An Observer-Based Psychoacoustic Procedure for Use With Young Infants

Lynne Werner Olsho,
Elizabeth G. Koch, and
Christopher F. Halpin
University of Virginia

Elizabeth A. Carter
Virginia Commonwealth University

This article describes an observer-based technique for assessing auditory capacities of infants from 3 to 12 months of age. This technique, referred to as the Observer-based Psychoacoustic Procedure (OPP), combines features of the Forced-choice Preferential Looking technique developed by Teller (1979) and of Visual Reinforcement Audiometry (Moore, Thompson, & Thompson, 1975). The rationale behind the procedure and the specific techniques used in its application are detailed here. Psychometric functions and thresholds for pure-tone detection and frequency discrimination obtained with OPP are also presented. The results for 6-month-olds are compared with results from previous studies employing a visually reinforced head-turn procedure.

The number of studies of infant hearing development has increased dramatically in the past 10 years. Infant absolute auditory sensitivity has been examined in several laboratories (e.g., Berg & Smith, 1983; Sinnott, Pisoni, & Aslin, 1983; Trehub, Schneider, & Endman, 1980). Frequency and intensity difference thresholds have been estimated (e.g., Olsho, 1984b; Sinnott & Aslin, 1985). The infant's ability to detect sounds in the presence of competing sounds has also been investigated (e.g., Bull, Schneider, & Trehub, 1981; Nozza & Wilson, 1984; Olsho, 1985). The literature on infant speech discrimination and categorization is substantial (see Aslin, Pisoni, & Jusczyk, 1983, for a review).

Although interest in human auditory development has a long history (e.g., see Eisenberg, 1976), it was not until the development of visual reinforcement audiometry (VRA) by Moore, Thompson, Wilson, and their colleagues (Moore et al., 1975; Moore & Wilson, 1978) that this area started to grow. VRA is essentially an operant conditioning procedure in which the operant response is a head-turn. Head-turns are reinforced by the presentation of an interesting visual event, usually the acti-

vation of a mechanical toy. The acoustic stimulus serves as a discriminative cue: Head-turns in the presence of a specified sound, or set of sounds, are reinforced, whereas head-turns at other times are not. Once the infant has learned the response contingencies, a threshold for the detection of sound can be determined by varying sound intensity to find the level at which the infant stops responding. All of the psychoacoustic studies mentioned above used some variant of the conditioned head-turn technique to estimate infant thresholds. Eilers (1977) modified the technique for studies of speech sound discrimination, and Aslin, Pisoni, Hennessey, and Perey (1981) added an adaptive psychophysical procedure to increase testing efficiency.

The advantage of the conditioned head-turn procedure over those of previous auditory testing techniques is that it allows for multiple comparisons. For example, we can ask if the infant can discriminate among several speech contrasts, or we can vary a single stimulus dimension to determine a psychophysical threshold. Because stimulus parameters such as intensity can be varied for a single subject, thresholds can be determined readily. Other procedures, such as habituation-based paradigms, allow for only a single comparison in a session. Consequently, within-subject thresholds are difficult, if not impossible, to obtain.

The major disadvantage of the conditioned head-turn procedure is that infants do not reliably produce short-latency, conditionable head-turns before they are about 5 months old (Moore & Wilson, 1978). Because the onset of hearing in humans occurs prenatally (e.g., Grimwade, Walker, Bartlett, Gordon, & Wood, 1971), and because, in many respects, auditory function is probably quite mature by the sixth postnatal month, it is clear that many of the interesting events in auditory development are occurring at ages at which we have few tools for analyzing hearing.

Of course, electrophysiological measures such as the auditory brainstem response (ABR) and the cortical evoked potential can be used with younger infants (Hecox & Burkard, 1982; Klein, 1984). Some of these measures are also affected by infant state of arousal, one of the most important variables contributing to the insensitivity of behavioral measures in younger infants (Hecox, 1975). ABR appears to be more resistant to state effects but suffers from other limitations. ABR depends on synchronous activity in a population of nerve fibers at peripheral levels of the auditory system. It is also difficult to obtain frequency-specific information using ABR. Thus, ABR provides a highly restricted view of auditory function. (Hecox and Burkard, 1982, review the characteristics of the ABR in detail.) In any case, whether any electrophysiological response is necessarily a measure of hearing is arguable: Even if a correlation between electrophysiological and behavioral measures can be demonstrated in adult listeners, it is not clear that the same relations will hold in developing organisms.

Our goal was to develop a procedure that employs a behavioral response to sound, under the assumption that behavior is the best measure of "hearing." Although it was clear from systematic studies from other laboratories (e.g., Moore & Wilson, 1978) and from our own casual observations that young infants would not produce reliable conditioned head-turns, we also assumed that there were behaviors emitted by young infants that we could bring under stimulus control. We based this assumption on observations of infants in our laboratory as well as on published reports. For example, Watrous, McConnell, Sitton, and Fleet (1975) reported observing eye widening and brow movement in many infants in the 3–5 month age range in response to various moderate-level noisemakers, although the frequency of head-turns was quite low (about 7% of total responses).

A third influence on our thinking about young infants was the work of Teller and her colleagues (Teller, Morse, Borton, & Regal, 1974), who developed psychophysical measures of infant visual sensitivity. Teller (1979) has described in detail the Forced-choice Preferential Looking (FPL) technique for estimating infant acuity, color vision, and other visual functions. This technique is based on the infant's tendency to attend preferentially to certain visual stimuli, as evidenced by looking toward and continuing to look at the preferred stimulus. In Teller's procedure, an observer watches the infant while the infant looks at two visual stimuli. Rather than just recording gaze direction, however, the observer, who cannot see the two stimuli, must decide whether the specified stimulus (the "signal") is to the infant's right or left. In other words, this observer makes a two-alternative, forced-choice response on each trial, on the basis of the infant's behavior. Clearly, if the observer can reliably judge the location of the stimulus in such a paradigm, then the infant must be providing a reliable behavioral cue to indicate discrimination of the two stimuli.

There are numerous advantages to this procedure over simply quantifying the infant's response directly. First, the response is completely objective: On every trial the observer is either right or wrong about where the specified stimulus was located. Second, the negative effects of response variability can be greatly reduced. The observer can base his or her decision on

any infant behavior, including direction of first look, duration of looking, head and body orientation, or any other cue. If the infant does not give the same response to the same stimulus on every trial or if the infant makes different responses to different stimuli, the observer can still use whatever response the infant emits to make a decision. If we limit ourselves to direct measurement of specific infant responses, it is impossible to record all of the responses that the infant might make, and, consequently, the visual response may appear to be highly variable from trial to trial and from infant to infant. Moreover, the response on which the observer bases a judgment need not be articulated at all. Thus, as Teller (1985) points out, FPL makes full use of whatever sensory cues are available to the observer. Third, because the observer's psychophysical behavior is more readily manipulated than the infant's, a wide variety of psychophysical analytic techniques become available. For example, observers could be asked to vary their criteria for a response systematically or to give confidence ratings of their judgments, so that independent estimates of sensitivity and criterion could be made. It is conceivable that the observer could make magnitude estimates of stimulus brightness on the basis of the infant's behavior. Teller (1979) discusses the logic behind the procedure and its psychophysical properties in some detail.

We hoped that if, in fact, 3-month-old infants were responding to sound with eye opening, brow movements, or other observable activities, we could use a technique similar to FPL to assess auditory sensitivity. We chose to make our new procedure as similar as possible to our original conditioning procedure (e.g., Olsho, 1984b, 1985; Olsho, Schoon, Sakai, Turpin & Sperduto, 1982a, 1982b) so that we could compare thresholds obtained using conditioned head-turn responses with those obtained using an observer-based procedure for infants at 6 months of age. Any differences between the two types of thresholds, then, could be accounted for by the response measure, rather than by differences in the number of signal intervals, signal presentation mode, or other procedural variables. This means that we use exactly the same apparatus and, with few modifications, the same software that we used to condition head-turns. Now, however, instead of judging whether the infant has made a 45° head-turn on a trial, the observer judges whether a signal has been presented on a trial. (Further justification for the specific procedural details are given in the sections that follow.)

We can describe our general testing procedure from the perspective of either the observer or the infant (see Teller's [1979] discussion of "double psychophysics"). For the observer, the procedure amounts to a yes–no task with feedback on each trial. The observer starts a trial when the infant is in a ready condition, and an indicator light signals that a trial is in progress. The observer watches the infant and must decide whether a specified sound, the *signal,* was presented to the infant on that trial. At the conclusion of the trial, the observer is given feedback, that is, informed whether a signal was actually presented.

For the infant, the procedure is almost identical to the conditioned head-turn procedure. The infant sits on the parent's lap while listening to sounds presented monaurally over lightweight earphones. The infant's attention is maintained at midline by

an assistant manipulating toys. At certain times a signal is presented. If the infant responds in such a way that the observer correctly decides that a signal has occurred, a mechanical toy is activated to reinforce whatever response was made by the infant. Thus, the feedback available to the infant is more limited than that available to the observer. The reinforcer provides feedback for positive responses detected by the observer: If the reinforcer comes on, then the response was correct (a *hit*); if not, the response was incorrect (a *false alarm*). No feedback for negative responses or undetected positive responses is available to the infant during testing.

