Perception & Psychophysics
1989, 46 (4), 345-350

A comparison of three adaptive psychophysical procedures using inexperienced listeners

J. A. STILLMAN
The University of Auckland, Auckland, New Zealand

Thresholds for a masked 1-kHz tone were obtained in single sessions from 60 inexperienced listeners. A two-alternative forced-choice procedure was used in conjunction with one of three adaptive psychophysical techniques. These techniques comprised two staircase techniques targeting either 70.7% or 79.4% correct detection (Staircase-71 and Staircase-79), and parameter estimation by sequential testing (PEST), targeting 80% correct detection. Listeners were provided with a rationale for maintaining concentration at weak signal levels. Similar threshold values were obtained from the Staircase-79 and PEST groups in equal numbers of trials. The degree of oscillation in the level of the signal around the value finally chosen as the threshold was comparable for both staircase techniques. Subsequent fixed-level testing did not provide a true indication of the subjects' capabilities. The amount by which percent correct in fixed-level testing differed from expectations based on adaptive testing varied among the techniques. Additional thresholds were obtained in a second session from 30 of the original subjects. Thresholds with both staircase techniques improved by about 1 dB on retest, while thresholds with PEST were constant across sessions. The variability of the data compared well with that from studies involving experienced listeners.

In psychophysical investigations necessitating a datum at a particular location on the psychometric function, the threshold is often determined by means of an adaptive procedure. Such procedures have in common the fact that the value of the stimulus is adjusted on a trial-by-trial basis, depending on the subject's responses to preceding stimuli. Various adaptive methods used in psychoacoustics have been the subject of comparisons and analyses with respect to efficiency, reliability, or bias, using both empirical data (Hesse, 1986; Kollmeier, Gilkey, & Sieben, 1988; Shelton, Picardi, & Green, 1982; Taylor, Forbes, & Creelman, 1983) and computer simulations (see for example Emerson, 1984; Hall, 1981; Pentland, 1980; Taylor & Creelman, 1967). From such studies, the relative practical and theoretical merits of a number of procedures can be ascertained and evaluated against chosen experimental design criteria.

A feature of most of the experimental studies listed above is the use of well-practiced listeners. The untrained group in Taylor et al. (1983) is an exception. On occasion, normative data are required from groups of individuals from whom extended practice cannot be expected. Such groups might comprise, for example, young children or individuals with cortical lesions associated with hearing impairment. The purpose of this research is to examine the potential application of modern, rigorous psychophysical procedures to such groups, and the study comprises an experimental comparison, using naive listeners, of three adaptive psychophysical techniques commonly used in psychoacoustics. Each technique constitutes an adaptive version of the two-alternative forced-choice (2AFC) procedure. The comparison is intended to ascertain the validity and efficiency of measures obtained from untrained listeners in only one or two experimental sessions.

Two of the procedures each utilize one of the several possible rules for changing the level of the signal employed in an adaptive staircase technique (Levitt, 1971). The particular rules implemented are those leading to 70.7% correct detections and 79.4% correct detections. The third procedure is PEST (parameter estimation by sequential testing; see Taylor & Creelman, 1967). These three procedures, henceforth referred to as Staircase-71, Staircase-79, and PEST, were chosen because of their wide use. The same procedures were among those compared by Kollmeier et al. (1988), using experienced listeners.

The following conventions are observed throughout this paper: A *single* trial refers to a sequence in which the listener monitors two brief intervals of time and registers a decision as to which interval contained a faint auditory signal. A *block* of trials consists of all the trials undertaken by a subject for a single threshold estimate, although not all of these trials are used to compute the threshold. The actual number of trials contributing to a threshold estimate depends on the procedure, as will be explained in the following sections.

