

Developmental perspectives on the localization and detection of auditory signals

BRUCE A. SCHNEIDER, SANDRA E. TREHUB, and LEIGH THORPE
Erindale Campus, University of Toronto, Mississauga, Ontario, Canada

Responsiveness of 1-, 3-, and 5-year-old children and adults to octave-band noises at .4 and 10 kHz was assessed with a go/no-go version of visual reinforcement audiometry (VRA) (Moore, Thompson, & Thompson, 1975) and a two-alternative, forced-choice version (Suzuki & Ogiba, 1961; Trehub, Schneider, & Endman, 1980). Infants performed better on the two-alternative, forced-choice version in quiet and in noisy backgrounds, and adults performed better on the two-alternative, forced-choice version in quiet but not in noisy backgrounds. Performance on the two tasks was essentially equivalent for 3- and 5-year-old children. Superior performance on two-alternative VRA over go/no-go may be due to lesser cognitive demands in the case of infants and to the engagement of superior decision strategies in the case of adults.

Over the past two decades, visual reinforcement techniques (visual reinforcement audiometry, or VRA) have been used in developmental studies of basic auditory processes. In these techniques, the head-turning component of an orienting response (Sokolov, 1963) to sound is followed by the presentation of a visual reinforcer, usually an animated toy. The head-turning or localization response can be elicited by 5 or 6 months of age (Chun, Pawsat, & Forster, 1960), and reinforcement can maintain it at high levels for 30-40 trials within a single session (Moore, Thompson, & Thompson, 1975; Suzuki & Ogiba, 1961; Trehub, Schneider, & Bull, 1981). Indeed, the robust nature of this technique and its applicability in the early years have made it the method of choice for studying absolute thresholds (Berg & Smith, 1983; Olsho, 1985; Schneider, Trehub, & Bull, 1980; Sinnott, Pisoni, & Aslin, 1983; Trehub, Bull, Schneider, & Morrongiello, 1986; Trehub, Schneider, & Endman, 1980; Trehub, Schneider, Morrongiello, & Thorpe, 1989), masked thresholds (Bull, Schneider, & Trehub, 1981; Nozza & Wilson, 1984; Schneider, Trehub, Morrongiello, & Thorpe, 1989; Trehub, Bull, & Schneider, 1981), critical bands (Olsho, 1985; Schneider, Morrongiello, & Trehub, 1990), and binaural unmasking (Nozza, 1987; Schneider, Bull, & Trehub, 1988) in infants and young children.

At the present time, two versions of this technique are used with infants to determine absolute and masked thresholds. In the single-loudspeaker version developed by Moore et al. (1975), the infant sits on the parent's lap

with a loudspeaker located to one side. The experimenter sits facing the infant and initiates a trial when the infant is looking directly ahead. The signal is then presented for a limited duration and the experimenter indicates the presence or absence of a turn toward the loudspeaker within a fixed time following signal onset. Turns during this response interval are reinforced by the illumination and activation of toys near the loudspeaker. Control or no-signal trials are interspersed with signal trials to ensure that the infant is turning to the signal, not merely to the locus of interesting visual events. Adaptive procedures are typically used, and infants who turn on "too many" no-signal trials are often eliminated, although the number of such false alarms and other exclusion criteria vary considerably across studies (Berg & Smith, 1983: greater than 0.33 false-alarm rate; Sinnott et al., 1983: 2 or more false alarms).

In the two-alternative, forced-choice version (2AFC; see, e.g., Schneider et al., 1980, 1989; Trehub et al., 1980; Trehub et al., 1989), loudspeakers are located to the infant's left and right. As in the single-loudspeaker version, the experimenter faces the infant and initiates a trial only when the infant is looking directly ahead. In both versions, experimenter and parent wear headphones with masking noise to prevent them from detecting the signal. Test trials in 2AFC differ, however, in that a signal is presented on every trial on one of the two loudspeakers, the signal remaining on until the infant turns 45° to the left or right. Correct responses (turns toward the loudspeaker producing the signal) are also reinforced by a toy near the loudspeaker and incorrect responses result in a short intertrial delay before the next trial.

Despite their similarities, the techniques, as typically used in threshold studies, have a number of potentially important differences. In the single-loudspeaker version, the listener indicates the presence of a signal by a head turn and its absence by not responding. Essentially, this is a single-interval, go/no-go signal detection task (G/NG-D).

This research was supported by grants from the Medical Research Council of Canada and the University of Toronto. We thank Marilyn Barras, who conducted all testing for these studies. Leigh Thorpe is now at Bell Northern Research, Department 9Y72, P.O. Box 3511, Station C, Ottawa, Ontario K1Y 1H0. Requests for reprints should be sent to B. Schneider, Department of Psychology, University of Toronto, Erindale Campus, Mississauga Road, Mississauga, Ontario L5L 1C6, Canada.

(For a discussion of go/no-go procedures in animal psychophysics, see Blough & Blough, 1977.) By contrast, the two-loudspeaker technique involves localization rather than detection. A signal is presented on every trial (right or left side), and the listener's task is to indicate the sound source. In essence, then, it is a single-interval, two-alternative, forced-choice localization task (2AFC-L).¹ Although it is possible, in principle, to have equivalent signal durations in both versions, it has proved advantageous, in practice, to use flexible rather than fixed durations in 2AFC-L, simply waiting until an infant responds (Trehub et al., 1981). In the G/NG-D task, signal duration must be fixed so that infants can be scored as responding or not responding on each trial.

How can one relate performance on the localization task, which involves binaural processing, to that on the detection task, which does not? It is clear that failure to localize a sound would not preclude its detectability. With adult subjects and certain auditory stimuli, thresholds for lateralization (i.e., detecting the ear of presentation in an earphone experiment) are considerably higher than detection thresholds (Egan & Benson, 1966). Although the process of locating the source in a sound field differs in many ways from identifying the ear of presentation in an earphone experiment, a similar dissociation between detection and localization thresholds might occur in the sound field. Furthermore, the existence and extent of this dissociation might be age-related if the mechanisms subserving detection and localization mature at different rates. Indeed, several investigators have criticized the use of localization measures in the study of auditory sensitivity (Aslin, Pisoni, & Jusczyk, 1983; Clifton, Morrongiello, Kulig, & Dowd, 1981).

To compare these techniques, it is instructive to examine both tasks from a signal detection perspective. On any G/NG signal trial, the presence of a signal presumably initiates activity in the auditory nervous system over and above activity that continues in the absence of a signal (i.e., noise). The distributions of activity on no-signal trials (noise, or *N*) and on signal trials (signal plus noise, *SN*) are assumed to be normally distributed, with equal variance (see Figure 1), the difference between the means of these distributions (expressed in standard deviation units) being d' .