This technique differs from FPL in at least two important respects. The first difference involves the use of reinforcement of the infant's responses. Reinforcement is not used in FPL. Mayer and Dobson (1982) developed an operant variation of FPL, operant preferential looking (OPL), for use with infants over 5 months of age. For 5-month-olds tested with both FPL and OPL, reinforcement did not affect visual acuity measures significantly (Mayer & Dobson, 1982). At the same time, the visual-preference response is quite robust. Infants readily look at a patterned visual stimulus when it is paired with a nonpatterned stimulus (a blank), and this is reflected in the fact that an observer can achieve 100% correct response on some stimuli. The auditory response, however, is not so robust. The difference should not be surprising: The infant must orient toward the visual stimulus in order to see it, whereas most sounds can be heard regardless of orientation of the head and eyes. In a sense, the visual preference response can be thought of as self-reinforcing; orientation toward a sound source need not be. We felt that explicitly reinforcing the auditory response, then, would increase the likelihood and magnitude of a response. Schneider and Trehub (1985) also discuss the idea that the reinforcer serves to make the stimulus "significant" to the infant.

The second difference between FPL and the procedure that we report here is that FPL is a two-interval, forced-choice procedure, whereas the procedure described earlier is a single-interval, yes–no procedure. From a psychophysical perspective, there are distinct advantages to the two-interval situation. For example, suppose that detection thresholds for a tone are to be determined in an adaptive procedure, one in which the intensity of the tone is varied according to whether the subject reported hearing the tone on the last trial. If a single-interval, yes–no procedure is used, it is sensible to use responses only on signal trials to decide whether the intensity should increase or decrease on the next trial. However, if the subject has a tendency to say *yes*, regardless of signal occurrence, the threshold determined in this way will be artifactually improved. In a two-interval task, the subject applies the same tendency to say *yes* to both signal and no-signal intervals. Stimulus intensity on a given trial will be based on whether the subject was correct in choosing the interval containing the signal on previous trials. Thus, a bias toward saying *yes* will not affect threshold. Moreover, because the signal is presented on every trial, every trial provides information about sensitivity, whereas no-signal trials (in the yes–no procedure) provide information about response criterion, not sensitivity. In developmental investigations, the obvious problem with the yes–no procedure is that individuals of different ages may also differ in response bias. Even if all subjects are

trained to some low criterion (e.g., one or fewer positive responses on the last five no-signal trials), there is no guarantee that the subject will maintain this strict criterion over the course of the session, especially if the level of the stimulus is systematically changing over the session. Moreover, infants may well differ from adults in the degree and timing of criterion variation. Thus, age differences in response criterion may be reflected in age differences in threshold as well as in the reliability of the threshold estimate.

Clearly, the two-interval task is an elegant and apparently simple solution to this problem (e.g., Hertzog, 1980; Teller, 1979). Unfortunately, the application of this technique to the study of auditory development is not simple. In FPL, signal and no-signal intervals are spatially separated; the observer makes a *left* or *right* judgment. An obvious extension to audition would be to have sounds presented to the infant's right or to the infant's left and have the observer decide from where the sound was coming on the basis of some directional response made by the infant. In fact, conditioned head-turn studies have used a similar paradigm (e.g., Schneider, Trehub & Bull, 1980; Trehub, Schneider, & Endman, 1980). However, this approach suffers from at least three basic problems. First, because we have not yet been successful in getting 6-month-old infants to make clear directional responses to sounds presented via earphones, this approach requires that the sounds be presented in the free field via a loudspeaker. Thus, age differences in the ability to perform in this situation may reflect differences in binaural processing, such as localization and separation of signal from noise. Frequency effects, furthermore, may be difficult to investigate, because binaural processing abilities are frequency-dependent. Second, there are a variety of psychoacoustic questions that may be difficult or impossible to examine in this context, most notably questions about the development of discrimination between sounds. Different stimuli would have to be presented simultaneously from two speakers, resulting in physical interactions between the sounds in the environment and interactions between the responses to the two sounds in the auditory system. Finally, besides any age differences in auditory processing that might exist, there is a strong potential for differences in the tendency to make directional responses across age during early infancy (e.g., Clifton, Morongiello, Kulig, & Dowd, 1981).

Another approach is to employ a two-interval, forced-choice procedure in which two successive observation periods occur on each trial. The observer would then judge the interval during which the signal was presented. However, if the signal occurred during the first interval and if the infant responded to it, the reinforcer could not be activated until the second interval had ended. This delay might prevent the infant from learning the contingencies between sound, response, and reinforcement. For reasons already discussed, reinforcement seems to be very important to the success of the procedure.

Our solution to the bias problem was to train observer–infant pairs in the single-interval task to a relatively low rate of incorrect positive responses (i.e., false alarm rate) and monitor the rate throughout testing, returning to the training procedure or stopping the procedure should the rate rise above criterion. This

approach has also been employed in earlier conditioned head-turn studies of 6-month-olds (e.g., Olsho, 1985; Sinnott & Aslin, 1985; Sinnott, Pisoni, & Aslin, 1983).

In the remainder of this article, we describe details of the procedure that we refer to as the Observer-based Psychoacoustic Procedure (OPP). We present data on pure-tone thresholds and on frequency-discrimination thresholds for 6-month-olds and compare these with thresholds obtained in our laboratory and with those published by other investigators using the conditioned head-turn procedure with infants of this age. In addition, we present psychometric function data obtained for 3-, 6-, and 12-month-olds for pure-tone detection and frequency discrimination using OPP to show that the technique yields reliable and reasonable data over a relatively large age range.

Procedural Details

General Procedure

The laboratory set-up for OPP is very similar to that generally used for conditioned head-turning. The parent sits holding the infant facing a window in the test booth and a video camera. A table is positioned in front of the parent and infant. An assistant, the "toy waver" in our laboratory, is also seated at the table at a right angle to the left of the parent and infant. The mechanical-toy visual reinforcer, enclosed with lights in a smoked plexiglass box, is located at a right angle to their right. The observer sits in the adjacent control room, watches the infant through the window or on a video monitor (or both), and makes judgments on a trial-by-trial basis.

Infant placement. In the majority of sessions we have the infant sit on the parent's lap. For younger infants, the parent may need to support the infant in an upright, centered position, although we ask the parents not to try to restrain the infant. Some infants stand on the parent's lap; a few sit or stand on the table. The only constraints on the infant's position are the length of the earphone cord, the infant's position relative to the visual reinforcer, and the observer's view of the infant. In some situations we allow the infant to have a pacifier during the session, because trials are long enough so that sounds would not be completely masked by the intermittent sounds of the infant's sucking. We take a break if the infant wants to take a bottle. The general rule is to impose as few restrictions on the infant as possible.

In accord with that rule, we use very small, lightweight earphones with the infants.[1] The earphone is covered by a small foam cushion and is placed in the concha at the opening of the ear canal. Few infants object to this earphone: We have had no rejections from 3- or 6-month-olds and only one rejection among 12-month-olds. Because it is quite easy to displace this earphone, it is secured to the infant's ear with micropore tape, and the cord is taped to the back of the infant's shirt. Secure placement of the earphone is critical: Substantial reduction in the level of sound delivered to the infant can result if the position of the earphone changes. With this procedure the earphone can be left in place during breaks, unplugged and still taped to the infant's shirt.

Toy waving. The toy waver stays with the parent and infant throughout the session and fulfills several roles. This assistant's primary job is to engage the infant with toys that are manipulated on the table top. It is not strictly necessary to maintain the infant's attention at midline because we are not necessarily conditioning head-turns. However, having a relatively restricted range of "baseline" behavior helps the observer to improve the signal-to-noise ratio in making judgments. Moreover, because many infants respond to sounds with head-turns toward the visual reinforcer, the toy waver, or the parent, having the head centered at the beginning of a trial does help to some extent. In order to keep some infants from lunging across the table to grab the toys, we sometimes give them a simple toy (e.g., a plastic ring or bead) while the toy waver maintains their visual attention with a more complex toy on the table. As investigators who have used conditioned head-turns have often noted, there are many tricks that the toy waver must use to keep infants engaged with the toys, but not so engrossed with them that they fail to attend to the sound or the visual reinforcer.

Another of the toy waver's jobs is simply to keep the infant from getting bored with the whole procedure. Few infants would tolerate 20 min in this position on the parent's lap—even with sounds and an intermittent mechanical toy—without having something to keep them occupied between trials. In addition, the toy waver often provides extra reinforcement when the infant is doing well in the procedure (e.g., by saying "Good girl!" when the infant seems to be getting the reinforcer to come on fairly consistently). As in the head-turn procedure, if an infant doesn't turn toward the reinforcer during the training procedure, the toy waver will direct the infant's attention toward it when it comes on.

