PESTI: A Procedure for Estimating Individual Thresholds in Infant Listeners*

SANDRA E. TREHUB, DALE BULL, BRUCE A. SCHNEIDER, AND BARBARA A. MORRONGIELLO

University of Toronto

An adaptive procedure for rapid estimation of adult thresholds (PEST: Parameter Estimation by Sequential Testing; Taylor & Creelman, 1967) was modified and applied to the estimation of auditory thresholds in 6-month-old infants (PESTI). This procedure yielded similar absolute thresholds to those obtained in previous research with the method of constant stimuli. Advantages of the present procedure are outlined.

In the past decade, there has been considerable growth in our knowledge of infant auditory perception. This has led, on the one hand, to considerations of the nature of auditory processing in infancy (e.g., Jusczyk, 1981; Kuhl, 1983; Werker & Tees, 1984) and, on the other, to systematic efforts to ascertain the limits of their sound detection abilities (e.g., Berg & Smith, 1983; Sinnott, Pisoni, & Aslin, 1983; Schneider, Trehub, & Bull, 1980; Trehub, Schneider, & Endman, 1980). In the latter domain, the major impediment to progress continues to be methodological—namely, how to obtain reliable information from nonverbal subjects who have limited response repertoires and even more limited attentional resources.

Despite the inherent difficulties, several attempts to establish auditory detection thresholds for infant listeners have met with some measure of success for infants 6 months of age and beyond. Typically, head turning toward a sound source has been the target response, and this response has been reinforced by the presentation of an interesting visual event (e.g., the illumination and activation of a mechanical toy). Whereas some investigators have employed unidirectional head turning toward a single sound source and a fixed response interval (Berg & Smith, 1983; Sinnott et al., 1983; Wilson & Thompson, 1984), others have required the infant to turn toward one of two loudspeakers in conjunction with an unlimited response interval (Schneider et al., 1980; Trehub et al., 1980). The use of the two-alternative, forced-choice procedure avoids problems of response bias that could potentially generate underestimates of threshold in the single-response procedure. Moreover, the unlimited response interval

* This research was supported by grants from the Medical Research Council of Canada and the University of Toronto. The authors extend thanks to K.J. Kim for technical assistance.

Correspondence and requests for reprints should be sent to Sandra E. Trehub, Centre for Research in Human Development, University of Toronto, Mississauga, Ontario, Canada L5L 1C6.
permits a flexible accommodation to idiosyncratic patterns of responding (Trehub, Schneider, & Bull, 1981).

The psychophysical method used to determine infant detection thresholds has also varied across investigations. For example, Schneider et al. (1980) and Trehub et al. (1980) used the method of constant stimuli, which involves the repeated presentation of several predetermined intensity levels in random order. Intensity levels are selected to cover the range from very poor to near perfect performance, and the obtained psychometric functions provide the basis for estimating threshold values. The large number of trials required for threshold determination typically precludes the possibility of threshold estimates for individual infants. Thus, it may be necessary to sacrifice the information about psychometric functions provided by the method of constant stimuli (Schneider & Trehub, 1985) in order to obtain threshold estimates for individual infants.

In contrast, Berg and Smith (1983), Sinnott et al. (1983), and Wilson and Thompson (1984) have used adaptive techniques, which means that the selection of intensity levels is adapted to, or determined by, infants' performance over the course of the test session. In particular, they have used up-down or staircase methods as opposed to simple ascending or descending intensity values (i.e., the method of limits). Such staircase procedures have been used extensively in adult psychophysics, especially when it is critical to minimize the number of test trials (Cornsweet, 1962; Levitt, 1970; Taylor & Creelman, 1967). Various staircase procedures differ in their decision rules for when to change intensity levels, what levels to present, when to terminate testing, and how to calculate threshold.

