SESSION II
CONTRIBUTED PAPERS: ALGORITHMIC METHODS

William R. Uttal, *University of Michigan*, Presider

Simultaneous independent threshold estimates: Multiple PEST*

C. DOUGLAS CREELMAN and HOWARD L. KAPLAN
University of Toronto, Toronto, Canada

An adaptive threshold estimation procedure, PEST, has been extended to track sensitivity to many independent signals simultaneously within a single experimental run. This paradigm has been implemented on a PDP-8/S computer, and enables investigation of a number of otherwise experimentally intractible substantive problems. The logic of the new experimental paradigm is described, and some examples are given of its use in auditory research. Results of some computer simulations show the relative efficiency of the procedure and evaluate alternative ways to summarize data.

Consider an experiment which asks the following kind of question: To what extent is human sensitivity decreased if the O attends to one of many alternative possible signals, compared to when he must attend to only one possibility? Another version of the same problem in experimental design might be: To what extent is sensitivity affected if signals are unpredictably presented with one from a set of different kinds of masking or interfering inputs?

EARLIER APPROACHES

One way to investigate this sort of problem has been to set all stimuli to the same physical intensity and to compare performance when signals are presented alone to performance when signals are intermixed. In an earlier study (Creelman, 1960), performance was averaged across the set of alternatives, without separate analyses for different possible signals. More recently, with the same experimental design, performance has been reported separately for each alternative signal (Creelman, 1972). If the possible signals are themselves of different detectability, performance changes become impossible to assess for the very easy and very difficult alternatives. This problem led Green (1961) to preadjust levels on the basis of preliminary data which determined thresholds for each of the signals separately.

ADAPTIVE MEASUREMENT: PEST

Our current approach is based on the adaptive psychophysical procedure PEST (Taylor & Creelman, 1967). Under PEST, the level of a stimulus is adjusted, in a two-alternative forced-choice (2AFC) detection experiment, to yield an estimate of that intensity which results in a preselected proportion of correct responses. Let us, for convenience, call the stimulus level found this way a "threshold." The PEST procedure has a set of rules for deciding, on the basis of a sequential likelihood ratio test (Wald, 1947), whether trials at a particular stimulus level yield a proportion which is lower or higher than the target percentage. PEST invokes a set of rules for changing the stimulus level appropriately in steps of generally decreasing size. When PEST calls for a step which is smaller than a specified minimum size, we call the event a "convergence." In the classical implementation of PEST, the experimental run would terminate at this convergence, and we would use the last signal level as an estimate of threshold.

The subroutines to implement PEST on the Digital Equipment Company PDP-8 series of computers can be acquired from DECUS (Kaplan, Taylor, & Creelman, 1972). A call to these subroutines must specify the address of a list of variables elsewhere in memory from the routines themselves, with a separate list for each individual being run. The PEST history list includes the current stimulus level, current step size, an indicator for the most recent direction of stimulus change, and other necessary data. Only seven locations are required to keep the history for any one O, plus a list of initializing constants, which is ordinarily the same for all Os. The history for a Wald sequential ratio test requires only one variable, in addition to three constants which define the test being used. As the arithmetic required to operate on these lists consists entirely of additions, subtractions, and logical decisions performed on single-precision integers, the program is quite feasible for real-time computation on even a slow, small computer.

MULTIPLE PEST

It is a simple extension to keep a set of separate PEST history lists for each of a number of potential stimulus conditions for a single O, or multiple sets of lists for different Os all run concurrently. We have been able to pack such history lists into four 12-bit words per O per stimulus condition. This extension is the core of our

*This research was supported by the Canadian National Research Council under Grant APA-111.