In signal detection theory, it is assumed that an observer (i.e., listener) locates a criterion, *C*, along the decision axis to maximize payoff. If an observation falls to the right of the criterion, the observer judges that a signal is present; if an observation falls to the left, the observer judges that it is absent. It is also assumed that the observer is free to locate the criterion at any point along the decision axis, with the location influenced by the payoff for correct and incorrect decisions, as well as the relative probability of signal and no-signal trials. In the typical application of G/NG-D procedures with infants, the listener is reinforced for correct detections, but there is no penalty for false alarms (i.e., head turns on no-signal trials). Moreover, when adaptive procedures are used, the

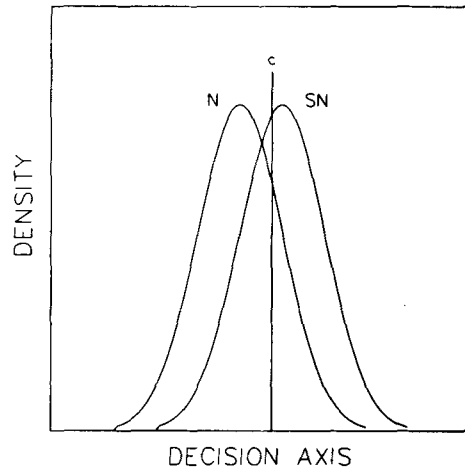


Figure 1. Hypothetical distributions of events associated with no-signal (noise) trials and signal-plus-noise trials. The observer's criterion is indicated by the vertical line at *c*.

probability of signal trials generally exceeds that of no-signal trials (Sinnott et al., 1983), which leads the rational observer to maximize reinforcement by moving the criterion far to the left. The only consequence of adopting this strategy is exclusion from the study if the observer's false-alarm rate exceeds some predetermined value.

From this perspective, one might expect all infant data gathered in this manner to be excluded on the basis of false-alarm and related criteria (e.g., not responding on selected probe or suprathreshold trials). Indeed, Berg and Smith (1983) excluded 36% of their 10-month-olds and 54% of their 14-month-olds, whereas Sinnott et al. (1983) excluded 41% of all test sessions with 7- to 11-month-olds. Perhaps another aspect of the G/NG-D procedure helps keep the false-alarm rate in check—namely, the experimenter's active manipulation of toys in front of the infant. In fact, Berg and Smith (1983) describe the experimenter as being an *entertainer* (pp. 412-413). Although this may preclude an elevated false-alarm rate, it may also compete with the signal for the infant's attention. Nevertheless, no-signal trials are likely to promote boredom and fussiness for infant participants, and entertainment could alleviate such boredom. By contrast, 2AFC-L, with signal presentation on every trial, may cause less boredom. Moreover, in the G/NG-D situation, returning to midline after responding (i.e., readiness for the next trial) results in a 0.5 or greater probability of signal presentation, with its associated potential for reinforcement. With 2AFC-L, midline orientation has a much greater potential payoff, a 1.0 probability of signal presentation and reinforcement potential.

In short, 2AFC-L avoids some of the difficulties of G/NG-D but has a potentially fatal flaw if localization performance is inferior to detection performance. In the present study, we examined performance in both tasks by 1-, 3- and 5-year-old children and adults as they attempted

to detect .4- and 10-kHz octave-band noises presented in quiet (Experiment 1) and in a background of masking noise (Experiment 2).

During the test phase of both tasks, 16 trials of one type were interspersed with 16 trials of another type. In the G/NG-D and 2AFC-L tasks, Type 1 trials consisted of stimulus presentation on the left loudspeaker and Type 1 responses consisted of left head turns (or left-button presses). The distribution of events evoked by Type 1 signals is shown for G/NG-D tasks in Figure 2A and for 2AFC-L tasks in Figure 2B. Also shown in Figure 2 are the distributions of events evoked by Type 2 signals in both procedures. Note, however, that Type 2 signals and Type 2 responses differ for the two tasks. In the G/NG-D task, Type 2 events are those that occur in the absence of signals, and Type 2 responses consist of no responses during the designated response interval. In the 2AFC-L task, Type 2 events are those elicited by signals presented over the right loudspeaker, while Type 2 responses are right head turns (or right-button presses). The separation between Type 1 and Type 2 distributions need not be the same, because Type 2 distributions could be different for the two tasks. Because the stimuli are presented over loudspeakers, the two ears may be involved in detection and/or localization responses. When a signal is presented over the left loudspeaker, the sound arrives first at the left ear, and is less intense at the right ear because of the shadow cast by the head. Therefore, listeners could base their detection and/or localization responses on, for example, the amount of energy in the left ear, interaural amplitude differences, or interaural time differences. In the G/NG-D procedure, interaural differences are unlikely to aid detection and probably do not contribute to decisions about the

presence of a signal. For example, the auditory system may simply count the number of neural impulses during a trial and base its decision on that count, ignoring interaural differences in this or any other parameter. In the 2AFC-L procedure, however, the auditory system uses interaural differences to localize the sound. Clearly, then, the separation between the two distributions, and indeed the nature of the decision axis, may differ in these tasks. In Figure 2, this possibility is indicated by representing the two distributions as having unequal differences between their means. Whether or not the separation between the two distributions (in normal deviate units or d') actually differs for the two tasks remains an empirical question.

EXPERIMENT 1

Listeners were tested in two sessions at one of two frequencies (.4 and 10 kHz), with a single stimulus intensity presented in a session. They were tested in the G/NG-D task in the first session and in the 2AFC-L task in the second, with 32 test trials in each session. In Session 1, the signal was presented on half of the trials, the other half being no-signal trials. In Session 2, the signal appeared on the left loudspeaker on half of the trials and on the right loudspeaker on the remaining trials. Performance in both sessions was measured in terms of d' .

Method

Subjects

Infant and child participants were recruited primarily from letters sent to local families. Adult participants were university students who responded to posted notices. Of the 238 participants recruited for this experiment, 6 were not tested because of colds or congestion on the day of testing (4 at 1 year and 2 at 5 years); 8 were eliminated because of documented hearing losses, the presence of ventilation tubes, or because of a family history of early hearing loss (5 at 5 years and 3 adults); 3 failed to train (2 at 1 year and 1 at 3 years); 10 failed to complete one of the sessions because of fussing or crying (3 at 1 year and 7 at 3 years); and 4 of the 5-year-olds declined to participate in the second session. The 1-year-old children were tested on the G/NG-D and 2AFC-L tasks at different times separated by no more than 12 days. Some of these children (7) were unable to return for the second testing session, and their data are not included. Older children and adults were tested on both tasks on the same day. The remaining 200 participants consisted of four age groups, each with 50 participants. Half of each age group was tested with the .4-kHz signal and half with the 10-kHz signal. The age groups were 12 months (± 2 weeks); 3 years (± 3 months); 5 years (± 3 months); and young adults (18-25 years).