In principle, staircase methods are particularly appropriate for threshold determination in infants, where efficiency or brevity of the test session is of prime importance. In fact, such methods present the only feasible alternative when individual estimates of threshold are required. In practice, however, staircase methods have been used with infants for group threshold estimation (Berg & Smith, 1983; Sinnott et al., 1983), and the selection of parameters has been determined arbitrarily rather than logically or empirically. Thus, while an adaptive procedure offers the lure of obtaining infant threshold estimates in relatively few trials, it is important to evaluate its success with an independent criterion. For example, one could compare thresholds obtained with an adaptive procedure to those obtained with another psychophysical technique, such as the method of constant stimuli.

The purpose of the present investigation was to provide an empirical basis for the selection of test parameters in one adaptive technique and to determine whether the resulting thresholds were comparable to those obtained previously with a nonadaptive psychophysical technique. Of several possible adaptive procedures, we selected Parameter Estimation by Sequential Testing (PEST), the staircase procedure developed by Taylor and Creelman (1967). PEST seemed potentially appropriate for infants because of its reported suc-
cess in concentrating test levels in the region of interest (i.e., threshold) as well as its reasonable tolerance of occasional lapses of attention (Hall, 1981).

PEST specifies the rules of sequencing but leaves open the selection of certain critical parameters such as the starting level, initial step size, maximum permissible step size, and criterion for intensity change. These parameters were manipulated in the present investigation. The result of our adaptation of PEST rules for the estimation of absolute thresholds in 6-month-old infants was PESTI, or PEST for infants.

We used PESTI in conjunction with the response measure that Trehub et al. (1980) had used with the method of constant stimuli. Infants were presented with a test signal located to their left or right, and the signal remained on until the infant turned 45° or more in either direction. Correct responses were reinforced by the presentation of visual stimuli. The use of this response measure offered the possibility of direct comparisons between the group thresholds obtained by Trehub et al. (1980) and those obtained by averaging individual thresholds from the PESTI procedure.

**EXPERIMENT 1**

Basically, PEST searches for the stimulus level that corresponds to any performance criterion, such as 75% correct. It accumulates evidence on correct and incorrect responses at a given intensity level to guide decisions about increasing or decreasing signal intensity in order to move closer to the targeted response probability. The decision history (i.e., continuations and reversals) determines the magnitude of subsequent increments or decrements in intensity level. The final estimate (i.e., threshold) is achieved when the next change in stimulus level would be smaller than some predetermined minimum step size, 2 dB in the present case.

Prior to the application of the PESTI rules, we used a procedure suggested by Creelman (personal communication, 1980) to get quickly into the range of an infant’s threshold, namely QUIR. In the QUIR phase, we used a fixed and relatively large step size of 8 dB. Correct responses on two of three successive trials at a given intensity level resulted in an 8-dB decrease in intensity on subsequent trials. Incorrect responses on two of three successive trials resulted in an 8-dB increase in signal intensity, the termination of QUIR, and the initiation of the PESTI phase. Thus, infants entered PESTI at the last intensity to which they responded reliably in QUIR. In the PESTI phase, we applied Taylor and Creelman’s (1967) rules to determine when to step up or down as well as what step size to use. A sample test session is illustrated in Figure 1. Note that subject M.A. moved quickly through the training and QUIR phases until she made two successive errors at 8 dB (QUIR trial 13), at which time she began the PESTI phase.

PEST looks for an intensity level such that the probability of a correct response at that level is equal to a preselected value. In the present experiment, this preselected value, \( P \), was set at .75. If, for example, the current intensity
level was much higher than the target level, then the proportion of correct responses should be significantly greater than .75. Thus, if we observe that an infant is responding correctly nearly 100% of the time, we know that we should decrease intensity. Over successive trials at the same intensity level, PEST keeps track of the number of correct responses at that level, and compares that number with the expected number of correct responses ($N \times P$, where $N$ is the number of trials at that level). When the absolute value of the difference between the obtained and expected number of correct responses equals or exceeds a criterion value, $W$, then a change in intensity occurs on the next trial. In the present experiment, $W$ equalled .75. For example, the expected number of correct responses after two trials at, say, 40 dB, is $2 \times .75 = 1.5$. If the observed number of correct responses is 1, then the absolute value of the difference between observed and expected number of correct responses is .5, which is less than the deviation limit of $W = .75$, and, consequently, no change in stimulus intensity is made. By contrast, if the listener responds incorrectly on the first trial at a particular level, then the absolute value of the difference between the expected (.75) and the observed (0) number of correct responses equals the deviation criterion of .75. Consequently, the signal level would be stepped up for increased detectability. Thus, subject M.A.'s error on the initial PESTI trial at 16 dB necessitated an increase in signal intensity.