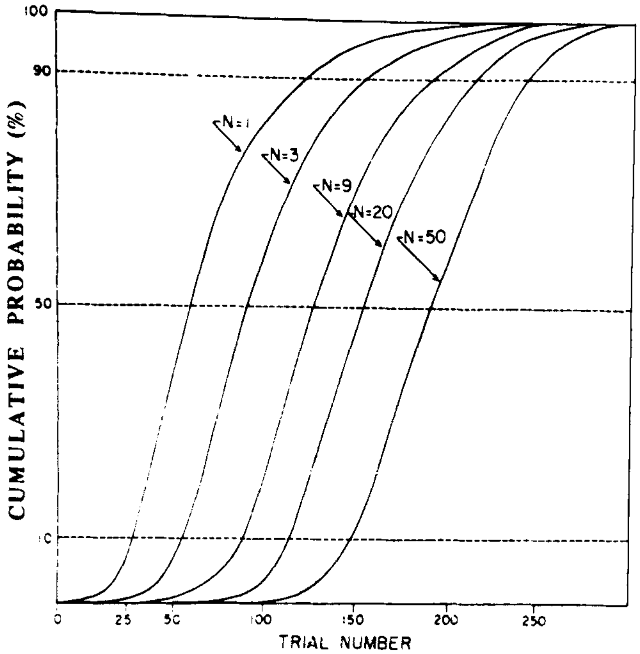

Fig. 1. Probability that all of N simultaneous PEST runs have converged at least once each, as a function of trial number.

solution to the problem of running experiments with multiple possible stimulus events within a single experimental run.

Our experiments are run on a PDP-8/S computer under the control of PSYCLE (Creelman, 1971), a language for computer control of the selection, timing, and presentation of stimuli in discrimination and detection experiments. Data from each 2AFC trial are gathered by the PSYCLE operating system, and on the basis of each O's performance on that trial, the PEST routines calculate the signal level to be used when the same conditions are presented again. These results are saved in the history lists for each O. Before the succeeding trial is begun, the next stimulus is randomly selected and the history for that condition is consulted to determine the current testing level for each O.

Our auditory experiments use a digitally controlled Wavetek Model 157D function generator with both frequency and amplitude set by the computer. Since we have only one generator and we run three Os at once, each with his own stimulus level, stimulus presentations cannot be exactly simultaneous. A single experimental trial has signal presentations interleaved, so that the Wavetek produces signals for each O in rapid sequence within the trial. Our driving programs for the generator allow amplitude control to within ¼ dB and frequency control to three significant decimal digits.

When tracking many thresholds at once under PEST, the gathering of data cannot cease when the stepping rules call for the first convergence, as is the case when a single threshold is wanted. The time to achieve termination is a random variable, and the run must continue until all conditions have given thresholds, with all possible conditions equally likely to be presented on a trial.

In this situation, PEST is run under "RAT" mode (Rapid Adaptive Tracking), where testing is continued even after the first convergence, by taking a minimum allowable step whenever the stepping rules call for anything smaller. Thus, throughout an experimental run, all stimulus conditions are equally likely, and the further into the run we have gone, the more likely it will be that any stimulus selected will be presented in the region of its own threshold.

Summarizing Data

Estimating an average threshold from the data under RAT-mode tracking is somewhat of a problem compared with the usual procedure, where the threshold is taken simply to be the signal level at the first convergence. We have gone separate ways in developing schemes to summarize the data, lacking at the outset any reason to prefer one to the other. The multiple PEST programs have a background task, typing out on the Teletype each change of level for each condition being run when it occurs. Steps which represent initial convergence of a PEST run are marked, as are all subsequent convergences. One of us simply used the PEST history to extract the testing level at each convergence and averaged those levels for a threshold estimate. Testing under this scheme was terminated when each stimulus condition had yielded at least one convergence for each of the three Os. The other author simply waited for a decent number of trials (about 150) and then began to sample the current level of each of the concurrent PEST runs at a fixed rate, irrespective of the stepping history. It should be clear that under either scheme the successive estimates of threshold are correlated, and that the summary estimate is therefore less precise than one obtained from an equal number of independent estimates. However, the number of trials needed to generate each estimate is considerably less than that needed for independent estimates.

MONTE CARLO SIMULATIONS

We have recently completed some simulation experiments, assessing these alternative schemes. In the simulations, the O is represented by a theoretical psychometric function which is placed randomly relative to the intensity at which testing is started and which is a realistic approximation of what our Os might give.

A first question for the simulations was the expected length of an experimental run under the first data-summarizing scheme, where testing continues until all conditions have converged at least once. In Fig. 1, we show the cumulative probability that all of a set of simultaneous PEST runs will have terminated as a function of the number of trials per run. The number of simultaneous PEST histories is the parameter. When a single threshold is being estimated, the left-hand curve of the figure shows that the median number of trials to