Apparatus and Stimuli

The physical configuration of the apparatus was identical to that of Schneider, Trehub, Morrongoello, and Thorpe (1986). Octave-band stimuli were produced by directing the output of a noise generator (General Radio No. 1952) to a programmable bandpass filter (Brüel & Kjaer No. 1617) set to pass an octave band centered at .4 or 10 kHz. The rate of energy falloff on each side of the octave band was 30 dB per octave. Octave bands were chosen principally because comparison data were available over a wide age range (Schneider et al., 1986; Trehub et al., 1980; Trehub, Schneider, Morrongoello, & Thorpe, 1988). Although narrowing the bandwidth would provide better frequency specificity, it would also increase

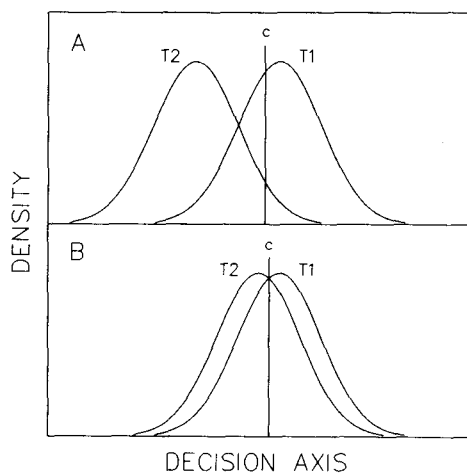


Figure 2. Hypothetical distributions of events associated with Type 1 (T1) and Type 2 (T2) trials. (A) T2 trials are noise trials and T1 trials are signal-plus-noise trials. (B) T1 trials are signal-on-left trials and T2 trials are signal-on-right trials. Because discrimination between the two types of trials in A and B is likely to be based on different events, the distributions are shown as having different separations between their means (different d' values).

amplitude variation in the sound field (Dillon & Walker, 1982). Octave bands minimize amplitude variation, thereby increasing the precision of the threshold measures while still retaining adequate frequency specificity.

The output of the filter was directed to two parallel circuits containing programmable electronic switches, attenuators, amplifiers, and loudspeakers. One circuit controlled stimulus presentation to the left loudspeaker, the other controlled presentation to the right loudspeaker. Stimuli were turned on and off with a linear ramp of 25 msec.

The loudspeakers were placed in an Industrial Acoustics sound-attenuating chamber (double-wall, measuring $3 \times 2.8 \times 2.8$ m), 1.8 m from the center of the listener's chair, which occupied one corner of the room. A chair for the experimenter was located in the corner opposite the listener's chair. Each loudspeaker was at a 45° angle to the listener's left and right. Below each loudspeaker was a four-chamber, smoked Plexiglas box with four different mechanical toys that served to reinforce correct responses. Adjacent to each loudspeaker was a portable color television set (Sony Trinitron, Model KV-1911), which was also used for reinforcement during the second session with 5-year-old children. The viewing screen was blacked out during periods of nonreinforcement. During the first session, a ceiling-length curtain was drawn over the right corner of the room, which prevented the participant from seeing the right loudspeaker, toy box, and television screen.

Calibration procedures were identical to those followed in Schneider et al. (1986). Only one stimulus intensity was employed at each frequency \times age condition. Stimulus intensities at .4 kHz were 33, 25, 25, and 10 dB SPL for 1-year-olds, 3-year-olds, 5-year-olds, and adults, respectively. Stimulus intensities at 10 kHz were 25, 25, 18, and 12 dB SPL for 1-year-olds, 3-year-olds, 5-year-olds, and adults, respectively. Levels were selected on the basis of previous studies (Schneider et al., 1986; Trehub et al., 1980) to produce 2AFC-L, d' values between 1.5 and 2.5.

Procedure

Session 1: G/NG-D. During the test session, 3- and 5-year-old children and adults were seated in a test chair equipped with a push-button on each arm. They were instructed to look directly ahead and to indicate the presence of the signal by pressing the left-hand button, a response that was automatically recorded. For the 1-year-olds, the parent sat in the test chair with the child on her lap facing the experimenter. The appropriate response to a signal was a left head turn of 45° or greater, which the experimenter recorded by means of a hand-held pushbutton. (When a second observer rescored a randomly selected subset of sessions from videotapes, inter-observer agreement was 100%.) During the test session, the parent and the experimenter wore headphones with continuous broadband noise to prevent them from discriminating between signal and no-signal trials.

Trials were initiated only when the listener (infant, child, or adult) was directly facing the experimenter. To achieve midline orientation with 1-year-olds, the experimenter employed a hand-held toy to attract the child's attention. Older children and adults were not instructed to hold their heads still during a trial. Once midline orientation was obtained, the experimenter pressed a button to initiate a trial. There were 16 signal trials randomly interspersed with 16 no-signal trials, with the constraint that no more than 3 signal or no-signal trials could occur in a row. Trial duration was 4 sec, as was signal duration on signal trials. If listeners turned to the left or pressed the left button during signal presentation, the signal was terminated and a "hit" (response, given a signal) was scored. If the listener did not respond within 4 sec after signal onset, the signal was terminated and a "miss" (no response, given a signal) was scored. If a response occurred within 4 sec after the onset of a no-signal trial, this was scored as a "false-alarm" (response, given no signal). Failure to respond during a no-signal trial was scored

as a correct rejection. Only "hits" were followed by reinforcers, which consisted of the illumination and activation of a mechanical toy for 4 sec. Another trial could begin directly after a correct rejection, if the child was looking directly ahead. Misses and false alarms were followed by a 4-sec intertrial interval.

To ensure that all of the children could perform the task of detecting the signal, a training criterion was employed with sound intensity well above threshold. During training, the intensity of the stimulus was initially 70 dB SPL and only signal trials were presented. Stimulus intensity remained at 70 dB until the participant made four correct responses in a row. Intensity was then reduced by 10 dB, and signal trials continued until the listener again made four successive correct responses. When this criterion was reached, the actual test series followed immediately. The training criterion for adults required only two successive correct responses at each training intensity.

Session 2: 2AFC-L. During Session 2, the curtain covering the right loudspeaker and reinforcement box was removed, and no-signal trials were replaced by right-signal trials. A left head turn (or left-button press) remained the correct response for left-signal presentations, and a right turn (or right-button press) was designated as correct for right-signal presentations. Unlike the G/NG-D procedure, the signals (left or right) remained on until the listener responded. Note that both the G/NG-D and 2AFC-L tasks have two types of trials: signal and no-signal trials in G/NG-D, and signal-on-left and signal-on-right trials in 2AFC-L. Note also that in both tasks, there are two distinct responses: left head turn and no head turn in G/NG-D, and right and left head turn in 2AFC-L. In G/NG-D, only head turns on signal trials were reinforced; not turning on no-signals trials, also a correct response, was not reinforced. Unlike the G/NG-D procedure, both kinds of correct responses (left response given signal on left, right response given signal on right) were reinforced. On successive trials during training, the signal alternated between loudspeakers. Once children had made four successive correct responses, the intensity was reduced to 60 dB and the alternation continued until the child made four successive correct responses at the lower intensity. When this training criterion was met, the test phase began. All aspects of the test phase were identical to those in Session 1 (G/NG-D), except that signal presentation continued until a response occurred and correct responses (right to a right-signal presentation and left to left-signal presentation) were reinforced by a 4-sec presentation of a mechanical toy located near the appropriate loudspeaker. The experimenter was unaware of the purpose of the experiment and, therefore, presumed to be equally unbiased in judging 45° head turns on G/NG-D and 2AFC tasks.