In order to determine the direction and magnitude of each intensity change, we adopted the four sequencing rules established by Taylor and Creelman (1967), limited the maximum change in intensity (step size) to 8 dB, and set the first step size in PESTI at 4 dB. In addition, the first trial following

---

**Figure 1.** A sample PESTI run for subject M.A. (female), 6 months, 7 days of age. Correct and incorrect responses are designated by C and I, respectively.
QUIR and every fifth trial thereafter was a probe trial at suprathreshold intensity (65 dB). The Taylor and Creelman (1967) decision rules are as follows: **Rule 1:** On every directional change in intensity (i.e., reversal), halve the step size. Since M.A. responded incorrectly on the initial PEST1 trial, the difference between obtained and expected number correct (0-.75) equalled $W$, which called for an increase in intensity by the initial step size value, 4 dB. Note that the initial increment in signal intensity when a subject exits from QUIR and enters PEST1 is not counted as a reversal. M.A. then responded correctly on the next three PEST1 trials at 19 dB, making the difference between obtained and expected number of correct responses (3-2.25) equal to $W$, necessitating a decrease in intensity. Since this decrease constituted a reversal, the size was halved from 4 dB to 2 dB. **Rule 2:** The second step in a given direction is the same magnitude as the first. Since M.A. continued to respond correctly at 17 dB (i.e., following the 2-dB decrement), the subsequent decrement would also be 2 dB. **Rule 3:** Whether a third successive step in a given direction is the same as or double the second depends on the sequence of steps leading to the most recent reversal. If the step immediately preceding that reversal results from a doubling of step size, then the third step in the same direction should not be doubled. If, however, the step leading to the most recent reversal is not the result of a doubling of step size, then this third step should be double the second. M.A. continued to respond correctly at 15 dB and required a further decrement in intensity. Since the step leading to her most recent reversal (i.e., the change from 19 dB to 17 dB) did not result from a doubling in step size, then this third step in the same direction was doubled from 2 dB to 4 dB. After 5 trials at 11 dB, the difference between obtained and expected number of responses (3-3.75) was equal to $W$ and led to a reversal and halving of step size. At 13 dB, M.A. responded correctly on one of three trials, generating a deviation of 1.25 and a required change of intensity. Since this constituted a second step in the same direction, the step size remained at 2 dB. After 15 trials at 15 dB, M.A. achieved a deviation between observed (12) and expected (11.25) correct responses of .75 and met the criterion for an intensity change. The next intensity change would be a reversal, requiring a halving of step size to 1 dB, which is less than the specified minimum step size of 2 dB, and called for termination of the session. **Rule 4:** For the fourth step in a given direction, always double the third step size. For example, if successive correct responding resulted in three 2-dB steps downward, then a fourth step down would produce a 4-dB decrement in signal intensity. Under all circumstances, however, the doubling of step size could only occur as long as the maximum step size of 8 dB was not exceeded. It should be noted, as well, that the initial increment in signal intensity that occurred when an infant exited from QUIR and entered PEST1 was not counted as a reversal.

**Method**

**Subjects.** Of 27 infants tested, the data from 9 were excluded because of failure to meet the training criterion described in Procedure ($n = 2$), evidence of a side bias in responding ($n = 1$), failure to complete the test session due to
fussing \((n=1)\), and equipment failure \((n=5)\). The final sample consisted of 18 infants \((10\ males,\ 8\ females)\) 6 months of age \((SD=6\ days)\).