The rules for changing the level of the stimulus with the staircase procedure specify a reduction in the level of the signal following two correct responses when the target level is 70.7% correct, and a reduction in the sig-

Thanks are extended to M. M. Taylor, who provided helpful information on the PEST psychophysical technique. Correspondence may be addressed to Jennifer A. Stillman, who is now at the Department of Psychology, University of Otago, P.O. Box 56, Dunedin, New Zealand.

nal level following three correct responses when the target level is 79.4% correct. In either case, the signal level is increased following a single incorrect response. A run comprises a sequence of changes in the level of the stimulus in one direction only. Within runs contributing to the final threshold estimate, the step size is fixed. Usually the first one or two turnaround points (stimulus values at which a change in direction occurs) are considered settling values and do not contribute to the threshold estimate. The average of the midpoints of every second run is accepted as the threshold level. This means that drifts in the location of the psychometric function during a trial block are ignored when the threshold is computed.

PEST uses information from the most recent testing level to decide whether to change the level of the stimulus. In accordance with a sequential likelihood-ratio test (Wald, 1947), a change occurs when the absolute value of the actual number of correct responses at the current testing level, $N(C)$, minus the expected number correct at the chosen probability, $P_t \cdot N(T)$ (where P_t is the desired probability and $N(T)$ is the number of trials completed at the current testing level), equals or exceeds a constant deviation limit, W. That is, when

$$|N(C) - P_t \cdot N(T)| \geq W.$$

The step size is variable throughout a block in accordance with predetermined rules (Taylor & Creelman, 1967) and depends on the response history up to the current trial. A block of trials terminates when the next step would be smaller than a chosen minimum step size. The level at which a succeeding stimulus would be presented is accepted as the threshold level and is not tested.

In the present study, no attempt is made to equate the number of trials across the different techniques, since the effort demanded from participants in the normal course of events is an important variable—particularly for groups of listeners with limited attentional resources. It is obvious a priori that for the same number of turnaround values the "1 up 2 down" rule used to obtain 70.7% correct will require fewer trials than the "1 up 3 down" rule used to obtain 79.4% correct. However, this advantage for the lower target must be evaluated against a loss of efficiency and an increased instability in the measures (Kollmeier et al., 1988).

METHOD

Subjects

Sixty undergraduate students at the University of Auckland volunteered to participate in the experiment. All were inexperienced in psychoacoustical procedures, and none was aware of any hearing defect.

Apparatus

The noise, produced by a General Radio noise generator Model 1381, was band-pass filtered with a Khron-Hite Model 3550R filter before being mixed with the signal. The signal was produced by a Data Pulse function generator (Model 410) and was routed to the gate through a Wavetek 852 filter and a Marconi TF 2162 attenuator. The level of the signal was set by a programmable attenuator, Charybdis Model D, under computer control. The computer (6809 processor) also recorded the listeners' responses and calculated threshold values. The signal and masker were presented monaurally through one member of a pair of TDH 49 headphones. The listeners were seated before a response panel in a sound-attenuating booth.

Stimuli

The masker was a band-pass noise centered at 1000 Hz with 3-dB cutoffs at 40 Hz and 1960 Hz. The noise was on continuously during the experiment, at a spectrum level of 35 dB SPL. The signal was a 1-kHz sinusoid gated with 100-msec linear onset and offset ramps. The duration of the signal was 400 msec.

Procedure

Part 1

The 60 participants in the experiment were randomly assigned to one of three groups, namely Staircase-71, Staircase-79, or PEST. Thus, for each procedure, data were obtained from 20 listeners. Sessions were planned so that both the instructions to the listeners and the collection of data could be completed within 1 h. By and large, testing was scheduled at times convenient to the participants, so that the collection of data from each of the three groups was interleaved and spread out over the entire 8-week period during which the experiment was conducted.

Listeners in each group completed five blocks of trials before being presented with two 50-trial blocks at a fixed level selected on the basis of performance in the adaptive procedures. The initial signal level for the first adaptive block of the session was the same for all listeners and was set at a level estimated to be about 12 dB above threshold. This level was chosen so as to provide the inexperienced listener with an adequate preview of the test stimulus while not unreasonably prolonging the time required to complete a block of trials. For each block of trials after the first block, the initial level of the signal was 12 dB above the threshold level obtained in the preceding block. This strategy, in combination with the rules applied to each procedure, meant that although the number of trials to threshold varied among the procedures, usually three above-threshold levels of the stimulus were visited prior to the first reversal.

