J Am Acad Audiol 1:138-145 (1990)

A Method for Determining Hearing Sensitivity in Infants: The Interweaving Staircase Procedure (ISP)

Richard S. Bernstein
Judith S. Gravel

Abstract

A computer-assisted adaptive (staircase) threshold acquisition procedure (the ISP) has been developed to determine hearing sensitivity of infants. A visually-reinforced head turn procedure is utilized. Thresholds at three frequencies are obtained with independent and concurrent staircases. Thus, motivation and attention to the task are the same for the three test frequencies. On each trial, the test frequency is randomly selected and then presented at a level determined by the response history at that frequency. The use of Probe and Catch trials permit learning and motivation to be continuously monitored. In the event of marked changes in motivation or attention, trial-by-trial history allows threshold estimates to be adjusted.

Key Words: Infants, pediatric assessment, audiometry, behavioral hearing tests

An assessment of hearing sensitivity must provide the audiologist with information characterizing any existing impairment or confirming the presence of normal auditory function. Minimally, threshold estimates should be obtained for several audiometric frequencies spanning the range important for speech understanding. While achieving this goal during a single test session is not usually problematic with most adults and older children, obtaining comparable audiometric information from infants is more challenging.

The pediatric behavioral assessment armamentarium has improved significantly since the introduction of an operant test procedure that utilizes visual reinforcement of responding. This procedure, now commonly termed Visual Reinforcement Audiometry (VRA), employs conditioning to establish easily observed and reliable responses to sound by infants (for a revie see Wilson and Thompson, 1984).

The simplest type of operant conditionin provides rewards when the desired behavior (e.g a head turn) occurs spontaneously. Over tim the subject forms an association linking the b havior (the operant) to the reward (the reinforc ment). Thereafter, the behavior will be pe formed frequently in order to obtain the rei forcement. In contrast, a more complex operaı procedure like VRA provides reinforcement *on* in the presence of a neutral stimulus (the di criminative cue). With this contingency in effe the infant should learn to listen for the di criminative cue and perform the behavior (i. respond) only when the cue is presented. Wh learning is complete, the neutral stimulus "cc trols" the occurrence of the behavior; the c "predicts" the availability of reinforcement.

Since operants are voluntary behaviors, t success of the conditioning procedure critica depends on the baby's desire (i.e., motivation) obtain the reinforcement. Associations will learned only if the infant is motivated to obta the reward. Moreover, when motivation declin even well established behaviors will cease to

Albert Einstein College of Medicine, Department of Otolaryngology, Clinical Research Center for Communicative Disorders, Rose F. Kennedy Center, Bronx, New York

Reprint requests: Richard S. Bernstein, Albert Einstein College of Medicine, Department of Otolaryngology, Clinical Research Center for Communicative Disorders, Rose F. Kennedy Center, 1410 Pelham Parkway, Bronx, NY 10461

performed. The learned behavior will be exhibited once again, however, if motivation is reinstated (Primus and Thompson, 1985).

In the VRA procedure, infants are taught to turn their head when a test sound (the discriminative cue) is presented. Head turning (the operant) is reinforced by the activation of a mechanical toy. Moore et al (1975) demonstrated that the presentation of an activated mechanical toy could serve as an effective visual reinforcer. When stimulus control has been established, hearing threshold can be determined by systematically varying the level of the discriminative cue (e.g., a pure tone) and regarding the head turn as an indicator that the infant heard the test sound. This hearing assessment procedure has been found to be effective for infants between 5 months (Moore et al, 1977) and 24 months (Thompson et al, 1989) of age.