Finally, the toy waver often helps the observer in trying to "read" a particular infant. Obviously, both the parent and the toy waver have to be kept "deaf" to the presentation of sounds to the infant.[2] At the same time, when the visual reinforcer is activated, the toy waver knows that the infant has just been pre-

---

[1] We have calibrated two earphones of this type, the Toshiba RM-3 and the Sony MDR-E242. In both cases, we found that the response of the earphone is relatively flat (within about 4 dB) over the range, from 500 to 4000 Hz, at which we do most of our testing. There is a peak in the earphone's response around 6000 Hz, on the order of 10 dB above the levels produced at higher and lower frequencies. At frequencies above 500 Hz harmonic distortion is not a problem. The amplitude of the first harmonic is at least 40 dB lower than that of the signal. Because of the ease of using these earphones with infants, they would seem an appropriate choice for many applications.

[2] When first using earphones with infants, we had assistants, who were sitting just next to but facing away from the infant, try to tell *signal* from *no signal* trials on cue from the control room. Because the assistants' performance did not exceed chance, we felt confident that the adults in the booth with the infant would not be able to hear the sounds presented to the infant as long as the earphone was seated properly. Because the adults are not masked during testing, they can hear sounds from the infant's earphone if the placement changes and adjust the infant's earphone as needed. Communication among the adults is also less cumbersome. Recently, we have found that with the small earphones, for some stimuli, the sounds are audible to the adults. For these stimuli we have the parent and toy waver wear head sets to keep them from hearing the stimulus. Although this makes things a little more awkward, it is not unmanageable.

sented with a sound, and may notice some behavior that he or she believes to represent a response to sound. Because the observer stays in voice contact with the toy waver, the toy waver can often suggest that the observer watch for that behavior on future trials.

The observer: Training and testing. The observer watches the infant through the booth window or on the video monitor (or both—the choice is left to the individual observer). Although the camera may give a close view of the infant's face, all three occupants of the booth can be seen through the window, and it is sometimes helpful to be able to "partial out" the effects of adult activity on the infant's behavior. Most observers seem to prefer to watch the infant through the window.

The observer begins a trial when the infant is in a ready state: quiet, attending to the toy on the table, usually looking up, facing toward midline. Starting trials at a consistent point in the course of the infant's behavioral variation helps to improve the observer's ability to identify signal trials. Sometimes the observer needs to request that the toy waver engage the infant in a specific way to achieve this. The observer also has to be careful not to start trials when one of the adults has just changed activities. Generally the observer will wait for things to settle down to a fairly stable rhythm to begin. Intertrial intervals typically range from about 5 to 40 s and are usually around 15 s.

When the observer starts a trial, either a *signal trial* or a *no-signal trial* occurs. The definitions of *signal* and *no-signal* depend on the task at hand. For pure-tone thresholds, a signal is defined as sound occurrence, and a no-signal as no sound. For frequency discrimination, a signal trial is one on which the frequency of the tone burst is changing, and a no-signal trial is defined as one on which no frequency change occurs. Signal trials occur on approximately 65% of trials. In either case, an LED flashes in the control booth for the duration of the trial to let the observer know that a trial is in progress. The observer must depress a button during the trial period to indicate a judgment that a signal occurred on that trial. If the observer thinks that a no-signal trial occurred, no response is made. If the observer correctly judges that a signal has occurred, the visual reinforcer is activated in the test booth as soon as the observer records the judgment. A trial-by-trial display on the computer monitor gives the observer feedback about what actually happened at the end of each trial.

Since beginning to test younger infants using OPP, we have increased the length of a trial from the 4–6 s we used for conditioned head-turns to 10 s. We have not systematically investigated observer response latency. However, 3-month-olds do seem to take longer to respond to the sounds than do 6-month-olds, and observers seem to take 1–2 s to decide whether some infant behavior actually represents a response to sound.

One important aspect of OPP is the training of the observers, and here two questions should be addressed: (a) How long does it take an observer to be able to score infants reliably? and (b) How long does it typically take for an observer to learn what a particular infant does in response to sound? (A related issue concerns the number of infants for whom we are unable to reliably estimate a threshold in a reasonable number of attempts.)

As far as the initial training of observers is concerned, we find that even people with little experience with infants can learn to observe in OPP within about 1 month. We train observers in a detection paradigm because most infants figure out the contingencies in this task fairly quickly. An observer-in-training moves from taking the role of toy waver, to watching a trained observer, and, finally, to observing under supervision. An observer is considered trained when his or her obtained thresholds are comparable with those of other observers, and when he or she responds *yes* on no more than 25% of no–signal trials on a regular basis.

The most common types of infant responses that observers report in making judgments are not surprising. They include eye movements, either toward the visual reinforcer or away from the toys manipulated by the toy waver; eye widening and other changes in facial expression; head turns, toward the reinforcer or, often, toward the parent or toy waver; and changes in activity level, either increases or decreases. Occasionally, infants will place a hand on the earphone during signal trials. A few infants develop what an adult might call superstitious behaviors: One infant consistently kicked her right leg on signal trials, and another would pat the table top. It is often the case, of course, that the observer cannot define the response(s) that are the basis for judgments.

Although it is helpful to point out to the trainees behaviors that they may not have noticed, the most important instruction that they should receive is to adopt a fairly conservative response bias. The most common training problem is a high false alarm rate. We are, in effect, instructing observers to adopt the goal of maximizing the hit rate for a fixed false alarm rate (i.e., the Neyman-Pearson decision rule). Although this rule is consistent with the goal of getting threshold as close to the infant's "true" sensitivity as possible, it is apparently not one that observers readily adopt.

Currently, we have trained seven individuals as observers. Individuals with prior experience with infants trained somewhat faster than others, but all seven were considered trained within 1 month. Thus, we would argue that the training procedure for observers need not be extensive or arduous.

In regard to the second training question, that of the time required for an observer to be trained on a specific infant, we find that for 6-month-olds, the number of trials required to reach training criterion is about the same as in the head-turn procedure. For example, in a frequency-discrimination task using conditioned head-turns, 6-month-olds take from 8 to 35 trials to reach a criterion of 80% correct (i.e., at least 4 of the last 5 signal trials correct and at least 4 of the last 5 no-signal trials correct). The average number of training trials in a session during which a threshold is obtained is about 18. In OPP the average number of training trials in a session during which a threshold is obtained is 22 for 3-month-olds, 22 for 6-month-olds, and 18 for 12-month-olds in a pure-tone detection task. For frequency discrimination, the averages are 19 training trials for 3-month-olds, 17 for 6-month-olds, and 24 for 12-month-olds.

We have also found that there are very few infants for whom data cannot be obtained using OPP. It is relatively common not to obtain a threshold for an infant during the first session; thus, it typically requires one session per threshold needed, and one or two additional sessions to complete testing for an infant. Approximately 40% of 3-month-olds and 20% of 6-month-olds do not reach training criterion during the first session, although

these rates are declining as we continue to use the technique. In our experience with conditioning head-turns, the corresponding rate is only about 15% for 6-month-olds. In a sample of thirty-four 3-month-olds and forty-seven 6-month-olds in a recent study, only three 3-month-olds failed to reach the training criteria within three sessions. The rate for conditioned head-turns typically runs from 5–10%, depending on the study. Subject attrition for OPP in our lab results almost exclusively from failure to get the parent to return to complete testing, and, in fact, we lose very few subjects at all.

In summary, we feel that OPP is a procedure that does not require special or extensive training. Although experience with infants may be helpful in learning to judge infant responses to acoustic stimuli, it is far from essential. Nearly all infants provide some sort of response that the observer can reliably use to establish auditory sensitivity.

Psychophysical Method

OPP is potentially applicable to a variety of psychoacoustic questions, using a variety of psychoacoustic methods. However, we have found that certain types of procedures tend to improve the sensitivity of the measure.

For example, a common procedure in infant auditory research is to have the infant discriminate a change from a repeating background. In frequency discrimination studies, for example, the infant hears tone bursts repeating at the standard frequency throughout the session; during a trial, the frequency of the tone burst changes to some comparison frequency, and response to the comparison frequency is reinforced. We were rather unsuccessful in estimating frequency-difference thresholds using this procedure in conjunction with OPP, although it worked quite well with 6-month-olds in the conditioned head-turn paradigm.

A discrete trials technique, on the other hand, worked very well. Here, rather than having the infant listen to a repeating background, the infant hears nothing between trials. When a trial occurs, for example in frequency discrimination, the frequency of the tone burst either changes on alternate bursts (signal trial) or remains constant on all bursts (no-signal trial). In a sense, this approach reduces some of the uncertainty for the listener. Because all sounds are clearly audible, at least in the discrimination experiment, the infant knows when to listen. This may reduce the signal-to-noise ratio for the observer and, consequently, result in lower threshold estimates.

We have used both adaptive techniques and the method of constant stimuli to obtain thresholds from infants using OPP. Adaptive techniques are those in which the difficulty of the task on a particular trial is determined by the infant's performance on previous trials. In the method of constant stimuli, stimulus parameters on a trial are determined randomly; a fixed set of stimuli is presented in different random orders to all infants. We have used only adaptive procedures for testing absolute sensitivity.