The sequence of events on each trial was as follows. Every trial began with a 500-msec warning light followed after 500 msec by two 400-msec observation intervals marked by lights. The two intervals were separated by a 500-msec pause. No further events occurred until a response was made. Responses were followed immediately by a 300-msec feedback light marking the interval in which the signal had occurred. The two blocks of trials in which the stimulus was presented at a fixed level immediately followed the adaptive testing.

Parameters pertaining to each procedure were chosen in accordance with the outcomes of previous studies and to accord with what appeared, from a survey of published work, to be common usage. Details are given below.

Adaptive staircase. A step size in the range 2 to 8 dB is commonly used prior to the first turnaround in a block of trials. A first-run step size of 5 dB was chosen for the present study, and the step size was reduced to 2 dB following the first turnaround. The termination of a block of trials occurs either after a set number of trials, typically between 40 and 80, or after a set number of turnarounds. In this study, a trial block terminated once the listener had produced 16 turnaround values, and the last 12 of these were used in the calculation of the threshold. Terminating a trial block after a set number of turnarounds, rather than after a set number of trials, ensures that each threshold is based on the same quantity of information. Choosing even numbers of turnaround values avoids bias in the estimate.

PEST. The initial step size was set at 4 dB, and the maximum permitted step size was 6 dB. The stopping rule found optimal by Shelton et al. (1982), 0.5 dB, was adopted. The program estimated the 80% point on the psychometric function, and a deviation limit of 1 was used. In addition to facilitating comparison with the staircase procedure aimed at 79.4% correct, the elected level had the advantage of being safely above the level at which the efficiency of the procedure begins to decline when a 2AFC task is used (Taylor & Creelman, 1967, Figure A2).

Fixed-level trials. During the two 50-trial blocks at a fixed level, the signal was presented at the average of the listeners' lowest pair of the final three threshold estimates determined with the adaptive procedure. The total number of correct responses from the two blocks thus constitutes percent correct.

Part 2

In Part 2 of the experiment, 10 listeners from each of the three groups were chosen on the basis of availability and tested a second time. The retests took place at least a week after the first session and involved repeating both the adaptive and fixed-level tasks. These retests were carried out so that a comparison with the earlier results might provide an indication of the validity of data gathered in an initial session with naive listeners.

Subject Variables

A lack of attention, especially on the part of inexperienced listeners, might occur when the level of the stimulus is close to threshold, because the listener is not motivated to do other than make a random guess as to the correct interval, and consequently no longer maintains concentration on the task. A better understanding of the task on the part of the subject should alleviate this problem. Accordingly, it was explained to each listener that even if he or she were unable clearly to distinguish the signal in the noise, the samples of sound in each observation interval would not be absolutely identical, so that on each trial there was a basis for a considered choice. This explanation was given interactively and adapted to the responses of each listener. Sometimes the spectrum analyzer was used to illustrate the differences between two samples of noise. The discussion was intended to convince the listener that although the difference between two observation intervals might be minimal, it would never be zero. Listeners were instructed that, if they felt unable to distinguish the signal in the noise, they were not to guess at the correct interval but were to choose whichever interval produced the stronger sensation. If subjects can be persuaded to concur with these instructions, inattention and bias at near-threshold levels of the stimulus should be minimized. The theory of signal detection underlies most psychophysical procedures and makes the assumption that in 2AFC tasks the subject's decision is based on the strength of an internal variable. However, without a rationale for such behavior, an inexperienced listener may see no point in continuing conscientiously to make a considered choice between the two intervals when the level of the stimulus is very low. Indeed, subjects frequently comment that they are "just guessing." In the grip of such a belief, listeners no longer concentrate sufficiently to identify the correct interval, even when the signal has reached a level at which, with effort, they could do so.