**Stimuli and Apparatus.** The stimulus, a 4,000-Hz, octave-band noise, was produced by filtering the output of a noise generator (General Radio, Model 1381) with a filter (General Radio 1952) set to pass an octave band centered at 4,000 Hz. The filter output was directed to two parallel circuits, one for the right and one for the left speaker. The first element of each circuit was an electronic switch under the control of a microcomputer (Commodore 4032). The output of each electronic switch \((\text{rise/decay time} = 25\ ms)\) was passed through a programmable attenuator \((\text{also controlled by the microcomputer})\) to a preamplifier. The preamplifier output served as left and right input to a stereo amplifier (SAE 2600). Thus, when the electronic switch for one of the two channels was turned on, the octave-band noise, whose intensity was controlled by the programmable attenuator, was presented over the appropriate loudspeaker \((\text{ESS-Heil, Model AMT1AM})\). The rate of falloff in energy on each side of the octave band was 30 dB per octave.

Testing took place in a double-walled sound-attenuating chamber \((\text{Industrial Acoustics})\). During testing, the infant sat on the parent's lap on a chair in one corner of the room. An experimenter sat on a chair in the opposite corner. Loudspeakers were positioned at an angle of 45° to the infant's right and left, approximately 1.85 m from the infant. On top of each loudspeaker was a smoked plexiglas two-chamber box that housed two mechanical toys and lights. These toys were used for reinforcing correct localization responses during testing.

Sound pressure levels were calibrated at the approximate location of the infant's head with a Bruel and Kjaer impulse precision sound level meter \((\text{Model 2204})\) and a .5-inch microphone. All measurements were made with the linear scale. Movement of the microphone around the area occupied by the infant's head produced readings within ±2 dB of the center reading. The background noise level, measured with a 2.54-cm microphone at the approximate position of the infant's head, was 16 dBA.

**Procedure.** Infants were tested during a single session that lasted approximately 40 min. As noted, infants were seated on the parent's lap on a chair in one corner of the testing chamber. An experimenter was seated on a chair in the opposite corner of the chamber, facing the parent and infant. A hand-held box interfaced with the computer allowed the experimenter to call for trials and record responses during a test session. Both the parent and the experimenter wore headphones over which masking noise was presented to prevent them from detecting the locus of the signal. A trial was initiated \((\text{i.e., signal presented})\) only when the infant was centrally positioned between the two loudspeakers, facing the experimenter. The signal remained on until a head turn of 45° or greater occurred. A head turn toward the signal loudspeaker resulted in a toy in the plexiglas box being illuminated and activated for 4 s. Head turns toward the non-signal loudspeaker \((\text{i.e., incorrect responses})\) resulted in a 4-s time-out period during which trials were not presented.
In order to insure that all infants could perform the task, a training criterion was employed with signal intensity well above threshold level. Initially, the signal was presented at 65 dB sound pressure level (SPL). Following correct responding on four consecutive trials, the intensity level was decreased by 10 dB to 55 dB. After four consecutive correct responses at 55 dB, the signal intensity was attenuated by an additional 8 dB to 47 dB, and the infant entered the QUIR phase of testing. During the training phase, the location of the signal alternated on successive trials between left and right loudspeakers.

During the QUIR and PESTI phases, the location of the signal was randomized with the constraints that no more than three successive trials were presented from the same loudspeaker and approximately half of the trials were presented from each of the right and left loudspeakers. During QUIR, all changes in intensity were in 8-dB steps. Correct responding on two of three successive trials at the same intensity resulted in an 8-dB decrease in signal intensity. Infants continued in QUIR until they responded incorrectly on two out of three successive trials at the same intensity level, at which time the signal intensity was increased by 8 dB, and infants entered PESTI.

Recall that the next trial and every fifth trial thereafter was designated as a probe trial to determine if the infant was attending to the signal. On these trials, the signal was presented at the original training intensity of 65 dB. Performance on probe trials had no effect on PESTI decisions. During PESTI, the maximum step size was set at 8 dB SPL and the minimum step size at 2 dB. Infants were considered to have completed testing when one of two conditions was met: (1) the minimum step size of 2 dB was exceeded, or (2) they accumulated 100 PESTI trials. In the latter case, threshold was defined as the next intensity value that would have been presented had a step size of 1 been allowed (i.e., the current intensity level plus or minus 1 dB). In the former case, threshold was defined as the next intensity value to be presented. Thus, in the case of M.A. (Figure 1), the next value that PESTI would have presented after meeting the termination criterion, 14 dB, is her estimated threshold for a 4,000-Hz, octave-band noise. Note that for all stimulus levels below 15 dB, M.A.'s performance was no better than 60% correct; for all levels above 15 dB, her performance was 80% or better. Note, too, that M.A. responded correctly on seven of the nine probe trials. It is likely that incorrect responses on these trials reflected momentary lapses of attention.