In frequency discrimination, where we have used both techniques to obtain frequency difference limens, the method of constant stimuli yielded better performance than an adaptive technique with OPP. Moreover, we have found that some adaptive techniques work better than others. A one-up, two-down technique yielded very poor thresholds in OPP. In that procedure, the frequency difference on the next trial increased if the infant missed the last frequency change, but decreased if the infant responded to the last two frequency changes correctly. On the other hand, a variation of the parameter estimation by sequential testing (PEST) rules for determining the frequency difference (Hall, 1981; Taylor & Creelman, 1967) gave thresholds for 6-month-olds in the same range as those we obtained in the conditioned head-turn procedure for 6-month-olds. This procedure uses more information than the last two trials to determine stimulus values on a given trial and tends to be more resistant to lapses of attention (Hall, 1981; Shelton, Picardi, & Green, 1982) on the part of the listener/observer.

Occasionally during testing, an observer's false alarm rate rises above what we consider a reasonable level. The criterion that we currently apply is 0.25. The program that we use to monitor the observer's responses and to control stimulus presentation estimates a "local" false alarm rate (i.e., it calculates the false alarm rate over the last four no-signal trials). If the rate rises above 0.25, testing is interrupted, and the observer gets a message indicating the current false alarm rate. If the false alarm rate has not dropped back to an acceptable level within eight more no-signal trials, the observer must either stop the session or retrain. If the observer decides (on the basis of the infant's state, etc.) that the session should continue, then the stimulus parameters on the next trial return to the level originally used in training, and the observer is required to achieve an 80% correct rate on both signal and no-signal trials before testing will continue. Once the observer has retrained, testing resumes at the point at which it was interrupted. If for some reason the false alarm rate for the whole session exceeds 0.25, even after retraining, then we do not use that infant's data. The percentage of sessions so excluded is about 16%, varying somewhat with age and type of test. In order to achieve a fairly reliable estimate of the false alarm rate, about 35% of all trials during testing are no-signal trials. This signal probability represents a compromise between the symmetric forced-choice (50% signal probability) case and the need to maximize the number of signal trials. Both the number of signal trials and the probability of signal occurrence would be expected to affect the reliability of the performance measure (McNicol, 1972; Ogilvie & Creelman, 1968).

In order to illustrate how we have used OPP in specific psychoacoustic experiments with infants, we will now describe the results obtained in pure-tone detection and in frequency discrimination.

Evaluation of the Procedure

Because nearly all of the data collected in our laboratory have dealt with frequency discrimination or with pure-tone sensitivity among 6-month-old listeners, we chose to develop OPP using these psychoacoustic tasks. Our goals have been twofold. First, we compared measures of sensitivity obtained using OPP with those obtained using conditioned head-turns. Next, we applied the technique to 3- to 12-month-old infants to determine

whether reasonable psychometric functions could be obtained from them.

Pure-Tone Sensitivity

Conceptually, the simplest application of OPP is the case in which the observer's task is to distinguish between trials on which the infant is presented a sound and trials on which no sound is presented. By varying stimulus intensity over trials, a psychometric function (i.e., the percent correct *yes* responses [i.e., hits] as a function of intensity) can be generated. In addition, threshold for detection of a pure tone can be estimated. We have used the technique as described above to obtain detection thresholds for 3-, 6-, and 12-month-olds, at frequencies ranging from 250 to 4000 Hz.

The first issue was whether OPP could give us thresholds for pure tones at least as good as the ones that we obtained from 6-month-olds using conditioned head-turns in a previous study (Olsho, 1985). We compared these with thresholds for 6-month-olds obtained using OPP. In both methods an adaptive technique (stimulus level varied from trial to trial depending on the listener's performance on previous trials) was used to estimate threshold. The conditioned head-turn thresholds were estimated using a one-up, two-down algorithm (i.e., stimulus level increased if the infant failed to respond on a signal trial but decreased if the infant responded on two consecutive signal trials). The OPP thresholds used a modification of the hybrid adaptive method described by Hall (1981)[3], but the rules were set up to estimate the same point on the psychometric function (0.707 hits) in both cases. The session continued until 50 signal trials had been presented, until the infant became too fussy or sleepy to continue, or until the false alarm rate exceeded 0.25, whichever came first. To be used, the session had to include at least 4 signal levels and at least one reversal in the direction of the staircase. About 30 signal trials were available, on average, to estimate each OPP threshold. If an infant's psychometric function was flat or nonmonotonic, it was not used.

The average thresholds and standard errors obtained by Olsho (1985) at 500, 1000, and 4000 Hz are shown in Figure 1; six infants are included at each frequency. Also in Figure 1 are average thresholds and standard errors for thirty-seven 6-month-olds using OPP: 12 infants at 500 Hz, 13 at 1000 Hz, and 12 at 4000 Hz. Thresholds estimated by the two methods are quite similar at all three frequencies.

Because we had relatively few data on pure-tone sensitivity using conditioned head-turns, we also compared our thresholds to those obtained in other laboratories for infants at this age. It is difficult to make detailed comparisons because studies differ not only in the specifics of the threshold estimation procedure, but also in the mode of stimulus presentation. Figure 2 summarizes data from three other studies in which similar testing procedures were used.[4] In order to take into account general differences among laboratories, we have replotted estimated thresholds for infants in each study relative to those of adults tested in the same study. In each case a yes–no procedure was used. All of the studies used an adaptive procedure to estimate thresholds. The adult thresholds from our laboratory are from an earlier (Olsho, 1985) study. The infants and adults were
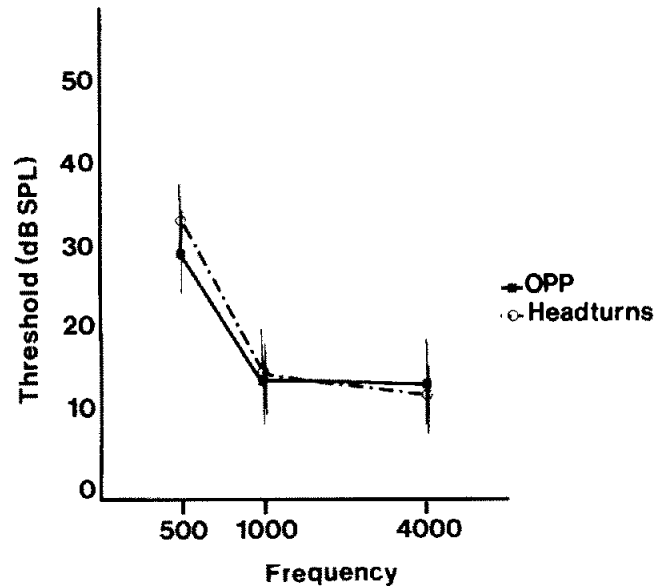

Figure 1. Pure-tone detection thresholds for 6-month-old infants as a function of frequency, obtained using two different techniques. (*Headturn* refers to the conditioned head-turn technique, *OPP* refers to the method described in this article. Error bars represent ±1 standard error.)

tested under similar conditions; the only difference was that the adults received verbal instructions to respond when they heard a sound. OPP appears to produce thresholds at 4000 Hz very similar to those reported in Nozza and Wilson's (1984) study. At 1000 Hz, OPP thresholds are about 6 dB lower than Nozza and Wilson reported. At 500 Hz, OPP thresholds are within 2 dB of those reported by Berg and Smith (1983). Given the differences in procedure among these three studies, then, the results seem to be in good agreement. The two latter studies, however, report thresholds 10–20 dB lower than those of Sinnott et al. (1983). Thus, the major difference seems to be between the Nozza and Wilson (1984), Berg and Smith (1983), and OPP studies, on the one hand, and Sinnott et al. (1983) on the other.

---

[3] Hall's (1981) method uses the PEST rules developed by Taylor and Creelman (1967) to determine signal level during testing but makes a maximum likelihood estimate of threshold at the end of a run (Hall, 1968). The hybrid has the advantages of both techniques: It brings signal level close to threshold within relatively few trials and, because the threshold estimate is based on all trials, it is relatively unaffected by momentary lapses of attention on the infant's part.