Results and Discussion

To examine potential differences in the separation of the two distributions in the two procedures, d' values were determined for each listener for both the G/NG-D and 2AFC-L tasks. In both cases, d' was defined as $ND[p(R1/T2)] - ND[p(R1/T1)]$, where $p(R1/T2)$ is the probability of Type 1 responses given a Type 2 signal, $p(R1/T1)$ is the probability of a Type 1 response given a Type 1 signal, and ND is the normal deviate for the corresponding probability. The d' transformation permits the assumption that the dependent variable is normally distributed, which might not be the case for percent correct scores. Given the small number of trials in this experiment, it is not unlikely that we would observe a hit rate of 1.0, given the true probability of a hit being less than 1.0. Similarly, it is possible to observe a false-alarm rate of 0 when the true false-alarm probability is greater than

0. We did not consider perfect scores to indicate infinite d' 's but rather to result from sampling fluctuations around a true probability of less than 1.0 for hits and greater than 0 for false alarms (Macmillan & Kaplan, 1985; Thorpe, Trehub, Morrongiello, & Bull, 1988). To circumvent such difficulties, we adjusted the calculations by adding $\frac{1}{2}$ to the number of responses on signal and no-signal trials and 1 to the number of signal as well as no-signal trials (suggested by H. Kaplan, personal communication, April 1986; applied previously by Thorpe et al., 1988; Trainor & Trehub, 1989). This adjustment treats all data uniformly and maintains the rank order of subjects.

Recall that task (single-loudspeaker vs. two-loudspeaker) was a within-subject variable, whereas age and frequency were between-subject variables, making this a split-plot design with one within-subject factor and two between-subject factors. In the analysis of variance corresponding to this design, main effects due to age and frequency and any interactions between age and frequency are not meaningful because of different intensity values used at different ages and frequencies. A similar argument holds with respect to frequency or to any interactions of age with frequency. What is of interest is a main effect due to the within-subject variable, task, as well as any inter-

actions between task and age, task and frequency, and task, frequency, and age. The analysis of variance revealed a main effect of task [$F(1,192) = 56.895, p < .0001$] and an interaction between task and age [$F(3,192) = 3.824, p < .025$]; but there was no significant interaction between frequency and task [$F(1,192) = .056, p > .8$], and no significant three-way interaction [$F(3,192) = .472, p > .7$]. The absence of a task \times frequency interaction indicates that performance differences between tasks were independent of test frequency for all age groups. However, the significant interaction between age and task indicates that performance differences were not independent of age.

Figure 3, which plots mean d' scores for the two tasks at different ages, indicates that there were larger d' scores for 2AFC-L than for G/NG-D for 1-year-olds and adults but not for the other two age groups. This was confirmed by Scheffé tests (Kirk, 1982) on differences in d' scores. Scheffé F values for comparisons of the two methods at the different ages were as follows: $F(1,192) = 58.9, p < .0001$ for 1-year-olds; $F(1,192) = 1.675, p > .1$ for 3-year-olds; $F(1,192) = 0.157, p > .5$ for 5-year-olds; and $F(1,192) = 19.85, p < .0001$ for adults. These results indicate that there were no significant differences

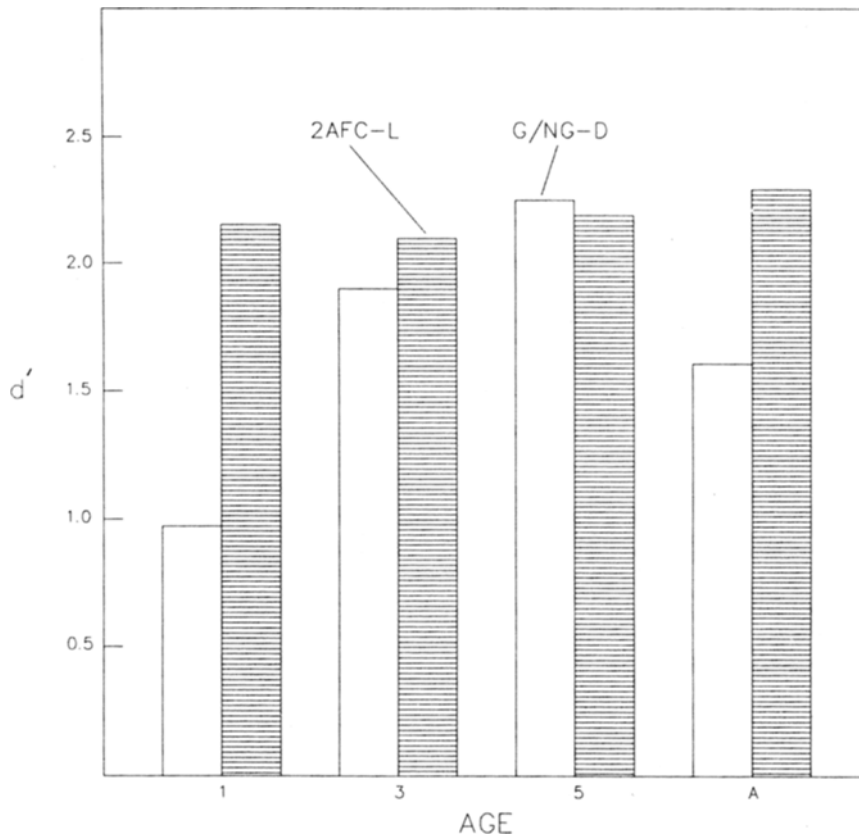


Figure 3. Performance differences in a quiet background, measured in d' units between 2AFC-L (two-alternative forced-choice localization) and G/NG-D (go/no-go detection) tasks at four different ages.

between tasks at 3 and 5 years of age but that 1-year-olds and adults performed better on the 2AFC-L task. When these analyses of variance were repeated on percent correct and arcsine transformed percent correct scores, there were significant task and age \times task interactions, with post hoc analyses showing task differences only for 1-year-olds and adults.

It is possible that the occurrence of 2AFC-L in the second session could have enhanced performance by prior practice on the G/NG-D task. This might have generated superior performance of 1-year-olds on 2AFC-L and might have obscured 3- and 5-year-olds' inferior performance on G/NG-D. To evaluate this possibility, 25 additional 1-year-olds were tested at .4 kHz and 25 additional 3-year-olds at 10 kHz, both with the 2AFC-L task. The mean d' value for these 1-year-olds was 2.20, compared to 2.10 for same-age children tested previously in the second session [$t(48) = .34, p > .25$]. The mean d' value for the 3-year-olds was 2.34, compared to 2.52 for same-age children in the second session [$t(48) = -.56, p > .25$]. This rules out practice or preexposure as the factor underlying the superior performance of 1-year-olds and the equivalent performance of 3-year-olds on the 2AFC-L task.