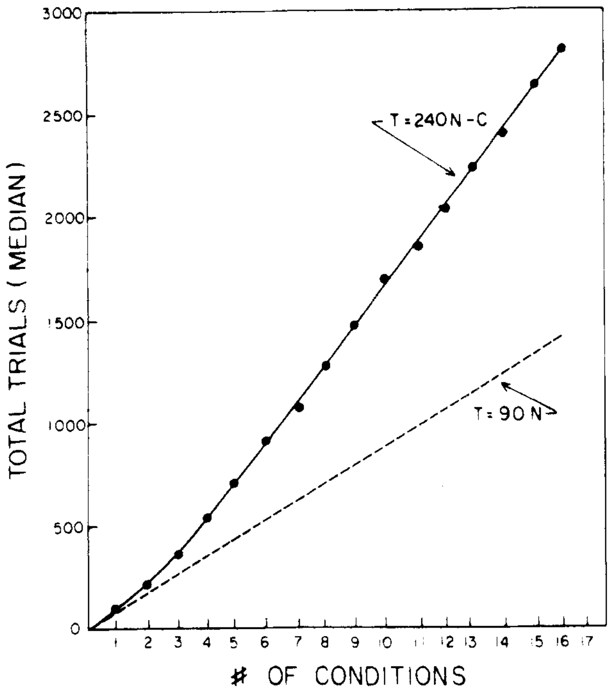

Fig. 2. Median total trials for each of three Os being run under 3n stimulus conditions, n per O.

reach convergence was about 60 trials. With three simultaneous PEST runs (i.e., with three Os each having a single threshold estimated), the median number of trials for the longest condition to converge is about 90 trials.

From these data, we can derive the expected total length of an experimental run as a function of the number of different thresholds being estimated for each of three Os. The result is shown in Fig. 2, where the dashed line shows the median total number of trials with N successive PEST runs, each allowed to converge at least once. The solid line shows our simulation result that each stimulus condition added to the simultaneous experiment adds about 240 trials to the median length of the total experiment. The abscissa of the figure goes to 17 conditions, the largest number we have tried to run simultaneously. The marginal cost in added experimental time will change with E's choice of minimum step size and other parameters of the sequential test used, but we feel confident that the overall pattern will remain the same for different choices of parameters. When multiple simultaneous threshold estimates are being gathered, there is considerable cost in the relative time to collect the data.

On the other hand, while waiting for the final stimulus condition to achieve convergence, the earlier-converging conditions continue to provide threshold estimates, raising the possibility of greater precision compared to a single convergence. Figure 3 shows the gain in information or precision (the reciprocal of variance) of the mean threshold estimate as a function of the number of trials in the experiment.

In this figure, the precision of the two summarizing procedures is compared. The right-hand ordinate is a guess, for our own benefit, as to realistic standard deviations, in decibels. The two sets of data are from the same identical set of 2,500 simulated PEST runs. The fixed-sample procedure allows experiments of known length and produces more precise estimates for the same number of trials.

We have investigated the effect of different fixed sampling rates on the stability of the average threshold estimate. The question might be phrased: How long can one wait between samples before beginning to lose information? Alternatively: How far apart must samples be before successive estimates are sufficiently uncorrelated for each new sample to add information? The answer seems to be that with the testing parameters we used, precision seems not to improve with sampling more often than once every 16 trials. Combined with the results on the relative efficiency of the two types of sampling scheme, this implies both that convergences must appear, on the average, less often than once every 16 trials and that each convergence cannot be a much more precise estimate of threshold than an arbitrarily selected trial.

The absolute values of the simulation results will vary, depending on the free parameters of the PEST procedure. Both the stringency of the sequential test and the size of the minimum step allowed will affect the precision of the threshold estimates and the number of trials taken to get the data. Overall, however, we find the tradeoff between accuracy and effort reported by Taylor and Creelman (1967) to hold here as well. With realistic assumptions as to the slope of the psychometric function (8 dB between .55 and .95 on the simulated 2AFC function), we feel reasonably certain that threshold estimates should be accurate to less than 0.5 dB in multiple estimation situations where about 250 trials are devoted to each possible stimulus

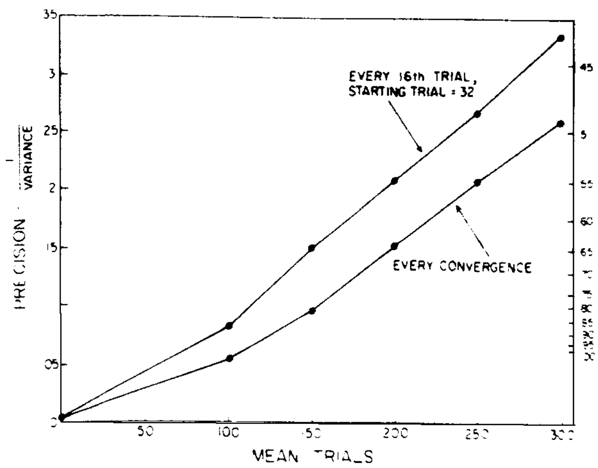

Fig. 3. Relative precision of two averaging schemes, with estimated standard deviation in decibels as alternate ordinate.