[4] We have omitted the data of Trehub, Schneider, and Endman (1980) from this comparison, because they used a two-alternative, forced-choice procedure to estimate thresholds for octave-bands of noise. In that study, infants were required to turn toward the location of a sound to receive visual reinforcement. Approximate thresholds for 6-month-olds relative to the average thresholds of two adults tested by Trehub et al. were 20 dB at 400 Hz, 23 dB at 1000 Hz, and 15 dB at 4000 Hz. We have not included data from Berg and Smith's (1983) study in which sounds were presented free field because no data were collected from 6-month-olds in these conditions.
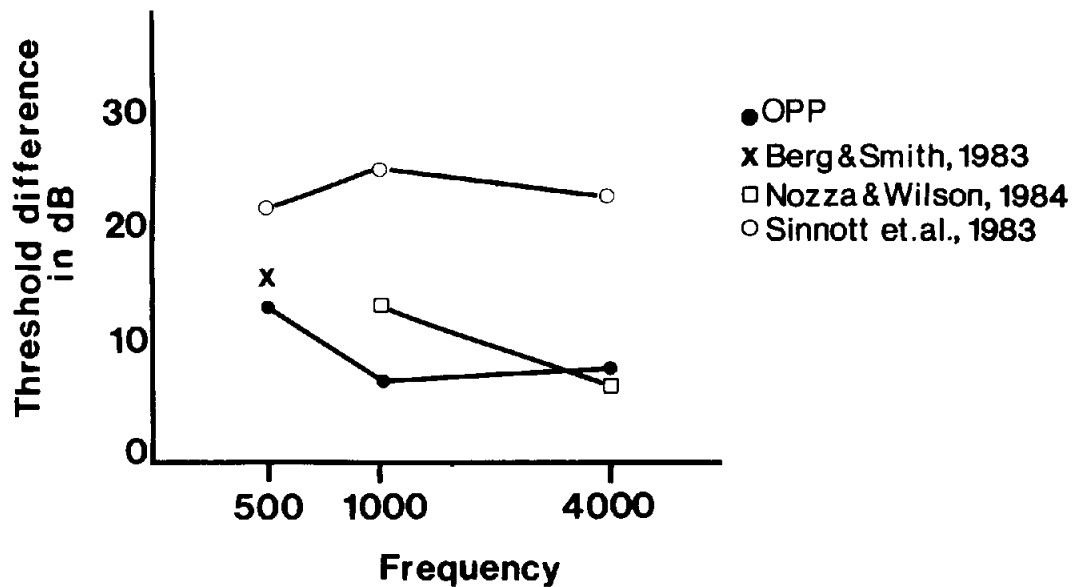
Figure 2. A comparison of pure-tone detection of 6-month-olds from four different studies. (Each point represents the difference between mean adult threshold and mean infant threshold obtained in that study at that frequency.)

Although it is difficult to make any firm statement about the sources of this difference, two rather obvious possibilities should be considered. First, in terms of mode of stimulus presentation, Sinnott et al. (1983) presented tones over loudspeakers, whereas the other three studies used headphones. Second, in terms of the average false alarm rate of subjects (see detailed discussion of false alarm rate in OPP, below), Sinnott et al. (1983) reported that, as a group, the 6-month-olds responded on fewer than 10% of the catch trials, and they excluded from analysis any session during which the infant responded on more than one of four to six catch trials. Berg and Smith (1983) reported an average false alarm rate of 16.9%, and Nozza and Wilson (1984) reported 11.5%. For the OPP infants whose thresholds are reported here, false alarm rate was 13.7%. It is impossible to say exactly how much of the difference between Sinnott et al. and the three other studies can be accounted for by these variables, but higher false alarm rates would be expected to lead to better infant thresholds.

We concluded from this comparison that OPP yields thresholds for 6-month-olds no different from those that we obtained from 6-month-olds in our laboratory using conditioned head-turn responses. Moreover, it appears that the technique yields thresholds similar to those reported in other studies using head-turns, especially in cases where variables such as false alarm rate and stimulus presentation are similar.

Psychometric functions. It is of obvious importance to demonstrate that OPP will yield reasonable data from 3-month-olds, because our primary purpose in developing the method was to be able to test younger infants. We could not compare 3-month-olds' threshold estimates from OPP with some independent measure of sensitivity. Therefore, the shape of the psychometric functions obtained from infants was the major criterion used to determine whether the method was satisfactory for

younger infants. We also examined false alarm rates as a function of age (that analysis is described in the following section).

Eighteen infants in each of three groups were tested: 3–5 months, 6–9 months, and 10–12 months of age. Performance was examined at three tone frequencies: 250 Hz, 1000 Hz, and 4000 Hz. Six infants listened at each frequency at each age. Because performance to the highest sound pressure level at 1000 Hz in the 3- to 5-month-old group did not reach 80%, we ran six additional infants at this age using a higher range of sound pressure levels. Both psychometric functions for the youngest infants at this frequency are included here. In all other respects, the procedure for testing was identical to that described above for threshold determination for 6-month-olds.

Because we had no a priori way to determine what levels should be used to obtain psychometric functions for 3-month-olds, we used an adaptive procedure, the hybrid technique mentioned earlier, to determine signal level on each trial. However, we made the starting level during testing relatively high: 40 dB over the thresholds for 6-month-olds that we had obtained previously. We made the initial step size 10 dB. The net effect was to spread observations over a relatively large region of the psychometric function rather than to concentrate them around threshold. We then chose those levels for which at least 4 trials were available from each infant tested at a given frequency. Thus, each point on the function is estimated from a minimum of 24 trials.

Functions constructed in this way tend to be rather shallow in slope. Even though the functions of individual infants are reasonably steep, because infants differ in sensitivity, the group function averages performance at levels at which some infants are well above threshold, whereas others are well below. We also caution against trying to estimate average thresholds from these functions. The adaptive method in this experiment was deliber-
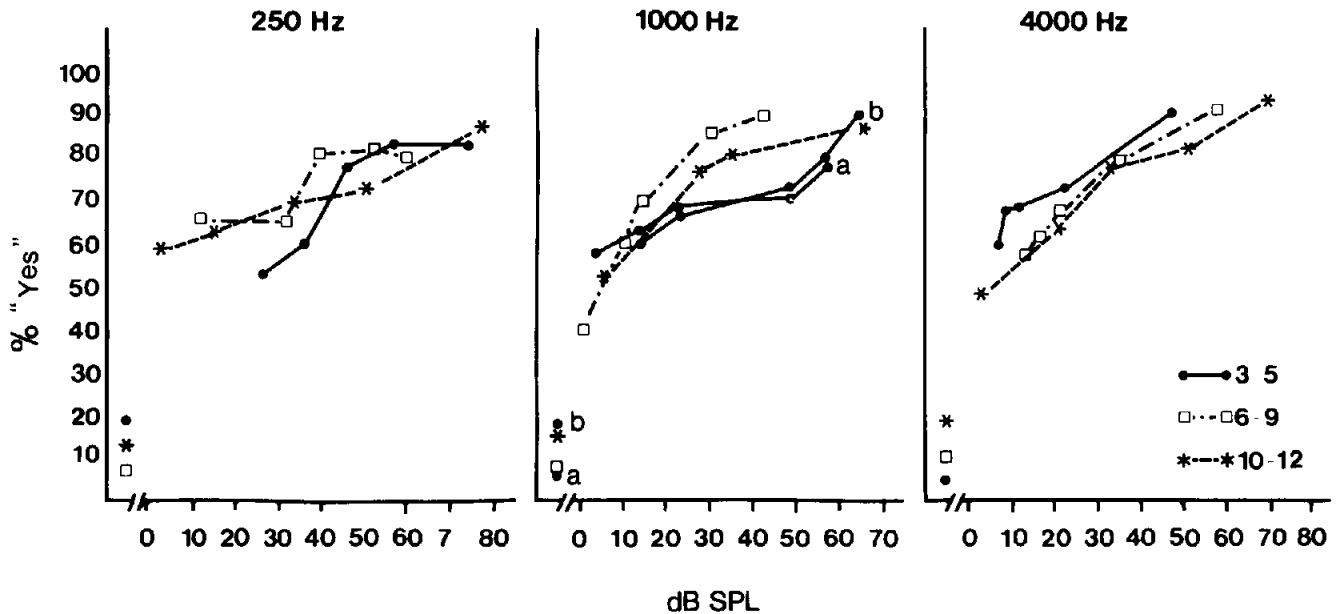
Figure 3. Psychometric functions obtained for three age groups of infants at each of three frequencies for pure-tone detection. (Two different groups of infants are included at 3–5 months of age, 1000 Hz; see Psychometric Functions section. False alarm rates are plotted at far left.)

ately designed to be a poor estimator of threshold. The levels closest to the individual infant's threshold tend to be excluded from the function because few infants received signals at those particular levels. Moreover, the average of the individual infant's thresholds is not necessarily equal to the threshold of the group average. The only purpose of this analysis was to determine whether the observer's ability to score infant responses improved with sound pressure level in a reasonable way.

The average psychometric functions that we obtained are shown in Figure 3. False alarm rate is plotted as the lower limit of the function (these points are discussed in more detail below). The first thing to notice about these functions is that they increase more or less monotonically with sound pressure level. We appear to have sampled over the mid- to upper range of the psychometric function, from a low of 40% to over 90% hits. At 250 Hz, 3- to 5-month-olds and 6- to 9-month-olds reached asymptotic performance at around an 85% *yes* response rate. At the other frequencies, the best performance levels averaged close to 90% at each age. Thus, in each case, we seem to be looking over a range of performance that would span threshold.

The fact that average asymptotic performance of observers did not approach 100% in the procedure may constitute a problem for OPP. Finney (1971) showed that if the limits of such a function do not approach 0 and 100%, then the variance of the sampling distribution for the parameters of the function fit to these data points will be increased. The lower limit of a yes–no psychometric function is the false alarm rate, so the precision of our threshold estimate is already reduced somewhat. Failing to reach 100% hit rate is likely to reduce the precision further. The confidence interval around the threshold also depends on the slope of the (individual) psychometric function. On the basis of the average individual function slope, following McKee,

Klein, and Teller's (1985) graphical-analysis approach, the 95% confidence limits associated with these threshold estimates should be less than ±4 dB, if the function ranges from 0 to 100%. For the case in which false alarm rate is about 10% and the upper asymptote is 90% the effect is not great; the confidence limits might be increased to ±5 dB. If false alarm rate is as high as 15% and the upper asymptote reaches only 85%, the limits may be as great as ±8 dB. Thus, substantial effects of limiting the range of the psychometric function would be expected only in the worst case observed here. At the same time, given the variability of infant behavior, if we can make procedural changes that will reduce the variability inherent in the statistical properties of the data, then it is certainly worthwhile to make those changes.