Instructions given to listeners must of necessity vary depending on the subject pool. For example, presenting the task as a game is an effective way to ensure that children maintain motivation to make a correct response (Abramov et al., 1984; Irwin, Stillman, & Schade, 1986). However, some simple form of the explanation used above is useful, because at low stimulus levels even motivated subjects may feel that deliberate choice will not advance either their own or the experimenter's objectives. (Obviously, although intermittent lapses of attention are likely to increase the length of a trial block, and so are not in the subject's best interest, any adaptive procedure will ultimately limit the number of correct responses that the listener can make.)

A sense of unease may also adversely affect the performance of a listener unused to psychoacoustic procedures. For example, a subject may feel intimidated in the presence of the numerous pieces of technical apparatus found in a typical psychoacoustic laboratory. Furthermore, the interior of a sound-attenuating booth constitutes a sterile and somewhat inhospitable environment. On the assumption that a relaxed subject will be able to cooperate more effectively with the investigator's instructions, efforts were made to put each participant at ease in the laboratory. These attempts consisted of purposely spending a few moments engaging the subject in general conversation unconnected with the experiment, and of paying attention to the interior of the sound-attenuating booth. To improve the experimental environment, a small vase of flowers was placed on the table that contained the response panel, and, in addition, a scenic poster was affixed to the cabin wall. This poster was positioned to one side of the response panel, so as not to distract the subject during the actual running of a block of trials. In addition, a small book of jokes was placed on the table for listeners to read while the data were printed out and the apparatus reset between blocks of trials.

RESULTS AND DISCUSSION

Part 1

Threshold Signal-to-Masker Ratio

A split-plot factorial analysis of the data from 60 listeners showed, as expected, a significant difference among the three procedures with respect to threshold signal-to-masker ratio [$F(2,58) = 19.55, p < .0001$]. The mean signal-to-masker ratios were Staircase-71 = 14.23 dB; Staircase-79 = 16.02 dB, and PEST = 16.02 dB. An increase of approximately 9% in the value of the psychometric function thus resulted from a 1.79-dB increase in the signal-to-masker ratio. This result is consistent with expectations for a masked 1-kHz tone. The analysis showed that thresholds did not differ significantly across the 5 blocks of trials for any of the three procedures. The data are displayed in Figure 1, which shows the threshold signal-to-masker ratios in each of the 5 blocks of trials for each procedure. Each data point represents the aver-

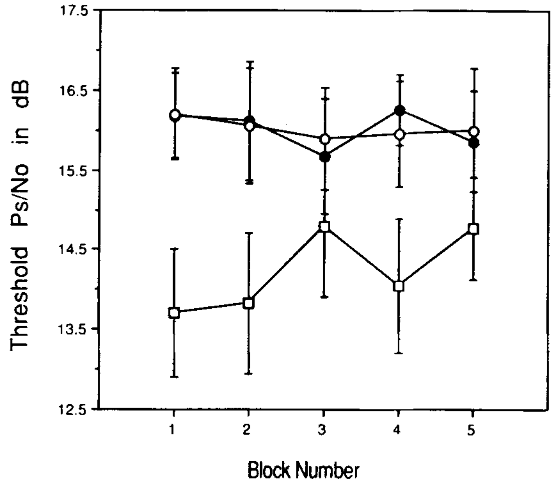

Figure 1. Threshold signal-to-noise ratio (P_S/N_O) in dB, as a function of block number with target percent correct for three psychophysical techniques as parameter. The signal was a 1-kHz sinusoid centered in a broadband masking noise. Each data point represents the average from 20 listeners. The different procedures are represented by the following symbols: open squares, Staircase-71; closed circles, Staircase-79; and open circles, PEST. Vertical bars are the 95% confidence intervals.

age from 20 listeners. The individual average standard errors of the threshold levels over observers were: Staircase-71 = 0.75 dB, Staircase-79 = 0.47 dB, and PEST = 0.54 dB. These statistics compare favorably with those obtained by Shelton et al. (1982) over 16 blocks of trials with experienced listeners, especially since this statistic is the average individual standard deviation divided by the square root of the number of blocks.