EXPERIMENT 2

Experiment 1 revealed that performance on 2AFC-L was superior to that on G/NG-D for infants and adults and was not demonstrably worse for 3- and 5-year-olds. Because the 2AFC-L task has been used to determine masked (Bull et al., 1981; Schneider et al., 1989; Trehub et al., 1981) as well as absolute thresholds, it is important to establish whether the same findings prevail for signals presented in noise. Accordingly, we conducted a similar comparison of G/NG-D and 2AFC-L in the presence of continuous background noise.

Under certain conditions, detection in noise is superior to lateralization performance (Egan & Benson, 1966) because of binaural unmasking. Studies of masking level differences (see Durlach & Colburn, 1978, for a review) have revealed that the detectability of a masked signal in one ear can be enhanced by presenting the same masker to the other ear as well. In a typical experiment, the threshold for a pure tone in noise is determined under two conditions. In one, the tone and noise are both presented to the same ear; in the other, the tone is presented to one ear and the noise is presented diotically (i.e., same noise in both ears). It is generally found that the threshold for the monaurally presented tone in a diotic noise background is lower than the threshold for the same tone in a monaural noise background. The difference in threshold between these two conditions is called the masking-level difference (MLD).

In the 2AFC-L masking procedure, noise is presented from both loudspeakers, but the signal is presented from only one. Although localization differs from lateralization, the present experimental context bears some similar-

ity to the second condition of a masking level difference experiment, so that binaural unmasking might occur.² If, in the G/NG-D procedure, the noise were presented only over a single loudspeaker, there would be no opportunity for binaural unmasking. To ensure that binaural unmasking was equivalent in both cases, the masker was presented over both loudspeakers in the G/NG-D and 2AFC-L tasks of the present experiment.

Method

Subjects

Of the 262 subjects recruited for this experiment, 30 were eliminated because of colds, documented hearing losses, the presence of ventilation tubes, or a family history of early hearing loss (8 1-year-olds, 8 3-year-olds, 8 5-year-olds, and 6 adults). In addition, 4 failed to train (1 1-year-old and 3 3-year-olds) and 16 failed to complete the experiment (8 1-year-olds, 5 3-year-olds, and 3 5-year-olds). The 1-year-olds and some of the 3-year-olds were tested on the G/NG-D and 2AFC-L procedures on separate days. Some (12) were unable to return for the second test session, and their data are not included. Older children and adults were tested on both procedures on the same day. The 200 participants in the final sample consisted of four age groups, each with 50 subjects. Half of each age group was tested with the .4-kHz signal and half with the 10-kHz signal. The age groups were 1 year (± 2 weeks), 3 years (± 3 months), 5 years (± 3 months), and young adults (18–25 years).

Apparatus and Stimuli

The apparatus was identical to that described in Schneider et al. (1989). In Session 1, a broadband noise (spectrum level = 0 dB SPL) was presented continuously throughout the session over both loudspeakers. An equalizer was used to ensure that the spectrum of this noise was approximately flat in the frequency range of .25–12.5 kHz. In Session 2, the same broadband masker was used. The intensities of the .4-kHz stimulus were 40, 35, 30, and 22 dB for 1-year-olds, 3-year-olds, 5-year-olds, and adults, respectively. At 10 kHz, the intensities were 45, 45, 40, and 34 dB for 1-year-olds, 3-year-olds, 5-year-olds, and adults, respectively.

Procedure. Methods of testing in both sessions were identical to those of Experiment 1, except that the curtain was not drawn over the right loudspeaker for the G/NG-D task to preclude interference with the background noise from the right loudspeaker. The only other difference between the present experiment and Experiment 1 consisted of the different stimulus levels necessitated by the presence of background noise.

Results and Discussion

Hits and false alarms were adjusted, as in Experiment 1, before determining d' values. As in Experiment 1, main effects due to age and frequency and their two-way interaction were considered irrelevant. An analysis of variance revealed a main effect of task [$F(1,192) = 7.837, p < .01$] and an interaction between task and age [$F(3,192) = 8.930, p < .0001$], but there was no significant interaction between frequency and task [$F(1,192) = 0.444, p > .5$], and there was no significant three-way interaction [$F(3,192) = 0.726, p > .7$]. The absence of any interaction between task and frequency indicates that performance differences between tasks were independent of test frequency for all age groups. However, the significant interaction between age and task indicated that,

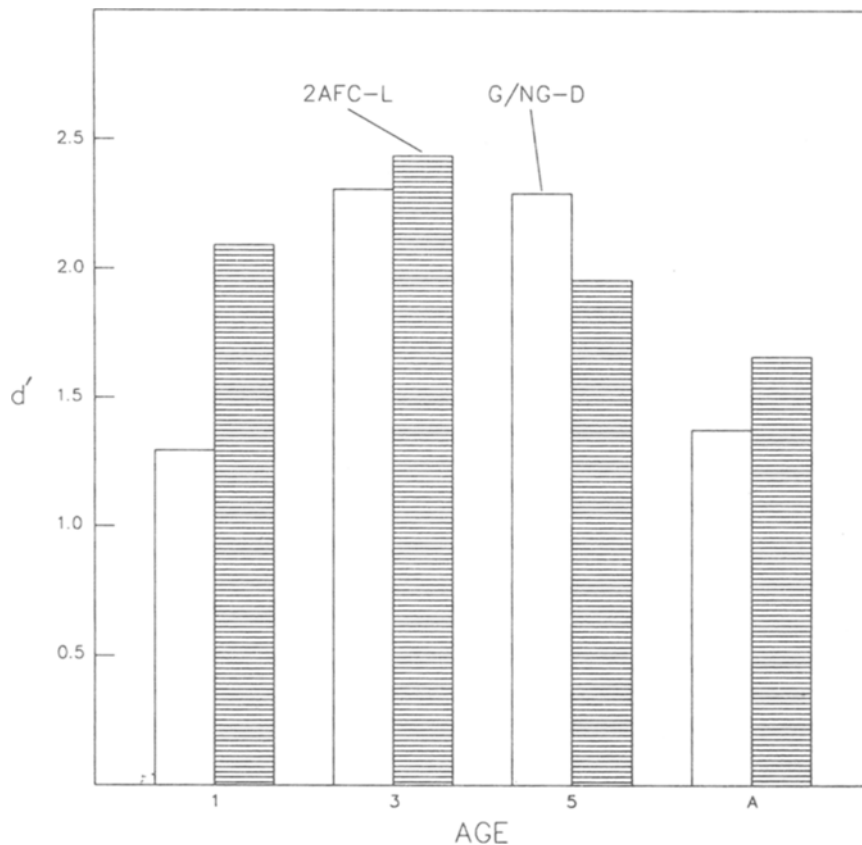


Figure 4. Performance differences in a noisy background, measured in d' units between 2AFC-L (two-alternative forced-choice localization) and G/NG-D (go/no-go detection) tasks at four different ages.

as in Experiment 1, performance differences were not independent of age. Figure 4, which plots mean d' scores for the two tasks as a function of age, indicates that, as in Experiment 1, d' scores for the 2AFC-L task were larger than those for the G/NG-D task for 1-year-olds. Note also that, although adult performance on 2AFC-L was better than on G/NG-D, this difference was not as large as before. Finally, 5-year-olds appeared to perform marginally better on G/NG-D than on 2AFC-L. These impressions were confirmed by Scheffé tests of the difference in d' scores at each age. Scheffé F values for a comparison of the two methods at the different ages were as follows: $F(1,192) = 26.076$, $p < .0001$, for 1-year-olds; $F(1,192) = 0.683$, $p > .25$, for 3-year-olds; $F(1,192) = 4.607$, $p < .05$, for 5-year-olds; and $F(1,192) = 3.288$, $p > .05$, for adults. When these analyses of variance were repeated on percent correct scores and arcsine transformed percent correct scores, there were significant task and age \times task interactions, with post hoc analyses showing task differences only for 1-year-olds. Thus, the task difference at 5 years of age can be regarded as marginally significant.