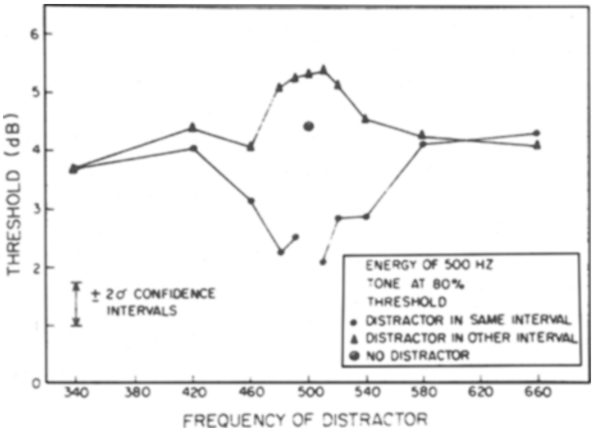

Fig. 4. Threshold of a 500-Hz tone as a function of distractor frequency and temporal interval for a typical O.

condition, assuming thresholds which are stable and not systematically changing over time.

SOME APPLICATIONS

We have used the multiple PEST procedure to investigate two very different substantive problems in psychoacoustics. The first is the effect on sensitivity of having Os listen for many alternative frequencies at once, as described above. In this case, the procedure has paid its way by yielding data quite unlike any reported in the literature heretofore, and methodology has contributed directly to better understanding of the nature of auditory attention sharing. In brief, the data seem to show that listening for many alternative signals raises threshold by only 3 dB or so, irrespective of the number of alternatives being listened for and regardless of the range of frequencies over which they may appear. These data are quite unlike the pattern we see when signals are presented at fixed levels. The data put severe constraints on models for auditory attention sharing and pose interesting questions about why other experimental paradigms show much greater apparent loss.

The second application has been to estimate threshold when a 500-Hz signal in white noise can appear in a 2AFC trial along with another "distractor" tone, and the tone may appear either in the interval with the signal or in the other interval. The frequency of the distractor is unpredictable. Trials are intermixed such that any condition can happen on any one trial. The threshold intensity of the signal under these conditions is shown in Fig. 4, where the frequency of the interfering tone is plotted on the abscissa and the parameter is whether the extra tone was in the interval with the signal or in the opposite interval. The level of the interfering tone was constant at 0 dB on the graph, about 4 dB below the threshold of the 500-Hz tone presented without any distractor tone. One way of considering these data is that they provide a direct, graphic estimate of the dimensions of the auditory critical band.

OVERVIEW

The multiple PEST paradigm allows us to move one step closer to realistic settings and problems in psychophysics. We are certainly not the first to suggest that effects of context on performance can be more interesting and important than the performance itself. We have implemented a particularly interesting context, both to the E and to the O. An E is interested in signal levels at which performance changes with experimental conditions; an analogous interest holds for the O. An O may not choose to attend where he cannot perceive any stimulus, nor may he need to focus where the stimulus is intense enough to command attention anyway. The multiple threshold tracking technique increases the probability that the E and O share a common definition of the task.

REFERENCES

Creelman, C. D. Detection of signals of uncertain frequency. Journal of the Acoustical Society of America, 1960, 32, 805-809.

Creelman, C. D. Rapid response and flexible experimental control with a small on-line computer: PSYCLE. Behavior Research Methods & Instrumentation, 1971, 3, 265-267.

Creelman, C. D. Detecting signals of uncertain frequency: Analysis by individual alternatives. Journal of the Acoustical Society of America, 1972, 52, 167. (Abstract)

Kaplan, H. L., Taylor, M. M., & Creelman, C. D. PEST/WALD/PINIT: Adaptive psychophysics testing package. DECUS Program Library, No. 8-520, 1972.

Green, D. M. Detection of auditory sinusoids of uncertain frequency. Journal of the Acoustical Society of America, 1961, 33, 897-903.

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

Wald, A. *Sequential analysis.* New York: Wiley, 1947.