Our current research requires that we obtain threshold estimates for three test frequencies each time a baby's hearing is assessed. To meet this goal, certain intrinsic limitations of the VRA procedure had to be addressed. The foremost problem is that the infant's motivation to respond declines as the young listener loses interest in the reinforcers (the mechanical toys). Each presentation of the reinforcer reduces its subsequent effectiveness and, therefore, the number of trials that can be obtained during a single test session is limited. Unfortunately, the precise number of trials available to the clinician cannot be predicted: some infants will respond for only a few trials whereas others will participate actively throughout a long test period. Thus, it seemed ill-advised to assess the audibility of one tone and then another (as is traditionally done in VRA) since this systematic decline in motivation would make the tone assessed last appear to have a poorer threshold than the frequency assessed first. Estimates of auditory configuration would be confounded by this order effect. Moreover, the available trials might be exhausted before the last frequency's assessment was completed. Therefore, we needed a procedure that would rationally allocate the available trials among the three frequencies.

The Interweaving Staircase Procedure (ISP) was devised to interweave three adaptive threshold searches in a single test run. The procedure utilizes trial-by-trial randomization of test frequency to ensure that information is obtained at all three frequencies and that motivation does not systematically differ for the three tones. Thresholds are estimated with an efficient staircase procedure that virtually always yields estimates of hearing sensitivity. In addition, since learning and motivation are intimately tied to the validity of the hearing assessment, the ISP was designed not only to determine the extent of stimulus control but also to monitor the infant's motivation throughout the session.

This paper reports on some aspects of the development of the ISP for the clinical assessment of infant hearing. First, the accuracy of the procedure is examined by estimating thresholds from adults using both the ISP and conventional audiometry. Second, thresholds estimated using the ISP are examined in a group of babies whose auditory status is known. Finally, this paper presents a means for examining the validity of each session's threshold estimations and a way to "correct" ISP estimates when the infant's motivation declines during the session.

METHOD

Subjects

Ten adult subjects were recruited from the staff and faculty of the Rose F. Kennedy Center for Research in Mental Retardation and Human Development at the Albert Einstein College of Medicine, New York. The subjects ranged from 20 to 60 years of age (mean = 33.5, SD = 11.6). Subjects were selected without regard to peripheral hearing status.

Additionally, a group of babies ranging in age from 9 to 18 months served as infant subjects. All are graduates of either a Neonatal Intensive Care Unit or Well-Baby nursery. Thus, this infant population is similar to that encountered at many clinical facilities. These babies participate in a large multidisciplinary study of auditory development. Longitudinal electrophysiologic, behavioral, and acoustic reflex findings indicate all members of the cohort have normal cochlear sensitivity. On the test day, both tympanometry and pneumo-otoscopic examination were performed in order to categorize each infant as having either normal or abnormal middle ear function (otitis media).

Instrumentation

A personal computer (Apple IIE) interfaced to a custom-built logic system is situated outside the double-walled, sound-treated test booth. The equipment controls all aspects of the experimental procedure, activates the reinforcers when appropriate, and provides a record of the trial-by-trial performance of each listener. Communication with the computer from within the test booth is accomplished by the activation of hand and foot switches connected to the computer's interface.

A visual reinforcement unit is located in one corner of the test booth. The display is constructed of plywood and stands 5 feet high. Three separate compartments house different mechanical toys which, when appropriate, may be independently activated and illuminated by the computer. The door of each compartment is made of dark-smoked Plexiglas so that the infant may view the activated toys only during periods of reinforcement. The construction of the doors allows the toys to be exchanged easily.

Test signals are generated by three adjustable pure-tone oscillators (Coulbourn S81-06). The output of the selected oscillator is led to an electronic switch (Coulbourn S84-04), programmable attenuator (Coulbourn S85-08), and then to a bandpass filter (Wavetek 852). After suitable gating (25 ms rise/fall times), attenuation, and filtering, the test signal is led to one of the tape input terminals of a two-channel clinical audiometer (Grason-Stadler GSI-16) located within the test chamber. Routing the test signal to the audiometer allows experimental stimuli to be presented via different transducers (earphones, bone oscillator, external speaker) without requiring external impedance-matching networks.

The thresholds reported in this paper were obtained monaurally using a TDH-49P earphone mounted in a MX41/AR cushion on a standard headband. The earphone was acoustically calibrated with a NBS 9-A earphone coupler (Bruel and Kjaer 4152) and soundlevel meter (Bruel and Kjaer 2203). Tone levels are expressed in dB HL (re: ANSI, 1969).