Trials Per Block

As expected, there was a significant difference between the procedures with respect to number of trials per block [$F(2,57) = 25.05$, $p < .0001$]. A Tukey hsd test confirmed that the difference was between Staircase-71 and the other two techniques. The average numbers of trials per block were: Staircase-71 = 54, Staircase-79 = 75, and PEST = 80. For each procedure, the number of trials in the first run is easily manipulated by changing the initial step size, or initial sequence. Therefore it is useful to compare the number of trials needed to complete a block once the first turnaround has occurred. Since, on average, three signal levels were visited prior to the first incorrect response, a further 48 trials were required for Staircase-71 (54−6), 66 trials for Staircase-79 (75−9), and 65 trials for PEST (80−15). The respective rules for a change in the level of the stimulus determine that the expected ratio between the number of trials per block after the first reversal with Staircase-71 and Staircase-79 will be 3:4.[1] The obtained ratio of 48:66 is close to this value and indicates that the tracking behavior of both groups was comparable. In each case, the number of trials is approximately twice the minimum possible for 16 turnarounds.

The average number of trials per block with PEST was influenced by the occasional occurrence of very long runs. Although the number of trials per block was not significantly greater with PEST than with Staircase-79, occasional very long trial blocks with PEST were a problem. If the number of trials in a block reached 150, the block was terminated and restarted. This occurred on 3% of the blocks. On these occasions, the original 150 trials were added to the number of trials completed on the second attempt. The average number of trials per block given above (80 trials) is thus inflated when compared to the average length of a trial block on the 97% of occasions when a repeat was unnecessary (75 trials). Although the number of long blocks was small, their effect was not inconsiderable. If a subject has a limited attention span, or cannot remain beyond the allotted time, then the possibility that a long run might occur has an adverse effect upon the experimenter and thus upon the relaxed atmosphere in which the experiment is conducted. Furthermore, the termination of a long block of trials without a result has a negative influence on the subject. This problem might be overcome by a modification to the PEST procedure (Kalpan, 1975, cited in Taylor et al., 1983) that involves averaging every *n*th trial after the first reversal. However, if lapses of attention are anticipated, as might be the case with inexperienced listeners, such an estimate could involve bias.

Percent Correct in Fixed-Level Blocks

For each group, the average percent correct in fixed-level trials fell short of the targeted value. The obtained values were Staircase-71 = 66.2%, Staircase-79 = 75.3%, and PEST = 72.0%. Only Staircase-71 and Staircase-79 differed significantly from each other on a Tukey hsd test. Thresholds with adaptive testing overestimated the listeners' performances in fixed-level trials by 4.5% and 4.1% for the staircase techniques, and by 8% for PEST. The standard deviation for percent correct was 6.8% for Staircase-71, 8.5% for Staircase-79, and 8.6% for PEST.

Underestimates of the capability of the listeners with fixed-level testing have also been found in studies involving experienced listeners (Kollmeier et al., 1988; Shelton et al., 1982). Such outcomes are usually attributed to the fact that in fixed-level testing at weak stimulus levels, the listener tends to lose track of the signal and becomes unable to focus attention on the signal frequency. The effect might be expected to be greater with inexperienced listeners; however, using curves fitted to both fixed-level and adaptive data from four experienced listeners, Taylor et al. (1983) estimated that the listeners could be expected to achieve 69% correct on fixed-level trials for signal levels that yielded 80% correct with PEST. Their estimate is thus slightly below the 72% correct achieved by the group of inexperienced listeners whose data are reported here.