The functions of the 3- to 5-month-olds do appear to be reasonable. At the two lower frequencies, the average functions for the younger infants are shifted toward higher sound pressure levels by about 10 dB, but in other respects the functions for the 3- to 5-month-olds are not very different from those for the older infants. The slopes are quite similar, at least at 250 and 4000 Hz, and the best levels of performance achieved do not seem to be lower than those achieved for either 6- to 9-month-olds or 10- to 12-month-olds. Thus, we concluded that OPP works as well for young infants as it does for older ones and, because we had already established that OPP works as well for 6-month-olds as conditioned head-turns do for tone-detection thresholds, we felt that OPP was a promising technique for assessment of absolute sensitivity in younger infants.

False alarm rates. False alarm rate is used here to refer to the proportion of no-signal trials on which the observer judges that a signal was presented (i.e., a false positive response). Psychometric functions like those presented above are based on the

Table 1

False Alarm Data for Pure-Tone Detection
Sessions Using OPP

Age (in mo.)	N	Total no. sessions	Excluded (%)	Total no. no-signal trials	False alarm rate[a]
3	34	74	20	405	.110 (.078)
3[b]	17	30	12	196	.107 (.022)
6	37	70	17	325	.137 (.058)
6[b]	18	32	9	210	.081 (.017)

Note. OPP = Observer-based Psychoacoustic Procedure.
[a] After exclusion of sessions for which rate exceeded 0.25; figures in parentheses represent standard error. [b] Infants included in psychometric function study.

proportion of signal trials on which the observer judges that the infant responded to a signal. Clearly, if the observer responds *yes* fairly frequently, even when no signal is presented, such a function presents a distorted view of the infant's sensitivity. The nature of the distortion depends on the way that false alarm rate varies with the signal. If it is simply the case that the observer's false alarm rate remains constant over the course of a session regardless of the signal level, then the effect will be to displace the function toward lower values, yielding a better threshold. If the false alarm rate varies with signal level, then it is difficult to predict the effect. It is likely that both the position and the slope of the function will be affected. Even more important, if false alarm rate varies with age or with the interaction of age with other variables, then we cannot legitimately make threshold comparisons between age groups.

Because of the nature of an adaptive psychophysical procedure, it is difficult to establish false alarm rate as a function of, say, sound pressure level for individual infants. Among other reasons, there are typically few no-signal trials that could be assigned to a particular level for this purpose. We are able, however, to make statements about the false alarm rate for an individual infant over the entire test session and to compare these for different ages.

Recall that during testing, the observer is required to keep "local" false alarm rate below .25 and that we do not use sessions for which false alarm rate exceeds this limit. Table 1 summarizes the number of sessions excluded for this reason, as a function of age and frequency. The number of sessions so excluded is quite small and it does not seem to be higher for younger infants. Table 1 also lists the average false alarm rates in detection sessions for the 3- and 6-month-olds in the psychometric function study as well as for the 6-month-olds in the threshold study. The average rate for 3-month-olds is slightly higher than that for 6-month-olds in the psychometric function study. However, given the variability in rate, this difference is not statistically significant, nor is it likely, even if real, to lead to a significant difference in thresholds. In addition, if we combine the 6-month-olds from the two OPP studies, the average is about the same as that for the 3-month-olds. The OPP false

alarm rates are within the range reported for 6-month-olds in the conditioned head-turn detection studies cited earlier (Berg & Smith, 1983; Nozza & Wilson, 1984; Sinnott et al., 1983). We feel this congruity provides rather convincing evidence that an observer using OPP can do as well in terms of maintaining as strict a response criterion as a 6-month-old does in producing head-turns in response to sound.

Frequency Discrimination

Although a variety of psychoacoustic measures are based on detection of sounds under different conditions, there are many capacities that are typically measured in a discrimination paradigm. Again, because we were concerned with frequency discrimination by 6-month-old infants in earlier studies, we used that task to evaluate the usefulness of OPP for testing discrimination. In frequency discrimination, we examine the listener's ability to distinguish between tones of different frequencies as the frequency difference between them is varied. It proved more difficult to obtain frequency discrimination thresholds (i.e., frequency difference limens [FDLs], the smallest discriminable change in the frequency of a pure tone) than it was to obtain detection thresholds from infants.

The greater difficulty with frequency discrimination may not be surprising. The stimulus situation is illustrated in Figure 4. The infant hears a stimulus on every trial. On half of the trials, the frequency of the repeating tone burst changes on alternate bursts; on the remaining trials, the frequency of the tone burst remains constant through the trial. The observer has to learn to distinguish trials on which the stimulus stays at one frequency from trials on which the stimulus frequency is changing. It is not uncommon to see an infant producing easily observable responses on both types of trials at the beginning of a frequency discrimination session. However, because the visual reinforcer is activated only following a frequency-change trial, we hoped that the infants would stop responding on no-change trials or learn to respond differently to the changing stimulus. Although our preliminary findings seemed to indicate that the infants did this, sensitivity using OPP was not very good. In initial attempts (Olsho, 1984a), we obtained FDLs of 3% to 32%, even for 6-month-olds, whereas we reported FDLs in this age group of 1% to 4% in previous conditioned head-turn studies (Olsho et al., 1982a, 1982b; Olsho, 1984b).

However, two procedural changes resulted in the FDLs we describe below. The duration of the tone bursts presented to the infants was increased from 200 ms to 500 ms, and we switched from an adaptive procedure to the method of constant stimuli. The infants were still trained using a large frequency change, but, during testing, on frequency-change trials they heard one of six different frequency changes, randomly ordered over trials. No-change trials occurred on about 35% of trials during testing using this procedure.

A cumulative normal curve was fit to each infant's data points. The value was taken as the value of Δf at which 70.7% of the responses would have been *yes*. This would correspond to the definition of threshold that we used in the earlier studies. If the infant's function was flat or undefined, the data were not
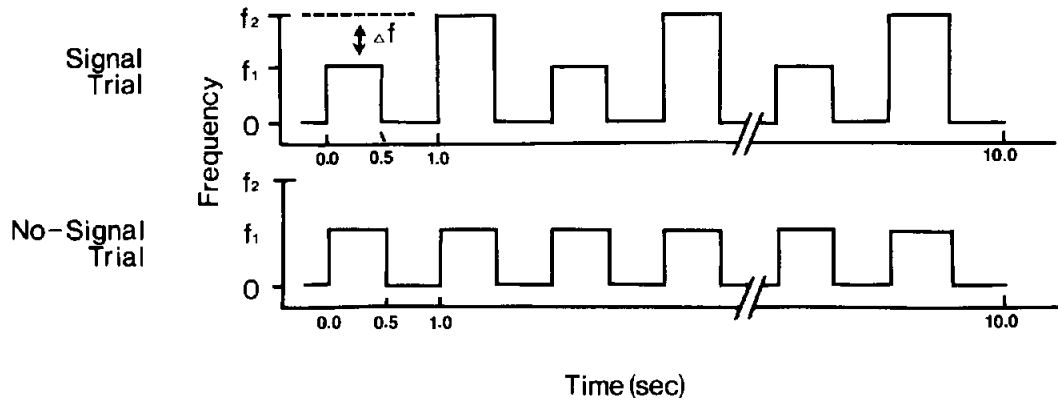
Figure 4. Stimulus conditions for frequency discrimination experiment.

used. This occurred in three 3-month-olds and in three 6-month-olds in the sample of 52 infants.

The first comparison we made was to our previously published FDLs for 6-month-olds. Figure 5 shows the relative FDL as a function of frequency, using OPP for 6-month-olds and the same measure for 5- to 8-month-olds from Olsho et al. (1982b) and from Olsho (1984b). Also plotted is the relative FDL reported by Sinnott and Aslin (1985). Although Sinnott and Aslin (1985) used an adaptive procedure that would converge on the 50% correct point, theirs is the only other published study of infant FDLs.

The FDLs we obtained from 6-month-olds using OPP were never as low as those reported by Olsho (1984b). At the same time, the FDLs at 1000 Hz, obtained using OPP for infants in

this age range, agree quite well with those reported by Olsho et al. (1982a) and by Sinnott and Aslin (1985). Moreover, the shape of the function relating FDL to frequency parallels that reported by Olsho (1984b). Why performance in the Olsho (1984b) study is better than that in the other studies is not clear. However, we felt justified in concluding that OPP produced reasonable FDLs for 6-month-olds, on the basis of the agreement between studies at 1000 Hz and the similarity to Olsho (1984b) as a function of frequency.