Results and Discussion

Excluding training and probe trials, the number of test trials to the PESTI stopping rule ranged from 17 to 79 with a mean of 40. These were distributed between approximately 10 QUIR trials and 30 PESTI trials. Infants responded correctly on 90% of all probe trials, indicating a relatively high level of attention to suprathreshold stimuli.

Individual thresholds are shown in Figure 2. Thresholds were dispersed over a broad range, -4 to 56 dB, with a mean of 29.4 dB SPL (SD = 18.7 dB), and a relatively flat distribution. These thresholds are somewhat higher than
Figure 2. Distribution of threshold estimates for Experiment 1.

the 18-dB threshold reported by Trehub et al. (1980) for 6-month-old infants tested with the same response measure, but with the method of constant stimuli. Note, however, that Trehub et al. (1980) defined threshold as the 65% detection level as opposed to the 75% detection level in the present experiment. When the 75% level is calculated from the psychometric functions of the earlier study (27 dB), the difference between the estimates is essentially eliminated.

The flat distribution and high degree of variability shown in Figure 2 revealed unacceptable precision in our threshold estimates. One possible reason is our choice of $W = .75$. This choice means that if an infant responded incorrectly on the first trial at a particular level, the intensity was immediately increased. Thus, an error or two due to inattention would have led to inappropriate increments or reversals and inaccurate estimates of threshold. It is likely that a more stringent criterion for intensity change would promote reductions in variability and greater precision in threshold estimation. Because an increase in $W$ would lead to an increase in the number of test trials, adjustments of other test parameters would be necessary to maintain the total number of trials at a reasonable level.

**EXPERIMENT 2**

The purpose of this second experiment was to increase the precision of the threshold estimate and to minimize the number of trials per session. Accordingly, three parameters were changed. First, a further QUIR rule was implemented: when a decrease in signal intensity in QUIR resulted in a signal level of 15 dB SPL or less, QUIR was terminated, and PEST1 began on the next trial at the intensity level that would have been used in QUIR (i.e., 8 dB below the previous intensity level.) The additional QUIR rule was expected to improve the efficiency of the procedure in terms of the information gained per trial. It is likely that a signal level of 15 dB or less is approaching threshold and that PEST1, because of its step size rules, would be more efficient than QUIR for converging on threshold once the infant begins to make errors. Thus, subject M.A. would have exited QUIR at 23 dB and would have begun PEST1 at 15 dB.
Second, the maximum step size in PEST1 was reduced from 8 dB to 4 dB SPL. To the extent that QUIR is successful in getting the infant quickly into the threshold region, smaller steps should facilitate convergence on the threshold value.

Third, the deviation criterion was increased from \( W = 0.75 \) to \( W = 1.0 \). Increasing this criterion necessitated the accumulation of more data on which to base a decision to change signal intensity in PEST1. For example, in the PEST1 phase of Experiment 1, M.A.'s error on the first presentation at 16 dB produced an immediate increase in signal intensity on the next trial (i.e., expected-obtained number of correct responses = .75). It can be seen that six PEST1 trials (three at 19 dB, three at 17 dB) with perfect performance intervened before she returned to the 15 dB intensity level. Given that fluctuations in attention might be expected in infants, this rapid reaction of PEST1 to minimal performance data seemed inappropriate. At \( W = 1.0 \), at least two successive errors at the onset of any new intensity level would be required for a change in signal intensity. With this modification, fluctuations in performance would have less direct impact on the course of a testing session.

The only additional change introduced in the present experiment was the rule that an incorrect response on a probe trial resulted in the presentation of an additional probe trial. Failure on two successive probe trials resulted in a brief rest break. If the experimenter resumed testing, then the first trial presented was always a probe trial.