GENERAL DISCUSSION

The results of Experiments 1 and 2 provide no evidence that performance is better on detection as opposed to localization tasks for .4- and 10-kHz octave-band signals presented in quiet or in masking noise. On the contrary, where performance differences do exist, as is the case for infants and adults, these differences favor the 2AFC-L task.

The fact that the 2AFC-L procedure yields superior performance for infants may seem surprising in light of the lower thresholds that have been reported in G/NG-D studies (e.g., Berg & Smith, 1983; Wilson & Moore, 1978) as opposed to 2AFC-L studies (e.g., Schneider et al., 1980; Trehub et al., 1980; Trehub et al., 1989). These lower thresholds have often been interpreted as indicating the superiority of G/NG-D compared to 2AFC-L. For example, Wilson and Moore (1978) reported warbled-tone thresholds in the sound field that were as much as 7-8 dB lower than 2AFC-L thresholds for same-age listeners (about 1 year of age). Berg and Smith's (1983) pure-tone thresholds in the sound field were also lower

than 2AFC-L thresholds although Sinnott et al.'s (1983) pure-tone thresholds were approximately equivalent to 2AFC-L thresholds. Because these G/NG-D studies differed from the 2AFC-L studies in stimulus (tones vs. octave-band noises) as well as task, threshold differences may be due, in part, to stimulus differences.

Dillon and Walker (1982) have indicated the difficulty of specifying sound pressure levels of pure tones in the sound field (as opposed to warbled tones or band-limited noises). To some extent, then, reported differences in measured threshold may be attributable to differences between actual tonal intensities during testing and those measured in the infant's absence. Because the sound pressure level of octave-band noises in the sound field varies less with location than is the case for pure tones, specification of sound intensity would be less problematic. Furthermore, Robinson and Whittle (1964) have found lower thresholds for pure tones than for octave-band noises for frequencies of up to .4 kHz. It is possible, then, that threshold differences between G/NG-D and 2AFC-L studies stem from some combination of factors associated with measurement and signal differences.

A second possibility is that lower thresholds in G/NG-D studies may result from the practice of ignoring false-alarm rates in the computation of threshold.³ Signal detection theorists (Egan & Clarke, 1966; Green & Swets, 1974) have long argued that, in single-interval experiments such as G/NG-D, measures of sensitivity that ignore the false-alarm rate rely excessively on the observer's criterion (see Egan & Clarke, 1966, pp. 236-240). Wilson and Moore (1978), Berg and Smith (1983), and Sinnott et al. (1983) used adaptive procedures and calculated thresholds from hits alone, despite the fact that their false-alarm rates differed. For example, false-alarm rates were considerably lower in Sinnott et al. than in Berg and Smith (1983), possibly because of a 15-sec intertrial interval following false alarms in Sinnott et al. (1983) and their exclusion of sessions with more than one false alarm. By contrast, false alarms in Berg and Smith (1983) did not affect intertrial intervals, and sessions were excluded only if false-alarm rates exceeded 33%. If d' values are computed at the threshold intensities reported in these two studies,⁴ the relevant d' value in Berg and Smith (sound-field conditions, 10- and 14-month-old infants) is 1.69 compared to 2.16 in Sinnott et al. (1983) (1-sec tones). Thus, the higher pure-tone thresholds reported by Sinnott et al. may simply reflect their more stringent criterion for threshold.

Because threshold differences in previous studies could arise from different stimuli (pure tones vs. octave-band noises), false-alarm rates, or criteria for data retention, comparisons of these and other studies with different test conditions and data treatment are fraught with potential difficulty. On the other hand, the present study indicates that when similar stimuli are used with the same number of Type 1 and Type 2 trials, infants perform better on the 2AFC-L task but young children perform equivalently on

the two tasks (with the possible exception that 5-year-olds perform slightly better on G/NG-D in noise).

One possible reason for superior infant performance on the 2AFC-L task is that the stimulus remained on until a response occurred. In the present G/NG-D task, stimulus duration (and the corresponding response interval) was limited to 4 sec. If infants require more time to reach a decision, we might expect better performance on the 2AFC-L task because of its unlimited response interval. Indeed, comparisons of 2AFC-L tasks with limited or unlimited response durations have revealed superior performance of 1-year-olds in the unlimited response case (Trehub et al., 1981). Because the response interval was limited in the G/NG-D task, a long decision or reaction time would have been scored as a miss, leading to underestimates of infants' ability to detect the signal. However, detailed examination of data from the 2AFC-L case revealed that superior performance was not entirely due to the flexible response interval. Despite the fact that infants tested on 2AFC-L had received reinforcement on correct trials with long decision times, one can nevertheless explore the implications of excluding such trials from the data set, or treating them as errors or misses. Simply excluding trials in Experiment 1 with a response time greater than that permitted in the G/NG-D task (4 sec) results in a change in proportion correct for the .4-kHz stimulus from .83 to .87 compared to .65 on G/NG-D; at 10 kHz, the proportion correct changes from .84 to .88 compared to .67 on G/NG-D. Therefore, applying such a response time criterion leads to the exclusion of 12.9% of trials and to a slight improvement in performance, which implies that misses occur more frequently with long- than with short-latency responses. Alternatively, if one treats such long-latency responses as misses, thereby applying a penalty on 2AFC-L compared to G/NG-D tasks, where appropriate training and feedback were provided, the following picture emerges. The proportion correct on 2AFC-L still exceeds that on G/NG-D, .75 compared to .65 for .4 kHz and .78 compared to .67 for 10 kHz. Clearly, then, there is more to the differential success of the tasks than duration of the response interval, although this factor may contribute to the difference.

A second possible factor underlying superior infant performance on 2AFC-L concerns the differential cognitive demands of 2AFC-L compared to G/NG-D. In the G/NG-D case, infants must learn not to respond on no-signal trials. Although verbal instruction in this regard is successful with older children, infants, for whom this is obviously precluded, require visual entertainment or distraction away from the loudspeaker and visual reinforcer. As noted earlier, the payoff for remaining at or returning to midline is lower in G/NG-D than in 2AFC-L, whereas signals and possible reinforcement are available with greater frequency in 2AFC-L. Thus, more entertainment may be required in G/NG-D to get the infant to return to midline. Therefore, G/NG-D, as typically applied with infants, might function as a divided-attention

task, potentially degrading performance. Although adults can perform well in simultaneous visual and auditory discriminations (Shiffrin & Grantham, 1974), infants may be less capable of doing so.