The extent to which the performance of the PEST group failed to approximate the target percent correct in fixed-level runs, relative to the Staircase-79 group, is puzzling, even if differences between the two groups in the quality of experience obtained with the preceding adaptive testing is taken into account. The experience of the groups differs in that whereas Staircase-79 presents listeners with levels bracketing the threshold, and terminates after a set number of turnarounds, PEST insists on performance at one particular signal level being maintained at the predetermined percent correct detections. Staircase-79 and PEST were targeted at almost the same percent correct in adaptive runs, and in fact produced the same threshold threshold signal-to-masker ratio, in approximately the same number of trials. The differences in the experience of the two groups ought either to lower adaptive thresholds with PEST relative to Staircase-79 or, because the PEST group was more familiar with threshold-level trials, predispose this group to better performance in the fixed-level runs relative to Staircase-79. The data are contrary to both expectations.

Part 2

The effect of a second session on the performance of 10 subjects from each of the three original groups is shown graphically in Figures 2 to 4. For each group of listeners, the figures show the threshold signal-to-masker ratio as a function of block number in each session. In Figure 4, one point is the average of 9 rather than 10 measures, because on the second block of trials in the second session, one listener's threshold differed by 10 dB from her average. Second-block estimates, however, were not used

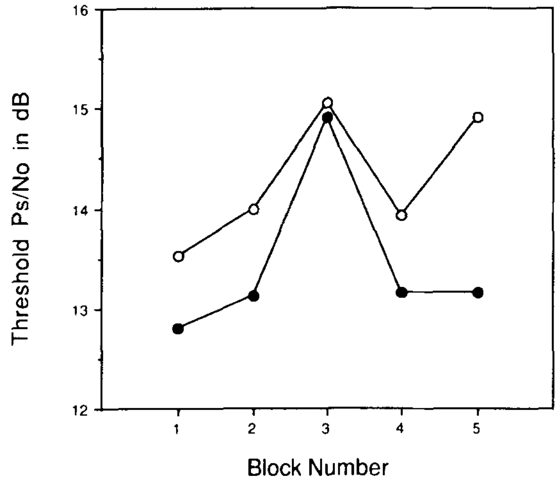

Figure 2. Threshold signal-to-noise ratio (P_S/N_O) in dB as a function of block number for a masked 1-kHz sinusoid when thresholds were measured using the Staircase-71 technique. Open circles are the data from an initial session and closed circles are the data from a second session. The data are from 10 listeners.

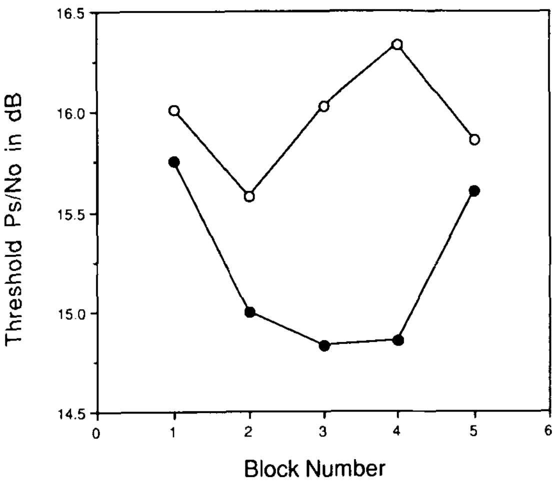

Figure 3. As for Figure 2, except that thresholds were determined using the Staircase-79 technique. Note the change of scale on the ordinate of the figure compared with Figure 2.

to determine the level appropriate to fixed-level testing. On the second session, a drop in the average thresholds is evident with both staircase procedures. The average of the lowest pair of the last three blocks of trials, used to determine the signal level for subsequent fixed-level testing, improved by 1.1 dB for both Staircase-71 and Staircase-79. On the other hand, the average fixed-level block with PEST was presented at a level 0.5 dB higher on the second than on the first occasion.

