run, and his identification number, the computer generates the stimulus materials, runs the adaptive psychophysical procedure, and as the *piece de resistance* terminates the test when "sufficient" (to be defined) data have been collected (Pollack, Headly & Maas, 1966).

Since I employ the procedure nearly continuously, around the clock and around the calendar, I often pause and wonder uneasily if things are going too well. A procedure that runs so smoothly must necessarily have some hidden drawbacks. The procedure clearly invalidates Pollack's law ("Nothing is easy")!

This report represents a second pause to reexamine possible difficulties with the PEST procedure. The previous examination (Pollack, 1968a) employed Uttal's (Uttal & Krissoff, 1966) well-known "gap" test. In that test, clear qualitative differences are reported when sufficiently large temporal discontinuities ("gaps") are introduced. An entirely different auditory task is here employed. This task is associated with the subtle changes in the perceived quality of auditory pulse trains when restrictions are imposed upon finite-state Markov generators. Under many conditions, clear qualitative differences are rarely reported. Presumably, such a test might be an extreme test of procedure. In addition, the present tests tend to avoid a serious difficulty encountered in the previous tests. The previous gap tests pushed the lower limit of temporal control (0.375 microsec) available, and thereby, resulted in an asymmetric or truncated distribution of responses about the final thresholds. Asymmetric distributions will necessarily yield shifts in mean thresholds as a function of parameters of the procedure. Dr. Martin Taylor was kind enough to point this out to me without exposing my errors to public view. In all fairness to PEST, however, the previous errors should be acknowledged.

METHOD

Underlying Task

A finite-state Markov generator developed sequences within defined statistical restrictions (Pollack, 1968b). The sequences were translated to interval-coded electrical pulse trains and converted to sounds by means of earphones (Koss PRO-4). The task of the listener was to pick out which one of four auditory pulse trains was generated by a different set of restrictions from the other three.

Sequences were defined in terms of: m, the number of finite states; b, the length of block over which each of the m states occurred equally often; and N, the total number of items in a sequence. For example, with m = 2, b = 6, N = 600, a sequence of 600 items was made up of 100 successive blocks of 6 items, each block representing a scrambling of 3 representations of each of 2 states. Typical blocks are: aababb, ababab, bbaaab, etc. In all tests, N = 600-612 intervals. Each pulse train was about 300 msec in duration. The states were encoded in terms of the interval between successive brief pulses. The mean interpulse interval averaged over all states was 0.502 msec.

A given observation consisted of four auditory pulse trains: three with a smaller block length, S, and one with a larger block length, L. The task of the listener was to identify which one of four pulse trains employed the larger block length.

Experimental Variable

PEST varied the difference in interpulse interval, DIPI, between successive states. When S responded incorrectly, DIPI was increased; when he responded correctly, DIPI was decreased. PEST manipulated DIPI to converge upon 50% correct response in a four-interval forced-choice test. An intuitive feel for the use of DIPI is that when DIPI = 0, all of the states of a sequence are identical and no sequential information can be gleaned from the sequence. It is assumed that discriminability among sequential restrictions is related to interstate interval differences, DIPI. DIPI was manipulated in steps of 3/8 microsec.

The rules for PEST were those given in the earlier publication (Pollack, 1968a, p. 206). At the advice of Dr. Douglas Creelman, I also incorporated Taylor and Creelman's Rule 4—a modification of the doubling rule—which permitted more efficient convergence. The essence of the PEST procedure is that the magnitude of variation of the manipulated variable, the step size, changes when the threshold region is approached by successive right and wrong responses.

The main problem is when to stop the adaptive procedure, i.e., when is sufficient information available to estimate the threshold? This problem is handled in the PEST procedure in terms of the exit criterion, E.C.—the magnitude of movement of the manipulated variable, the step size, at which the procedure is terminated for a given experimental condition. If a very fine criterion is demanded, we may take many more trials than if a coarser criterion is demanded. Moreover, it does not necessarily follow that a "more accurate" answer is obtained with a finer criterion. If the criterion is too fine, only a random combination of correct and incorrect responses might drive the manipulated variable below E.C.

In the PEST procedure, the movement of the manipulated variable, the step size, is in terms of multiples of E.C. In the

---

*This research was supported in part by Grant GB 6148 from the National Science Foundation. The writer is indebted to M. M. Taylor and C. D. Creelman for comments on the earlier examination of PEST, to Mrs. K. Weidner for supervising the experimental tests, and to Lon Radin for writing the computer program.

†Mailing address: Mental Health Research Institute, University of Michigan, Ann Arbor, Michigan 48104.

Table 1
Threshold DIPI Scores for Several Experimental Conditions and Exit Criteria (EC)

EC			4-6-2[a]	6-9-3	8-12-4
1[d]			146	106	62[b,c]
2				108	128*
4			241	228*	132*
8			618	369	283
16				589*	524*
	2-4-2	3-6-3	4-8-4	6-12-6	8-16-8
1	53	55	44	27	24
2		46	65		
4	75	77	73	84	48*
8	101	86	90	123*	123*
16		129	91		
	2-6-2	3-9-3	4-12-4		
1	36	30	21		
2		30	25		
4	46	37	18		
8	62	31	34		
16		76	28		

[a] *The experimental condition is designated in terms of S, L, m. Thus, 4-6-2 implies S = 4, L = 6, m = 2.*
[b] *Threshold DIPI expressed in 3/8 usec units for each entry.*
[c] *Average IPI was 0.502 msec for all conditions*
[d] *Initial starting values at 31, 63, 127, 255, and 511 temporal units for EC = 1, 2, 4, 8, and 16, respectively.*
**An asterisk designates experimental conditions in which the level of DIPI exceeded 1,966 units/(m − 1) or .737 msec/(m − 1) upon 20% or more trials.*

present tests, for example, the maximum step size is 32 times E.C. The initial step size is 32 times E.C. minus 1.