Method

Subjects. Of 33 infants tested, the data from 13 were excluded because of failure to meet the training criterion (n = 7), an incomplete test session due to fussing (n = 4), and equipment failure (n = 2). The final sample consisted of 20 infants (15 males, 5 females) 6 months of age (SD = 10 days). All infants were free of colds on the test date.

Stimuli and Apparatus. The stimuli and apparatus were identical to that described for Experiment 1.

Procedure. The training and testing procedures were the same as those reported for Experiment 1.

Results and Discussion

Only 4 of the 20 infants responded incorrectly on probe trials, resulting in 8 occasions that necessitated a second successive probe trial. Hence, we can conclude that the rate of inattention in the present experiment was reasonably low. Excluding training and probe trials, the mean number of PEST1 trials was 44. This compares with 30 PEST1 trials in Experiment 1. Thus, the additional QUIR rule and the reduction of maximum step size to 4 dB were not sufficient to offset the larger number of trials required by a stricter deviation criterion, \( W \).

Despite the large number of test trials, relatively few infants (approximately 17%) failed to complete the test session. This is a very conservative esti-
mate of subject attrition since there was no attempt to retest such infants. This attrition rate compares favorably with those of other investigators who have used adaptive procedures for estimating infant auditory thresholds (Sinnott et al., 1983: 40% for 7- to 11-month-olds; Berg & Smith, 1983: 54% for 6-month-olds, 36% for 10-month-olds). Because of numerous differences, including response measure and staircase algorithm, it is difficult to pinpoint the factors associated with higher attrition levels. Within reasonable limits, however, the length of a session *per se* may not be the principal factor that determines success or failure in testing. Rather, the nature of the staircase algorithm, with its schedule of reinforced and nonreinforced responses, may make a substantial contribution to the maintenance of infant motivation.

Figure 3 (top panel) shows the distribution of threshold estimates in the present experiment. Approximately 1/2 of the infants had thresholds in the range from 12-19 dB. Thresholds for the remaining infants are scattered over the range from 20-48 dB. Figure 3 (lower three panels) also shows the threshold estimates that would have resulted from terminating the test session after 20,

![Figure 3](image-url)
25, or 30 trials. For example, the panel that is labelled 30 trials or normal exit from PEST1 plots the thresholds for those infants who exited from PEST1 before 30 trials, along with the intensity value at trial 30 for those infants who had not yet satisfied the PEST1 rules. It is clear that, after about 25 PEST1 trials, there is very little change in the threshold estimate. In the interest of efficiency, then, it seems reasonable to terminate testing after about 25-30 trials, since very little change in the estimate occurs after this point. The mean threshold value obtained in this experiment was 25.4 dB (SD = 12.09), which compares favorably with the group threshold obtained by Trehub et al. (1980).

The change from the relatively flat distribution of Experiment 1 to the unimodal distribution of the present experiment, coupled with skewness toward lower threshold values, indicates superiority of the present QUIR and PEST1 parameters over those of Experiment 1. Despite the large number of PEST1 trials required for termination, the present parameters yield fairly consistent estimates of threshold if the session is terminated arbitrarily in as few as 20 PEST1 trials (mean threshold = 25.2, SD = 11.68) or 25 such trials (mean threshold = 25.8, SD = 11.57). Thus, it would seem reasonable to modify the PEST1 termination rule to include a maximum number of trials, say for example 25.

In summary, the present investigation provides a first step in the direction of an empirically determined adaptive procedure for infants. The PEST1 decision rules offer several advantages that favor their use with prelinguistic listeners. First, the subject attrition rate for 6-month-olds is relatively low. Second, our evidence suggests that PEST1 yields threshold estimates that are comparable to those obtained with the method of constant stimuli. Third, the possibility of terminating PEST1 after 20 or 25 trials without sacrificing accuracy generates the efficiency that is essential for clinical application. Finally, it should be noted that PEST1 is, in principle, applicable to any sensory mode and behavioral response measure and can be implemented in a low-tech environment.

REFERENCES