Psychometric functions. In regard to pure-tone sensitivity, psychometric functions for frequency discrimination were examined for 3- and 6-month-olds. These functions, showing percent correct as a function of Δf, are shown in Figure 6; the functions for each frequency are plotted separately. For both age
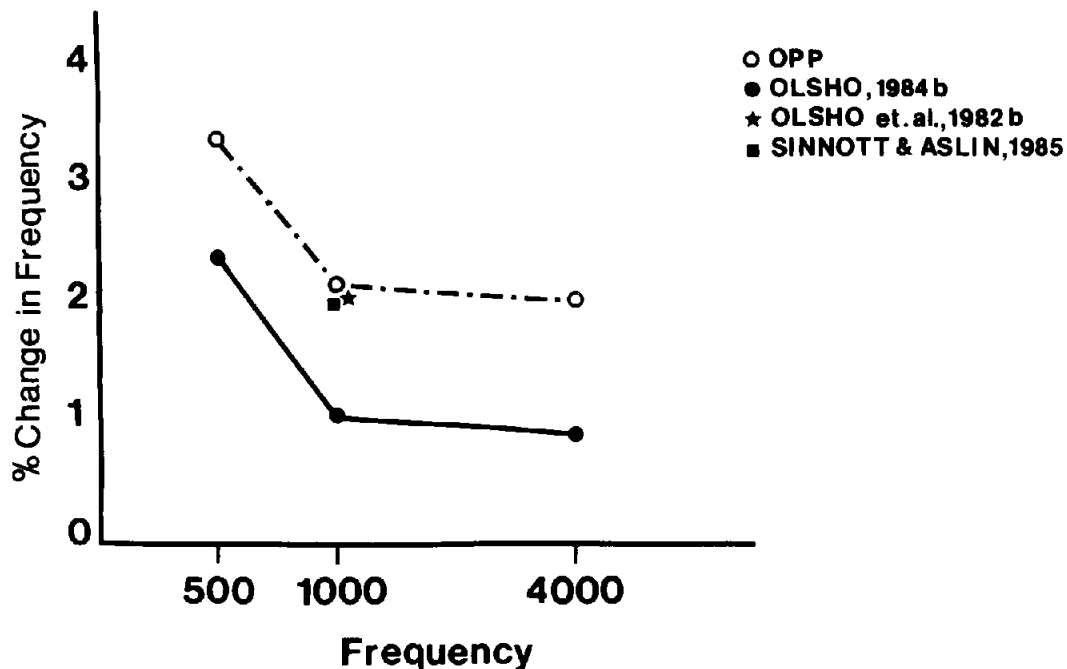

Figure 5. A comparison of frequency difference limens obtained in three studies.
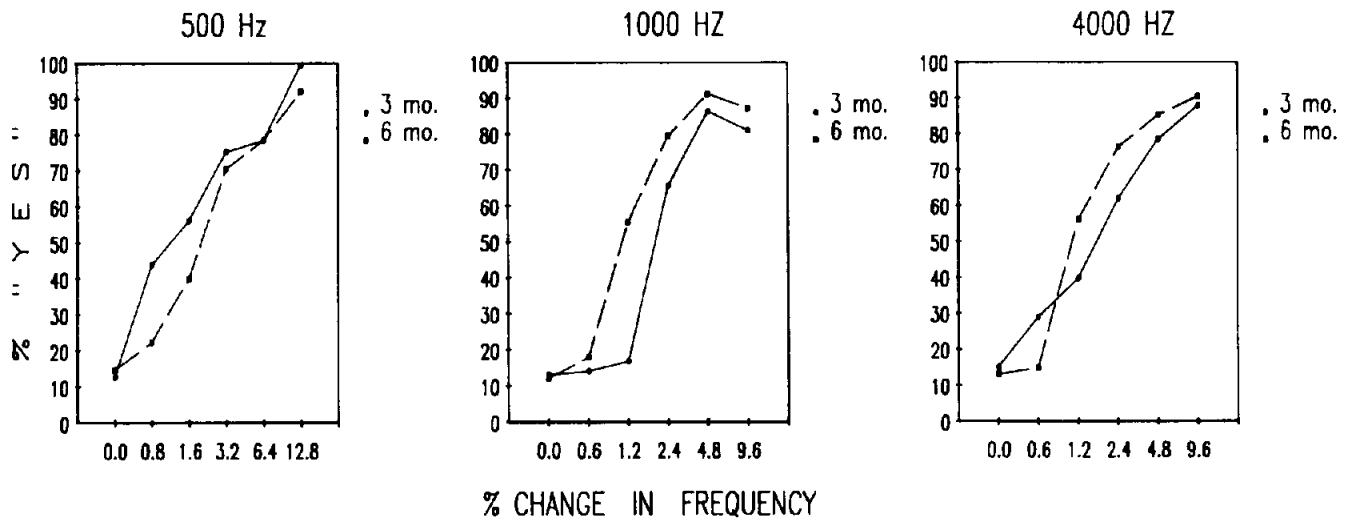
Figure 6. Psychometric functions in frequency discrimination for 3- and 6-month-olds at three different frequencies. (False alarm rates are plotted at 0% change in frequency.)

groups, percent correct increases in a nearly monotonic fashion with the size of the frequency change. The 3-month-olds seem quite similar to the 6-month-olds, at least in the slopes of the functions. Asymptotic performance ranges from an 85% to 95% hit rate; false alarm rates range between 10% and 15%. (The comments made earlier concerning the effects of restricting the range of the function on the precision of the threshold estimate must be kept in mind here as well.)

False alarm rates. False alarm rates in frequency discrimination using OPP tended to be a little higher than for pure-tone detection. Table 2 summarizes the number of sessions excluded for excessive false alarm rate and the false alarm rates for sessions in which the rate was below the specified limit. Although false alarm rate was 3% to 5% higher in frequency discrimination than in pure-tone detection using OPP, Sinnott and Aslin (1985) also reported a higher false alarm rate for frequency discrimination (about 15%) than Sinnott et al. (1983) reported for pure-tone detection (less than 10%). Again, we concluded that OPP produces data that are in line with conditioned head-turn results, and that it can do so with younger infants as well as 6-month-olds.

Table 2
False Alarm Data for Frequency Discrimination Sessions Using OPP

Age (in mo.)	N	Total no. sessions	Excluded (%)	Total no. no-signal trials	False alarm rate[a]
3	41	98	20	1614	.135 (.047)
6	33	87	11	1680	.137 (.046)

Note. OPP = Observer-based Psychoacoustic Procedure.
[a] After exclusion of sessions for which rate exceeded 0.25, figures in parentheses represent standard error.

Discussion

OPP represents a technique for psychoacoustic testing of young infants in a variety of tasks. It is relatively easy to train observers in the technique, and the training of specific infant-observer pairs is not difficult. Few infants are untestable using OPP. We have demonstrated that OPP is as sensitive a technique as conditioned head-turns for 6-month-old listeners and that it produces reasonable psychometric functions for 3-month-olds, at least for pure-tone sensitivity and for frequency discrimination. We are currently using the technique in several other tasks, including gap-detection and masked-detection thresholds. There are many other auditory capacities that could be assessed using OPP as well.

There are limitations to this technique, of course. The most essential limitation is that if an observer is unable to achieve above-chance performance on a particular infant, then it is not entirely clear that the infant can't hear. As Teller (1979) put it in discussing the FPL technique for infant visual psychophysics, "If the observer's performance is at chance, it follows that the information . . . was lost somewhere between the display and the observer, hopefully (but not necessarily) within the infant's sensory visual system" (p. 136). If we can demonstrate that the observer achieves very high levels of performance in at least one condition, but not in others, then we can argue, as Teller does, that the infant is at least as sensitive as our results indicate. Whenever the observer performs above chance we can be assured that the infant has responded in some reliable, observable fashion.

An additional point is that the observer–infant team's failure in some condition must be interpreted within the context in which it occurs. For example, if we are investigating pure-tone sensitivity and fail to surpass chance performance at low sound pressure levels, then it would not be unreasonable to suggest that the infant's response to very quiet sounds is so subtle that the observer is unable to detect it. Then, the observer's thresh-

old for detecting an infant response is higher than the infant's threshold for detecting a sound, and the infant's capacity would be underestimated. On the other hand, in considering a frequency-discrimination experiment, let us assume that we have equated stimuli of different frequencies along relevant dimensions such as effective intensity. If an infant responds to a frequency change at one frequency, but fails to respond to an equivalent change at another frequency, would it be reasonable to argue that the subtlety of the infant's response to frequency changes is frequency dependent? Unless there is evidence that the infant's responsiveness varies in some specific fashion with tone frequency, even after absolute sensitivity differences have been taken into account, then a stronger case could be made here that this is an auditory effect. In other words in many situations there is no a priori reason to suppose that the sensitivity of the observer–infant team should change over conditions. Even though performance in any given stimulus condition may underestimate the infant's capacity, relative performance levels in different conditions should not be affected.

Although the problem of interpreting negative results is inherent to the procedure, there are other limitations that might be overcome by future modifications. As we discussed earlier, we chose a single-interval approach: On each trial the observer responds *yes* or *no*. As we have pointed out, thresholds obtained in such a procedure will be influenced by the response criteria of the infant–observer pair. We control response bias effects by requiring the observer to maintain a false alarm rate below some relatively conservative level, but observers do not easily adopt this strategy. This may occur because signal trials are presented more frequently than no-signal trials and because we are forcing the observer to operate at a false alarm rate far below the actual frequency of signal trials (i.e., at a very conservative response criterion). If we were to adopt a procedure in which signal and no-signal probabilities were equal, we could then instruct the observer to maximize the percentage of correct judgments, and there would be no benefit to a response bias in either direction.