Procedure

The ISP uses three different types of trials: Test, Probe, and Catch (cf., Aslin, 1989) in which the probabilities of occurrence are set at 0.70, 0.15, and 0.15, respectively. Test trials are used for estimation of thresholds and consist of the presentation of four, 500-ms tone bursts (separated by 300 ms) at one of the three test frequencies (500, 2000, or 4000 Hz). Test frequency is selected at random by interrogating a random number algorithm. The computer determines the level at which the selected tonal signal will be presented.

Probe trials are trials in which the signal is presented at a clearly audible level. Probes are used to monitor motivation and also to ensure that the infant listener can receive reinforcement even when the test trials are being presented at

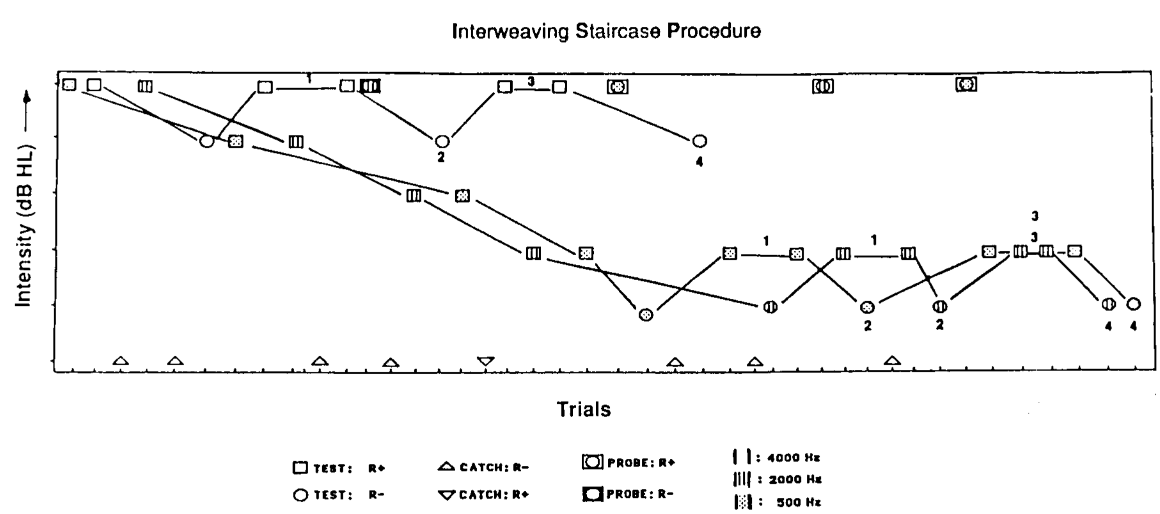

Figure 1 A simulated ISP trial-by-trial sequence of stimuli and responses that might be obtained from a listener with a high-frequency hearing loss. Although stimulus frequency is randomized, the computer tracks performance at each frequency as though it were the only frequency being assessed. Reversals in the direction of presentation level are indicated numerically (1–4) for each frequency. The occurrence of Probe and Catch trials are also depicted. (After Aslin, 1989.)

very low levels. Catch trials are trials in which the tone is omitted, thereby permitting stimulus control to be assessed. Responding on Catch trials provides an indicator of how well the baby learned the task. Probe and Catch trials are interspersed throughout the run. The only constraint imposed on the random selection is that not more than two Catch trials occur in succession.

Thresholds are determined with a variant of the up-down transformed response technique first described by Levitt (1971). Initially, the level near threshold is rapidly determined by reducing intensity by 10 dB following each correct response and increasing level by 10 dB following a miss. After the first miss, a "2–1" decision rule is implemented in order to estimate threshold (cf., Moschetto et al, 1980). Presentation level is based on performance in blocks of two trials: 2 successive correct responses (at that frequency) result in the level being attenuated by 10 dB (descending presentation level), a miss results in the signal level being increased by 10 dB (ascending presentation level). The "2–1" rule tracks the level yielding 71 percent correct. Each threshold is computed after four reversals (in the direction of presentation level) are obtained and is defined as the mean of the presentation levels visited after the first miss. Responses during Catch and Probe trials are not included in threshold computations. The computer terminates the run when all three frequencies have completed staircases. Performance records are maintained separately for each of the three test frequencies.