The finding of superior adult performance on the 2AFC-L task at both frequencies in quiet but not in noise is more problematic. Because 3- and 5-year-old children were not adversely affected by the greater cognitive demands of the G/NG-D task, it is unlikely that adults would be. Rather, adults may capitalize on sophisticated strategies that are marginally more effective in the 2AFC-L situation.

One example of such a strategy has been noted anecdotally by several adults tested on 2AFC-L tasks in the present as well as previous studies. They report discovering that one ear is more sensitive than the other, in that they become aware of more frequent errors on one side. As a result, they adopt a strategy of responding on the side of the good ear if they hear the signal there, the bad ear if they hear it there, and the bad ear again if they do not hear the signal but must nevertheless respond. If, for example, their left ear is less sensitive than their right, then on trials with nondetected signals, they presume the

signal to be emanating from the left. If such anecdotal reports are at all accurate, some proportion of adults may notice such correlations and behave appropriately, in contrast with children or less sophisticated adults, who respond randomly when the signal cannot be heard or localized. This strategy enhances performance on 2AFC-L but not on G/NG-D. In any case, effective decision strategies such as this might underlie differential adult performance on 2AFC-L and G/NG-D. It is important that this weak-ear strategy would be more effective in quiet than in noise, because ear asymmetries tend to be greater in quiet than in noise. Thus, such a strategy would generate a clear 2AFC-L advantage in quiet but not necessarily in noise.

It is also possible that adults as opposed to children map the stimuli onto the decision axis in different ways. As noted in our introduction, the factors contributing to the decision axis may differ in the two tasks, and, therefore, there is no a priori reason why 2AFC-L and G/NG-D should yield equivalent d' 's. If adults are more successful at mapping the additional information available in 2AFC-L compared to G/NG-D tasks, this might account for their

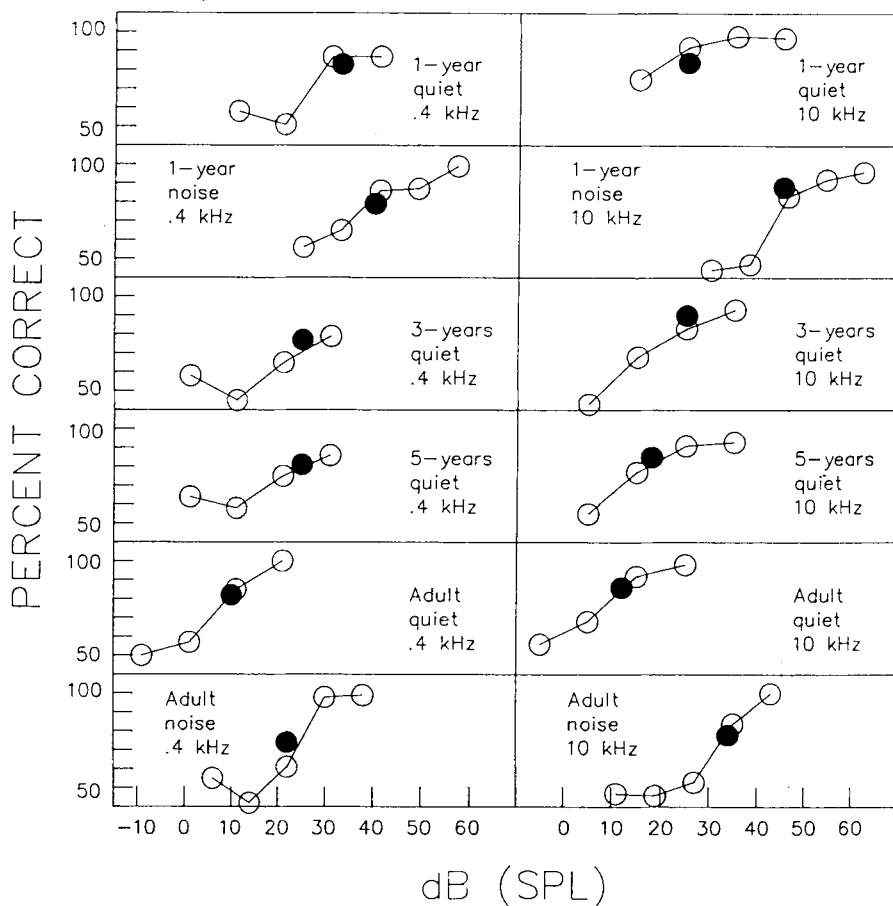


Figure 5. Percent of correct head turns in 2AFC-L (two-alternative forced-choice localization) as a function of sound pressure level for infants, children, and adults in quiet and in noise. Open circles are data from Schneider, Trehub, Morrongoello, and Thorpe (1986, 1989) and Trehub, Schneider, and Endman (1980). Filled circles are data from the present experiments.

superior performance in 2AFC-L. However, this factor fails to explain 2AFC-L superiority in quiet but not in noise.

In the present version of the 2AFC-L task, only a single intensity was presented in a session. In previous research with this technique, however, we presented four or five intensities at one frequency or several different frequencies during a session. To determine whether the number of frequencies and/or sound pressure levels affects performance in a test session, we plotted the average percent correct in the present experiment (filled circles) with psychometric functions obtained previously (Schneider et al., 1986, 1989; Trehub et al., 1980) with infants, children, and adults tested at the same frequencies and ages in quiet and in noise (see Figure 5). The data from the present study are compatible with psychometric functions obtained in previous experiments, suggesting that the number of frequencies and intensity levels in the 2AFC-L task have little effect on performance.

It is encouraging that the two different methods, when appropriately compared, give nearly equivalent results and that there is no interaction between stimulus frequency and task. Infants' superior performance of the 2AFC-L task, however, makes a strong case for its use with this age group in the determination of absolute and masked thresholds.