Procedure

A single parameter tape with 45 experimental conditions was prepared. The starting position within this tape varied across 13 Ss, but the sequential order of conditions within the tape was invariant. The experimental variables were scattered through the tape.

Under difficult listening conditions, PEST might call for a magnitude of DIPI which exceeded the capability of the program: $1{,}966/(m - 1)$ units of 3/8 microsec. When the called-for DIPI exceeded the program capability, a single observation was provided at the limiting condition. An incorrect response at the limiting condition terminated the trial. Terminated trials were later repeated. The results represent geometric means across two nonterminated thresholds by each of 13 listeners.

A qualitative feature of the testing deserves attention before examining the results. The assumption that discriminability among the sequential restrictions is related to DIPI is probably reasonable for low and moderate values of DIPI ($<$ 20% DIPI/IPI$_c$). However, at still higher DIPI levels, sequential discriminability appears to be less dependent upon DIPI, and a wide range of answers can emerge from the testing. An analogy is the use of color to bring out certain fine details in a picture. A small amount of color may help to bring out certain structures; more color may bring out more structures; but, beyond a certain color contrast, further color changes may produce little additional effect.

RESULTS

The results are summarized in Table 1. Each underlined experimental condition is designated in terms of S, L, m. Each entry is the DIPI threshold in units of 3/8 microsec as a function of the experimental condition. Separate rows represent conditions with a constant S:L:m ratio.

In general, as the exit criterion, E.C., is increased, the mean threshold is increased. This dependence is striking in the most difficult listening conditions (top row), but the dependence remains suggestive, even in the least difficult listening conditions (bottom row). Of course, when an appreciable fraction of the trials are terminated (conditions represented by an *) and the trials are repeated, an addition of bias enters into the experimental findings. Noteworthy is the reduced dependency upon E.C. when trials are not terminated.

DISCUSSION

How might we overcome the effect of the E's arbitrary choice of E.C. in the PEST procedure? Following the lead of Stevens (1955), I suggested in the initial examination of PEST (Pollack, 1968a) that an iterative approach might be employed. Such an approach was especially effective for removing the effects of stimulus bias in rating scale procedures (Pollack, 1965). Operationally, I employ the following procedures: I first guess at an appropriate value for E.C. On the basis of initial results, I then modify E.C. to approximately 10% of the tentative average threshold. On the basis of further tests, I readjust E.C. by the same rule. The process continues until the tentative threshold stabilizes. Consider how this might be employed when thresholds are monotonically related to E.C., as in 4-8-4 of Table 1. Had we started with an E.C. of 1, we might have obtained a ballpark estimate—depending upon sampling variability and the relative skill of the listeners—of about 44. Had we then repeated the test with an E.C. of 4, we would have obtained an estimate of about 73. Had we then repeated the test with an E.C. of 7, we would have obtained an estimate of about 90. Further modifications would not have changed the estimate. However, had we initially chosen an E.C. of 16, we would have converged upon a stable answer within a single iteration. In practice, only one or two revisions of E.C. are usually needed to stabilize PEST thresholds. In practice, an upper bound on E.C. is set to permit the listener at least three observations before the trial is terminated.

When thresholds are insensitive to E.C., as in Condition 4-12-4, the iterative approach cannot help stabilize the thresholds, but neither does the approach introduce special difficulties.

The iterative approach to the setting of parameters in the PEST procedure is obviously not a cure-all for all of the ills of psychophysical experimentation. Nor are adaptive programming procedures especially vulnerable to arbitrary decisions made by the E. I suspect that conventional psychophysical procedures are even more vulnerable to such decisions. I also suspect that the PEST procedure can be modified so that it internally adjusts its own parameters to conform with external rules. For example, if PEST exits in too few observations, E.C. might be reduced; if PEST exits too slowly, E.C. might be increased. For example, in the present tests, E.C. was doubled following each 16 successive observations without exiting. In summary, the marriage of efficient adaptive programming procedures with an iterative approach to the setting of their own parameters may be an ideal match for the attainment of relatively unbiased, efficient psychophysical procedures.

REFERENCES

POLLACK, I. Iterative techniques for unbaised rating scales. Quarterly Journal of Experimental Psychology, 1965, 17, 139-148.

POLLACK, I., HEADLY, P., & MAAS, E. Modest computer controlled psychoacoustical facility. Journal of the Acoustical Society of America, 1966, 39, 1248(A).

POLLACK, I. Methodological examination of the PEST (Parametric Estimation by Sequential Testing) procedure. Perception & Psychophysics, 1968a, 3, 287-289.

POLLACK, I. Discrimination of changes in Markov sequences with the first N-th orders statistically independent. Proceedings, 76th Annual Convention, APA, 1968b, 125-126.

SMITH, J. E. K. Stimulus programming in psychophysics. Psychometrika, 1961, 26, 27-34.

STEVENS, S. S. On the averaging of data. Science, 1955, 121, 113-116.

TAYLOR, M. M., & CREELMAN, C. D. PEST: Efficient estimates on probability functions. Journal of the Acoustical Society of America, 1967, 41, 782-787.

UTTAL, W. R., & KRISSOFF, M. Effect of stimulus pattern on temporal acuity in the somatosensory system. Journal of Experimental Psychology, 1966, 71, 878-883.

(Accepted for publication July 20, 1970.)