Thus, if improved performance is defined in terms of lower thresholds, listeners tested with both staircase techniques showed evidence of improvement on the second session. The amount of improvement did not depend upon which point on the psychometric function was being determined. Listeners tested with PEST, on the other hand, showed nearly constant thresholds over both sessions. However, none of the procedures produced statistically significant differences in threshold values between the first and second sessions when the individual thresholds were entered into a split-plot factorial analysis of variance.

Number of Trials

There was no significant reduction in the average number of trials per block between the first and second sessions. The results for the 10 listeners in each group who completed two sessions were: Staircase-71, 55 trials per block in Session 1 and 52 trials per block in Session 2; Staircase-79, 75 trials per block in both Session 1 and Session 2; PEST, 74 trials per block in Session 1 and 79 trials per block in Session 2.

Percent Correct in Fixed-Level Testing

Percent correct in fixed-level testing in Sessions 1 and 2 are shown graphically in Figure 5. The Staircase-71 group achieved comparable scores on both their first and second sessions, although on both occasions, fixed-level thresholds underestimated the capabilities of the listeners. There is no ready explanation for the poorer performance of the Staircase-79 group upon retest. Both staircase groups were presented with signals that were approximately 1 dB less intense on the second occasion; however, only the Staircase-79 group showed a decrease in performance. The better performance of the PEST group in Session 2 over Session 1 suggests that listeners had learned to detect the presence of the signal in the absence of frequent above-threshold reminders. As noted previously, this group received more presentations at threshold level with adaptive testing than did the other two groups, although this experience did not appear to enhance their performance relative to the Staircase-79 group in fixed-level trials following the first session. In considering the performance of this group relative to that of the other two groups, it should be noted that whereas the other

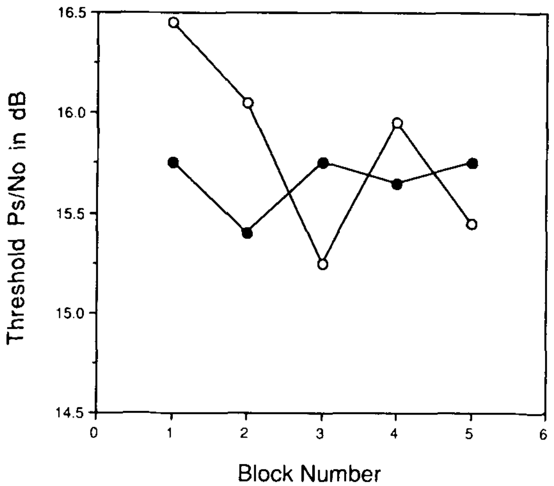

Figure 4. As for Figure 2, except that thresholds were determined using the PEST technique. Note the change of scale on the ordinate of the figure compared with Figure 2.

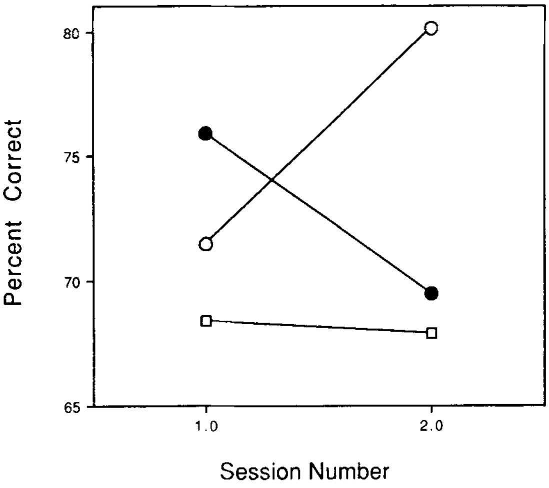

Figure 5. Percent correct in fixed-level trials for Sessions 1 and 2. The signal was presented at the threshold level determined with adaptive testing. The techniques used in adaptive testing were: Staircase-71, actual target 70.7% correct detections (open squares); Staircase-79, actual target 79.4% correct detections (closed circles); and PEST, actual target 80% correct corrections (open circles).

two groups received a less intense signal upon retest, the PEST group was tested with a signal that was on average 0.5 dB more intense on the second occasion. The between-subjects variability in percent correct remained approximately the same between sessions for the Staircase-71 group and was reduced on the second session for the other two groups. The standard deviations in percent correct were: Staircase-71, 6% first session, 7% second session; Staircase-79, 10% first session, 7% second session. The least variable data were obtained from the PEST group on the second session. The standard deviations for this group were: 10% first session, and 3% second session.