REFERENCES

- ASLIN, R. N., PISONI, D. B., & JUSCZYK, P. W. (1983). Auditory development and speech perception in infancy. In P. H. Mussen (Ed.), *Handbook of child psychology: Vol. II. Infancy and developmental psychobiology* (pp. 573-687). New York: Wiley.
- BERG, K. M., & SMITH, M. (1983). Behavioral thresholds for tones during infancy. *Journal of Experimental Child Psychology*, **35**, 409-425.
- BLOUGH, D. S., & BLOUGH, P. (1977). Animal psychophysics. In W. K. Honig & J. E. R. Staddon (Eds.), *Handbook of operant behavior* (pp. 514-539). Englewood Cliffs, NJ: Prentice-Hall.
- BULL, D., SCHNEIDER, B. A., & TREHUB, S. E. (1981). The masking of octave-band noise by broad-spectrum noise: A comparison of infant and adult thresholds. *Perception & Psychophysics*, **30**, 101-106.
- CHUN, R. W. M., PAWSAT, R., & FORSTER, F. M. (1960). Sound localization in infancy. *Journal of Nervous & Mental Disease*, **130**, 472-476.
- CLIFTON, R. K., MORRONGIELLO, B. A., KULIG, J. W., & DOWD, J. W. (1981). Developmental changes in auditory localization in infancy. In R. N. Aslin, J. R. Alberts, & M. R. Peterson (Eds.), *Development of perception: Psychobiological perspectives. Vol. I. Audition, somatic perception and the chemical senses* (pp. 141-160). New York: Academic Press.
- DILLON, H., & WALKER, G. (1982). Comparison of stimuli used in sound field audiometric testing. *Journal of the Acoustical Society of America*, **71**, 161-172.
- DURLACH, N. I., & COLBURN, H. S. (1978). Binaural phenomena. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception: Vol. IV. Hearing* (pp. 365-466). New York: Academic Press.
- EGAN, J. P., & BENSON, W. (1966). Lateralization of a weak signal presented with correlated and with uncorrelated noise. *Journal of the Acoustical Society of America*, **40**, 20-26.
- EGAN, J. P., & CLARKE, F. R. (1966). Psychophysics and signal detection. In J. P. Sidowski (Ed.), *Experimental methods and instrumentation in psychology* (pp. 211-246). New York: McGraw-Hill.
- GREEN, D. M., & SWETS, J. A. (1974). *Signal detection theory and psychophysics*. Huntington, NY: Krieger.
- KIRK, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Monterey, CA: Brooks/Cole.
- MACMILLAN, N. A., & KAPLAN, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, **98**, 185-199.
- MOORE, J. M., THOMPSON, G., & THOMPSON, M. (1975). Auditory localization of infants as a function of reinforcement conditions. *Journal of Speech & Hearing Disorders*, **40**, 29-34.
- NOZZA, R. J. (1987). The binaural masking level difference in infants and adults: Developmental changes in binaural hearing. *Infant Behavior & Development*, **10**, 105-110.
- 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 & Hearing Research*, **27**, 613-622.
- OLSHO, L. W. (1985). Infant auditory perception: Tonal masking. *Infant Behavior & Development*, **8**, 371-384.
- ROBINSON, D. W., & WHITTLE, L. S. (1964). The loudness of octave bands of noise. *Acustica*, **14**, 24-35.
- SCHNEIDER, B. A., BULL, D., & TREHUB, S. E. (1988). Binaural unmasking in infants. *Journal of the Acoustical Society of America*, **83**, 1124-1132.
- SCHNEIDER, B. A., MORRONGIELLO, B. A., & TREHUB, S. E. (1990). Size of critical band in infants, children, and adults. *Journal of Experimental Psychology: Human Perception & Performance*, **16**, 642-652.
- SCHNEIDER, B. A., TREHUB, S. E., & BULL, D. (1980). High-frequency sensitivity in infants. *Science*, **207**, 1003-1004.
- SCHNEIDER, B. A., TREHUB, S. E., MORRONGIELLO, B. A., & THORPE, L. A. (1986). Auditory sensitivity in preschool children. *Journal of the Acoustical Society of America*, **79**, 447-452.
- SCHNEIDER, B. A., TREHUB, S. E., MORRONGIELLO, B. A., & THORPE, L. A. (1989). Developmental changes in masked thresholds. *Journal of the Acoustical Society of America*, **86**, 1733-1742.
- SHIFFRIN, R. M., & GRANTHAM, D. W. (1974). Can attention be allocated to sensory modalities? *Perception & Psychophysics*, **15**, 460-474.
- SINNOTT, J. M., PISONI, D., & ASLIN, R. (1983). A comparison of pure tone auditory thresholds in human infants and adults. *Infant Behavior & Development*, **6**, 3-18.
- SOKOLOV, E. N. (1963). *Perception and the conditioned reflex*. New York: Macmillan.
- SUZUKI, T., & OGIBA, Y. (1961). Conditioned orientation reflex audiometry. *Archives of Otolaryngology*, **74**, 192-198.
- THORPE, L. A., TREHUB, S. E., MORRONGIELLO, B. A., & BULL, D. (1988). Perceptual grouping by infants and preschool children. *Developmental Psychology*, **24**, 484-491.
- TRAINOR, L. J., & TREHUB, S. E. (1989). Aging and auditory temporal sequencing: Ordering the elements of repeating tone patterns. *Perception & Psychophysics*, **45**, 417-426.
- TREHUB, S. E., BULL, D., & SCHNEIDER, B. A. (1981). Infants' detection of speech in noise. *Journal of Speech & Hearing Research*, **24**, 202-206.
- TREHUB, S. E., BULL, D., SCHNEIDER, B. A., & MORRONGIELLO, B. A. (1986). PESTI: A procedure for estimating individual thresholds in infant listeners. *Infant Behavior & Development*, **9**, 107-117.
- TREHUB, S. E., SCHNEIDER, B. A., & BULL, D. (1981). Effect of reinforcement on infants' performance in an auditory detection task. *Developmental Psychology*, **17**, 872-877.
- TREHUB, S. E., SCHNEIDER, B. A., & ENDMAN, M. (1980). Developmental changes in infants' sensitivity to octave-band noises. *Journal of Experimental Child Psychology*, **29**, 282-293.
- TREHUB, S. E., SCHNEIDER, B. A., MORRONGIELLO, B. A., & THORPE, L. A. (1988). Auditory sensitivity in school-age children. *Journal of Experimental Child Psychology*, **46**, 273-285.
- TREHUB, S. E., SCHNEIDER, B. A., MORRONGIELLO, B. A., & THORPE, L. A. (1989). Developmental changes in high-frequency sensitivity. *Audiology*, **28**, 241-249.
- WILSON, W. R., & MOORE, J. M. (1978, November). *Pure-tone earphone thresholds of infants utilizing visual reinforcement audiometry (VRA)*. Paper presented at the American Speech and Hearing Association Convention, San Francisco.

NOTES

1. This single-interval 2AFC task should not be confused with a two-interval 2AFC task, where the subject identifies which interval the stimulus appears in. Therefore, relationships between d' in 2I-2AFC and yes/no detection tasks do not apply here.

2. Even though the stimulus conditions in G/NG-D are the sound-field equivalent of an earphone experiment in which both signal and noise are monaural, and the stimulus conditions in 2AFC-L are the sound-field equivalent of a monaural signal with diotic noise, the interaural differences are by no means equivalent in earphone and sound-field cases, because of interactions among the left and right sounds in the sound

field (see Schneider et al., 1988). As a result, it is difficult to predict the outcome of the sound-field case from the earphone case.

3. When the G/NG task is used in discrimination as opposed to detection studies, a single stimulus is typically presented over a block of trials. Because adaptive methods are not used, there is no problem in estimating difference thresholds with this technique. Furthermore, there is no easy way to adapt the 2AFC-L task to discrimination studies. In discrimination studies, then, G/NG remains the method of choice.

4. In computing the d' values, we assumed that the false-alarm rate in these experiments did not vary with signal level during the session. This assumption is questionable, given that false-alarm rate often varies with signal level.

(Manuscript received November 6, 1989;
revision accepted for publication July 9, 1990.)