A simulated ISP run is represented in Figure 1. Depicted is the trial-by-trial sequence of stimuli and responses that might be obtained from a subject with a high-frequency hearing loss. Although the three tones are presented in random order, the computer tracks performance at each frequency as though it were the only stimulus being assessed. The sequence of Catch and Probe trials and the subject's response to each are also displayed.

Adult Procedure

ISP thresholds for adults were obtained in the same manner described below for infants except that verbal instructions replaced the training procedure. Illumination of the reinforcer provided the listener with trial-by-trial performance feedback. Conventional pure-tone audiograms were also obtained by an audiologist either before or after the ISP threshold estimates were obtained.

Infant Procedure

Training

The infant is seated on the caretaker's lap facing the examiner. To begin training, the examiner directs the infant's gaze to the midline by manipulating simple, quiet toys and then presents one of the tones while signaling the computer to activate one of the reinforcers (selected at random). The activation and illumination of the reinforcer usually elicits a head-turn. Gradually, the onset of the reinforcer is delayed (with respect to the onset of the tone) so that the infant learns to turn towards the display before it is available for viewing. Noncontingent (random) head turning is never rewarded. During training, the infant is exposed to all three test frequencies at various stimulus levels to facilitate the necessary stimulus generalization. Training continues until the infant learns to turn only *in anticipation* (i.e., before the presentation) of the reinforcer.

Threshold Acquisition

Prior to the ISP threshold search, the starting levels of each staircase are specified by the examiner. Selection of levels, which are "just clearly audible," maximizes the efficiency of adaptive threshold searches by ensuring that the majority of test trials are presented in the region of threshold. The examiner may use any knowledge of the infant's hearing ability (e.g., performance during shaping or relevant clinical history) to select appropriate starting levels.

The examiner directs the infant's gaze to the midline and then initiates an ISP trial. When a head turn occurs, the examiner indicates the response, and the computer delivers reinforcement if a tone had been presented. Reinforcement is not provided if a response is made to a Catch trial (a false alarm) or if the infant fails to respond to a trial containing a tone (a miss). After each trial, there is a short inter-trial interval.

During the ISP, both caretaker and examiner wear earphones. Masking noise gated on with the onset of every trial serves both to mark the trial interval and to keep the examiner naive as to the trial type presented. The caretaker listens to taped music.

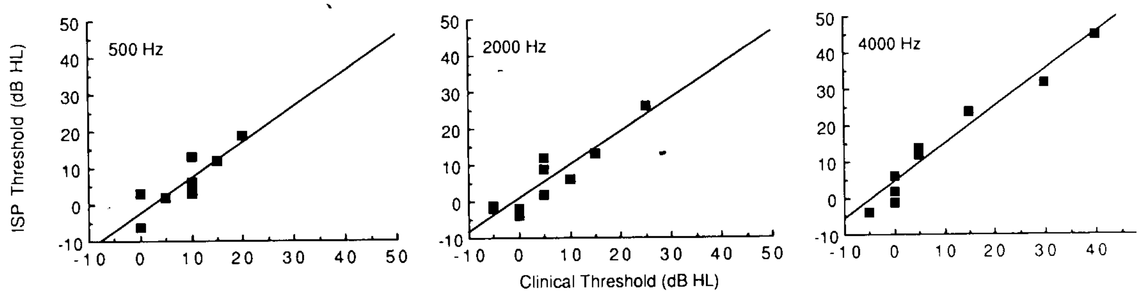

Figure 2 Comparison of clinical and ISP thresholds obtained from 10 adult listeners. Frequency is indicat in each panel. Solid lines are lines of best-fit (least squares).