SUMMARY

Following careful instruction, inexperienced listeners showed remarkably stable threshold estimates across blocks and sessions with each of three adaptive psychophysical procedures. The within-subject variability of the results compared favorably with data from studies in which experienced listeners participated. Variability was greatest with the staircase technique targeted at 70.7% correct, but this finding must be evaluated against the demand placed upon the subjects, in terms of the number of trials required. This might be important with certain groups of listeners.

Equivalent thresholds were measured by both PEST and Staircase-79. This finding reflects well on the validity of both techniques. The occurrence of occasional lengthy blocks of trials was a drawback with PEST. Nevertheless, with this procedure, naive listeners produced thresholds that were almost identical on each of two sessions.

Fixed-level testing produced different results for the three groups, both within and between sessions. Generally, fixed-level trials underestimated listeners' capabilities. An exception to this generalization is that on retest the listeners in PEST achieved the target level of performance.

These results argue for the validity of threshold measures obtained from unpracticed observers. The stability of the results in this study may have been aided by the careful instructions given to the subjects, and by attention to details affecting their welfare.

REFERENCES

Abramov, I., Hainline, L., Turkel, J., Lemerise, E., Smith, H., Gordon, J., & Petry, S. (1984). Rocket-ship psychophysics: Assessing visual functioning in young children. *Investigative Ophthalmology & Visual Science*, **25**, 1307-1315.

Emerson, P. L. (1984). Observations on a maximum likelihood method of sequential threshold estimation and a simplified approximation. *Perception & Psychophysics*, **36**, 199-203.

Hall, J. L. (1981). Hybrid adaptive procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America*, **55**, 1090-1091.

Hesse, A. (1986). Comparison on several psychophysical procedures with respect to threshold estimates, reproducibility and efficiency. *Acustica*, **59**, 263-273.

Irwin, R. J., Stillman, J. S., & Schade, A. (1986). The width of the auditory filter in children. *Journal of Experimental Child Psychology*, **41**, 429-442.

Kalpan, H. L. (1975). *Some tests of filtered-energy models for auditory detection of a sinusoid in noise*. Unpublished doctoral dissertation, University of Toronto.

Kollmeier, B., Gilkey, R. H., & Sieben, U. K. (1988). Adaptive staircase techniques in psychoacoustics: A comparison of human data and a mathematical model. *Journal of the Acoustical Society of America*, **83**, 1852-1861.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.

Pentland, A. (1980). Maximum likelihood estimation: The best PEST. *Perception & Psychophysics*, **28**, 377-379.

Shelton, B. R., Picardi, M. C., & Green, D. M. (1982). Comparison of three adaptive psychophysical procedures. *Journal of the Acoustical Society of America*, **71**, 1527-1533.

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

Taylor, M. M., Forbes, S. M., & Creelman, C. D. (1983). PEST reduces bias in forced choice psychophysics. *Journal of the Acoustical Society of America*, **74**, 1367-1374.

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

NOTE

1. The expected ratio arises because once the value of the signal has been reduced to the threshold level, the probability of a response sequence leading to an increase in the value of the stimulus is equal to the probability of a response sequence leading to a decrease in the value of the stimulus. Each oscillation about the threshold level with Staircase-71 requires two correct responses and one incorrect response, whereas the required sequence with Staircase-79 is three correct responses and one incorrect response.

(Manuscript received November 21, 1988;
revision accepted for publication April 4, 1989.)