In order to arrive at that point, however, two additional developments are necessary. First, some means of increasing the total number of trials obtained from each infant would be needed in order to accommodate the additional no-signal trials. This is a particular problem when many trials are required to estimate a threshold. For example, in the frequency discrimination studies we attempted to obtain 60 signal trials from each infant. That would mean about 20 no-signal trials at a signal probability of 65%, but about 30 no-signal trials at a signal probability of 50%. Changing the reinforcers used or, perhaps, the schedule of reinforcement might make it possible to obtain more trials. Second, additional data regarding infant auditory sensitivity are needed. If signal and no-signal trial probabilities are equal in a staircase procedure, the likelihood of a long string of unreinforced trials at near-threshold levels is quite high. Once we have a better idea of where the infant's threshold will fall, however, we will be able to use a restricted set of signal values, increase the number of trials for each value, and, by randomly ordering values, reduce the chances that many trials will be presented before reinforcement is available. Adding "probe" trials, on which signals at suprathreshold levels are presented but not

used to change the staircase, would be another alternative. Some means to accommodate the additional trials would still be required.

Another issue that we must address in the development of the technique is the failure to push the upper asymptote of the psychometric function to 100%. As discussed earlier, the statistical properties of threshold estimates based on functions in which the upper asymptote falls much below 100% are poor (Finney, 1971; McKee et al., 1985). At this point, we believe that this limitation can be overcome if we manipulate the reinforcer in such a way as to generally increase the magnitude of the infant's response. Clearly, the reinforcers that we have always used for 6-month-olds may not be the most effective for 3-month-olds. There may also be techniques for varying the reinforcer over the course of a session (e.g., by randomly activating one of two reinforcers) that would make it easier for the observer to interpret an infant's behavior.

In summary, we feel that OPP represents a useful and a potentially powerful technique for assessing auditory capacity in very young infants. It may even work for infants younger than 3 months of age (Trehub and Schneider [1986] have recently presented preliminary data from an observer-based technique applied to infants as young as 1 month of age). As such additional applications are made in different laboratories, the limits of OPP will be more firmly established.

References

Aslin, R. N., Pisoni, D., Hennessey, B., & Perey, A. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. *Child Development, 52,* 1135–1145.

Aslin, R. N., Pisoni, D. B., & Jusczyk, P. W. (1983). Auditory perception and speech perception in infancy. In P. H. Mussen (Series Ed.) & M. M. Haith & J. J. Campos (Vol. Eds.), *Handbook of child psychology: Vol. 2. Infancy and developmental psychobiology* (4th Ed.) New York: Wiley.

Berg, K. M., & Smith, M. C. (1983). Behavioral thresholds for tones during infancy. *Journal of Experimental Child Psychology, 35,* 409–425.

Bull, D., Schneider, B. A., & Trehub, S. E. (1981). The masking of octave-band noise by broad-band spectrum noise: A comparison of infant and adult thresholds. *Perception and Psychophysics, 30,* 101–106.

Clifton, R. K., Morongiello, B. A., Kulig, J. W., & Dowd, J. M. (1981). Developmental changes in auditory localization in infancy. In R. N. Aslin, J. R. Alberts, & M. R. Petersen (Eds.), *Development of perception: Psychobiological perspectives.* New York: Academic Press.

Eilers, R. E. (1977). Context sensitive perception of naturally produced stop and fricative consonants by infants. *Journal of the Acoustical Society of America, 61,* 1321–1336.

Eisenberg, R. B. (1976). *Auditory competence in early life.* Baltimore, MD: University Park Press.

Finney, D. J. (1971). *Probit analysis* (3rd ed.). Cambridge, England: Cambridge University Press.

Grimwade, J. C., Walker, D. W., Bartlett, M., Gordon, S., & Wood, C. (1971). Human fetal heart rate and movement in response to sound and vibration. *American Journal of Obstetrics and Gynecology, 190,* 86–90.

Hall, J. L. (1968). Maximum likelihood sequential procedure for esti-

mation of psychometric functions. *Journal of the Acoustical Society of America, 44,* 370.

Hall, J. L. (1981). Hybrid adaptive procedure for estimation of psychometric functions. *Journal of the Acoustical Society of America, 69,* 1763–1769.

Hecox, K. (1975). Electrophysiological correlates of human auditory development. In L. B. Cohen & P. Salapatek (Eds.), *Infant perception: From sensation to cognition* (pp. 151–191). New York: Academic Press.

Hecox, K., & Burkard, R. (1982). Developmental dependencies of the human brainstem auditory evoked response. *Annals of the New York Academy of Sciences, 388,* 538–556.

Hertzog, C. (1980). Applications of signal detection theory to the study of psychological aging: A theoretical review. In L. W. Poon (Ed.), *Aging in the 80s: Psychological issues* (pp. 568–591). Washington, DC: American Psychological Association.

Klein, A. J. (1984). Frequency and age-dependent auditory evoked potential thresholds in infants. *Hearing Research, 16,* 291–297.

Mayer, D. L., & Dobson, V. (1982). Visual acuity development in infants and young children, as assessed by operant preferential looking. *Vision Research, 22,* 1141–1151.

McKee, S. P., Klein, S. A., & Teller, D. Y. (1985). Statistical properties of forced-choice psychometric functions: Implications of probit analysis. *Perception and Psychophysics, 37,* 286–298.

McNicol, D. (1972). *A primer of signal detection theory.* Sydney, Australia: George Allen & Unwin Ltd.

Moore, J. M., Thompson, G., & Thompson, M. (1975). Auditory localization of infants as a function of reinforcement conditions. *Journal of Speech and Hearing Disorders, 40,* 29–34.

Moore, J. M., & Wilson, W. (1978). Visual reinforcement audiometry (VRA) with infants. In S. E. Gerber & G. T. Mencher (Eds.), *Early diagnosis of hearing loss* (pp. 177–213). New York: Grune and Stratton.

Nozza, R. J., & Wilson, W. R. (1984). Masked and unmasked pure-tone thresholds of infants and adults: Development of auditory frequency selectivity and sensitivity. *Journal of Speech and Hearing Research, 27,* 613–622.

Ogilvie, J. C., & Creelman, C. D. (1968). Maximum-likelihood estimation of Receiver Operating Characteristic curve parameters. *Journal of Mathematical Psychology, 5,* 377–391.

Olsho, L. W. (1984a, April). *Preliminary results of an observer-based method for infant auditory testing.* Paper presented at the International Conference for Infant Studies, New York.

Olsho, L. W. (1984b). Infant frequency discrimination. *Infant Behavior and Development, 7,* 27–35.

Olsho, L. W. (1985). Infant auditory perception: Tonal masking. *Infant Behavior and Development, 8,* 371–384.

Olsho, L. W., Schoon, C., Sakai, R., Turpin, R., & Sperduto, V. (1982a). Auditory frequency discrimination in infancy. *Developmental Psychology, 18,* 721–726.

Olsho, L. W., Schoon, C., Sakai, R., Turpin, R., & Sperduto, V. (1982b). Preliminary data on frequency discrimination in infancy. *Journal of the Acoustical Society of America, 71,* 509–511.

Schneider, B. A., & Trehub, S. E. (1985). Infant auditory psychophysics: An overview. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life.* New York: Academic Press.

Schneider, B. A., Trehub, S. E., & Bull, D. (1980). High-frequency sensitivity in infants. *Science, 207,* 1003–1004.

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.

Sinnott, J. M., & Aslin, R. N. (1985). Frequency and intensity discrimination in human infants and adults. *Journal of the Acoustical Society of America, 78,* 1986–1992.

Sinnott, J. M., Pisoni, D. B., & Aslin, R. N. (1983). A comparison of pure tone auditory thresholds in human infants and adults. *Infant Behavior and Development, 6,* 3–17.

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

Teller, D. Y. (1979). The forced-choice preferential looking procedure: A psychophysical technique for use with human infants. *Infant Behavior and Development, 2,* 135–153.

Teller, D. Y. (1985). Psychophysics of infant vision: Definitions and limitations. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life* (pp. 127–144). New York: Academic Press.

Teller, D. Y., Morse, R., Borton, R., & Regal, D. (1974). Visual acuity for vertical and diagonal gratings in human infants. *Vision Research, 14,* 1433–1439.

Trehub, S. E., & Schneider, B. A. (1986, April). *Testing auditory sensitivity in young infants.* Paper presented at the International Conference for Infant Studies, Los Angeles.

Trehub, S. E., Schneider, B. A., & Endman, M. (1980). Developmental changes in infants' sensitivity to octave-band noises. *Journal of Experimental Child Psychology, 29,* 283–293.

Watrous, B. S., McConnell, F., Sitton, A. B., & Fleet, W. F. (1975). Auditory responses of infants. *Journal of Speech and Hearing Disorders, 40,* 357–366.