RESULTS

Threshold Estimation Accuracy

Adults

Figure 2 presents scatterplots of both clinical and ISP thresholds for each frequency (500, 2000, and 4000 Hz) obtained from the 10 adult listeners. The solid lines are the lines of best-fit (least squares criterion). The mean slope of the lines is 0.98 with a mean y-intercept at 0.96. If both procedures had yielded identical thresholds, the lines would have slopes of 1 and y-intercepts at 0. The agreement between theoretical and empirical lines indicate that clinical and ISP estimates are similar. The average differences (ISP minus clinical) were −2.7, 0.4, and 4.8 dB at 500, 2000, and 4000 Hz, respectively. Indeed, the maximum difference between the procedures for any individual was only 9 dB. The ISP yielded sufficiently accurate estimates of thresholds in a reasonable number of trials (mean = 43.3, SD = 8.0).

Infants

Considered next are the threshold estimates obtained from 17 infants (mean age = 12.2 months, SD = 3.2) who completed the ISP. These infants demonstrated that they had learned the task by yielding low false-alarm rates (mean = 0.14, SD = 0.16). Moreover, their high rate of response to Probe trials (mean = 0.95, SD = 0.13) indicates that they were motivated throughout the ISP run. The procedure required on average 38.2 trials (SD = 8.1) and was usually completed in less than 15 minutes.

Mean thresholds from the 7 infants categorized as having normal middle ear function were 28.1, 23.7, and 21.6 dB HL at 500, 2000, and 4000 Hz, respectively. In contrast, 10 infants with otitis media (OM) yielded mean threshol of 40.9, 37.2, and 41.4 dB HL for the same thr frequencies. Plotted in Figure 3 are the mea thresholds for both normal and OM group Error bars indicate 1 standard error of the mea As might be expected, the group of infan with otitis media had poorer thresholds tha the group with normal middle ear functic (F[1,15df] = 10.08; $p < 0.05$). Indeed, tl threshold difference was statistically significa at both 2000 Hz (Scheffé F-test = 7.74; p<0.0 and 4000 Hz (Scheffé F-test = 11.02; p <0.0 Acoustic leakage (earphone-to-head couplin could have resulted in increased threshold vari bility, thereby limiting the power of the statis cal analysis at 500 Hz (see also Wilson ar Thompson, 1984).

The ISP threshold estimates for infan with normal middle ear function reported a somewhat poorer than previous reports (e.

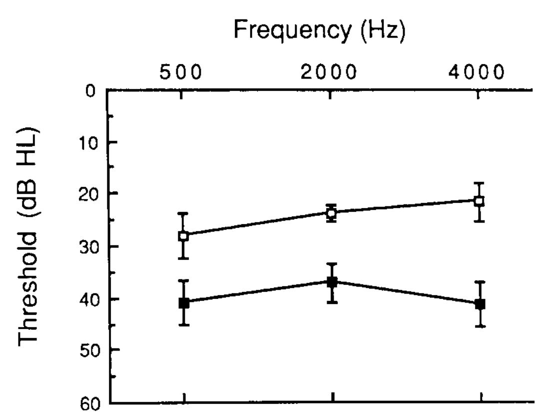

Figure 3 Mean thresholds at 500, 2000, and 4000 obtained from a group of normal hearing infants (op symbols; N=7) and a group of infants with otitis m dia (closed symbols; N=10). Error bars indicate 1 sta dard error of the mean.

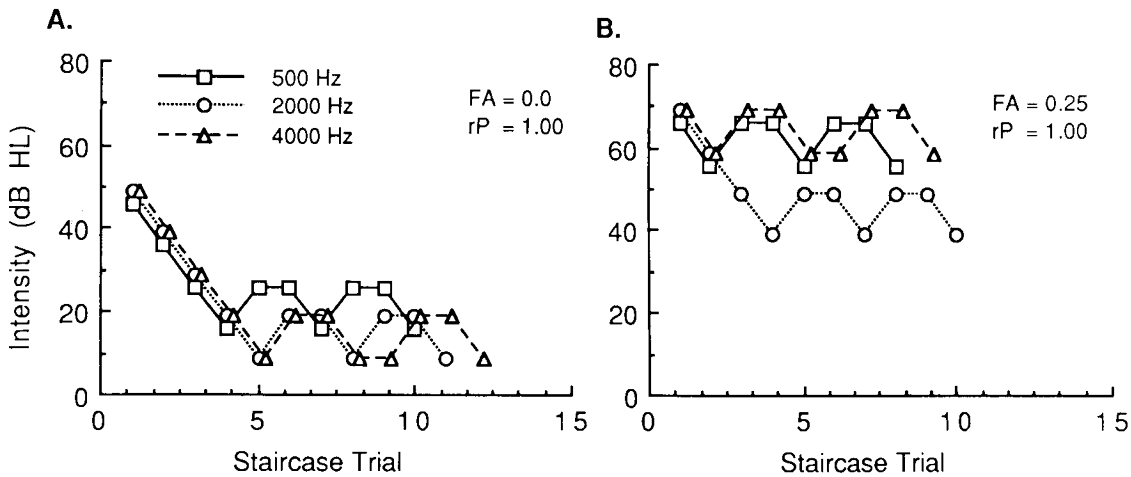

Figure 4 Simplified ISP trial sequences obtained from: *A*, a 9-month-old infant with normal hearing and *B*, a 13-month-old baby with otitis media. Rates of false alarms (FA) and responses on Probe trials (rP) are indicated.

Wilson and Moore, 1978; Nozza and Wilson, 1984). Several factors can account for the differences observed. First, the ISP determines the 71 percent rather than the 50 percent point on the psychometric function; therefore, threshold estimates should be several dB higher (Green and Swets, 1974) than obtained in previous studies. Secondly, methods in each of the laboratories differ in numerous ways such as staircase step-size, test signal properties, etc. Perhaps the most important difference, however, is the time required for training and threshold acquisition. Consistent with the purpose of their study, Wilson and Moore (1978) required, on average, two 15-minute sessions to obtain three threshold estimates. Similarly, Nozza and Wilson (1984) devoted an entire session to training and then utilized two additional sessions to obtain two threshold estimates. In contrast, the ISP was designed to estimate thresholds in the clinical situation in a single test session.

The thresholds obtained from babies with otitis media is consistent with known effects of the pathology. Indeed, the threshold elevations observed are similar to values obtained in infants (Gravel, 1989) and older children (Fria et al, 1985) using more traditional clinical procedures. Therefore, the ISP seems to be capable of detecting mild forms of hearing loss in infants such as that imposed by middle ear pathology.

Figure 4 presents the ISP results of two individual infants whose data were included above: (a) a normal hearing infant and (b) a baby with OM. Their performance is depicted in a simplified ISP report format. This representation was obtained by plotting the infant's response for each frequency as if it were the only one assessed. "Trial number" indicates the order of trials *within* an individual staircase and should be regarded as a relative index of the trial sequence. Rates of false alarms (FA) and responses on Probe trials (rP) are listed in the insets. These infants, regardless of hearing status, were clearly ideal subjects. Their low FAs coupled with their high rPs indicate that they were motivated throughout the session.

DISCUSSION

The results obtained from both infant and adult subjects indicate that the ISP is capable of accurately and efficiently estimating hearing thresholds. These listeners demonstrated low false-alarm rates (mean = 0.12 for adults and 0.14 for infants) in conjunction with high rates of responding on Probe trials (1.0 and 0.95; adults and infants, respectively). These findings suggest that subjects had both "learned" the task and were motivated to respond throughout the session.

Babies, however, must often be tested at times that are less than optimal. As a consequence, learning may be incomplete and/or the baby may not have sufficient motivation to complete the entire session. The ISP permits the extent of learning and motivation to be considered in the interpretation of the test results. As will be discussed below, valuable information on an infant's hearing status can be extracted from the trial-by-trial performance record.

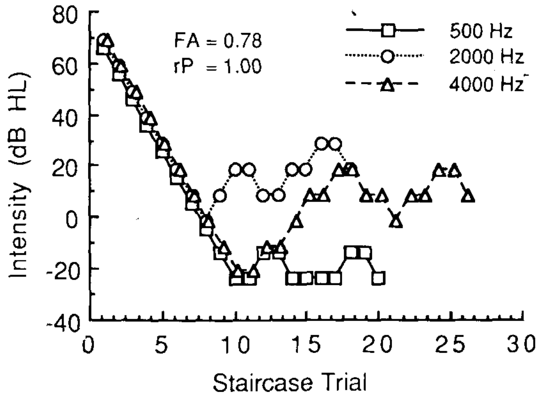

Figure 5 ISP trial sequence of a 13-month-old infant demonstrating a high false-alarm rate. This baby learned to turn towards the reinforcer, but not contingently. As a consequence, at two of the frequencies (500 and 4000 Hz) attenuation reached equipment limits (floor).

Stimulus Control

On occasion, despite seemingly successful training, an infant may respond frequently on Catch trials. While committing some false alarms is expected in any sensory task (cf., Green and Swets, 1974), an excessively high FA suggests that although the infant has learned to make the head turn, the tone is not being used as a discriminative cue. Consider, for example, an infant who attempts to obtain reinforcement by turning away from the examiner toward the display every several seconds. Sometimes, purely by chance, this infant turns while the tone is being presented and will see the activated toy(s). While this schedule of reinforcement would be intermittent, it may be sufficient. This baby is not attending to the presence or absence of the tone. Rather, this infant has learned a different rule governing behavior. The finding of a high false-alarm rate indicates that the threshold estimates are not valid, and the clinical evaluation is compromised. In fact, when a high FA is obtained, threshold estimates do not reflect hearing sensitivity. An example of a baby exhibiting this type of behavior on the ISP is presented in Figure 5.

Motivation

For threshold estimates to be accurate, it is assumed that the infant is motivated to respond throughout the procedure. In the ISP, this assumption is tested by interposing Probe trials within the run. Failure to respond on a Probe trial indicates either: (1) that the infant was momentarily distracted from the task or (2) that the infant was no longer motivated to see the visual reinforcer. Because of the nature of adaptive search techniques, low motivation yields inflated threshold estimates. Although occasional failure to respond to Probe tones is inevitable with very young listeners, a low rP for the session must be considered in the interpretation of the ISP run. When motivation is low throughout the run, the estimates should be regarded as an "upper boundary"; that is, the clinician may regard hearing sensitivity to be *at least as good as* the results indicate.

If motivation wanes only during the latter part of the ISP session, threshold accuracy can be improved by considering the manner in which staircase estimates converge on threshold. For example, observe the staircase tracings depicted in Figure 4. Recall that threshold is defined as the midpoint (mean) between the excursions. The midpoint does not substantially change if one or more reversals are excluded. Fewer reversals decrease the estimation *reliability*, but do *not preclude* use of the threshold. Thus, if inadequately motivated trials can be identified, the accuracy of the ISP can be improved by basing threshold computations *only* on the trials that occurred while motivation was high. Similar

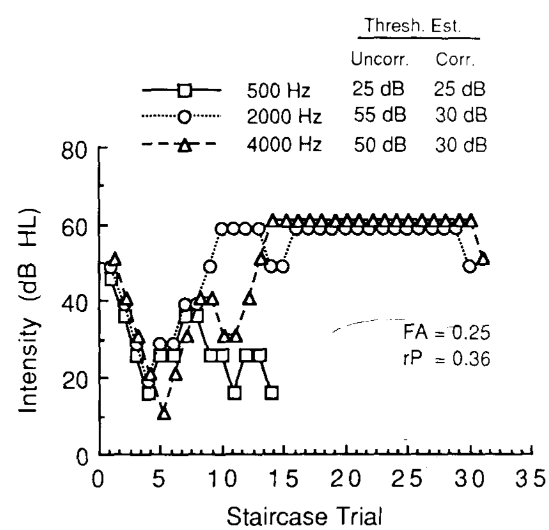

Figure 6 ISP trial sequence from an infant demonstrating declining motivation. Failure to respond during the latter part of the run caused the presentation level to steadily increase to the equipment limits (ceiling). Uncorrected and corrected threshold estimates are indicated for each frequency (see text).

considerations permit thresholds to be estimated when the session must be prematurely terminated.

Figure 6 presents an ISP run in which motivation declined during the latter part of the session. Note that declining motivation yielded a steady increase in presentation levels near the end of the run inflating threshold estimates. However, if threshold is recomputed based on the trials prior to the decline in motivation, threshold estimates change markedly at two of the three frequencies (four reversals were obtained at 500 Hz prior to the motivational decline). Displayed are both the uncorrected and corrected threshold estimates. The corrected thresholds more accurately represent this infant's hearing ability than did the original estimates.

In conclusion, when it is necessary to evaluate the audibility of several different stimuli, the use of the ISP ensures that motivational changes do not confound estimates of auditory configuration. Continuous monitoring of responding on Catch and Probe trials permit the ISP to detect inappropriate rules governing responding and to detect declining motivation. Under certain circumstances, thresholds can be corrected for the effects of motivational deficits. Inspection of the ISP performance record can be an important adjunct to sensitivity estimates, allowing the clinician to formulate a subjective clinical impression based on objective test results.

Acknowledgments. Work supported by NIH-NIDCD Grant #PO1 DC00223. Portions of this paper were presented at the Association for Research in Otolaryngology, St. Petersberg, FL, February, 1989 and at the American Academy of Audiology Meeting, Kiawah Island, SC, April, 1989. Thanks are due to Daphne Straws for assistance in data collection and Robin Cooper for assistance in the preparation of one of the figures. David Stapells and Martha Hildebrandt are acknowledged for their comments on an early draft of the manuscript.

REFERENCES

Aslin R. (1989). Discrimination of frequency transitions by human infants. *J Acoust Soc Am* 86:582–590.

Fria TJ, Cantekin EI, Eichler JA. (1985). Hearing acuity of children with effusion. *Arch Otolaryngol* 111:10–16.

Gravel JS. (1989). Behavioral assessment of auditory function. *Semin Hear* 10:216–228.

Green DM, Swets JA. (1974). *Signal detection theory and psychophysics.* New York: Kreiger.

Levitt H. (1971). Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49:467–477.

Moschetto CF, Bernstein RS, Raab DH. (1980). Noise-intensity discrimination as measured by up-down psychophysical methods. *J Aud Res* 20:113–118.

Moore JM, Thompson G, Thompson M. (1975). Auditory localization of infants as a function of reinforcement conditions. *J Speech Hear Disord* 40:29–34.

Moore JM, Wilson WR, Thompson G. (1977). Visual reinforcement of head-turn responses in infants under 12 months of age. *J Speech Hear Disord* 42:328–334.

Nozza RJ, Wilson WR. (1984). Masked and unmasked puretone thresholds of infants and adults: development of auditory frequency selectivity and sensitivity. *J Speech Hear Res* 27:613–622.

Primus MA, Thompson G. (1985). Response strength of young children in operant audiometry. *J Speech Hear Res* 28:539–547.

Thompson M, Thompson G, Vethivelu S. (1989). A comparison of audiometric test methods for 2-year-old children. *J Speech Hear Disord* 54:174–179.

Wilson WR, Moore JM. (1978). Pure-tone earphone thresholds of infants utilizing visual reinforcement audiometry (VRA). Paper presented at the *American Speech and Hearing Association Convention, San Francisco.*

Wilson WR, Thompson G. (1984). Behavioral audiometry. In: Jerger J, Ed. *Pediatric audiology.* San Diego: College-Hill Press, 1–44.