Development of absolute thresholds in chickens

Lincoln Gray[a] and Edwin W Rubel

Department of Otolaryngology, University of Virginia Medical School, Charlottesville, Virginia 22908

Absolute auditory thresholds were estimated in chickens at 0 and 4 days after hatching. Momentary suppressions of the chicks' regular peeping, following the onset of a tone, were used as indications of stimulus detection. In the first experiment a staircase procedure was used to estimate thresholds. The absolute thresholds of both ages were the same at low frequencies (250–500 Hz), but at higher frequencies (1–2 kHz) 4-day-old chicks had lower thresholds than the 0-day-old chicks. The estimates of thresholds at 1 kHz were corroborated in the second experiment with a method of constant stimuli. A more efficient modified method of limits was used to replicate the age by frequency interaction in the third experiment. These changing thresholds are likely to reflect a developmental process somewhere in the auditory system and not some nonsensory artifact for two reasons: (a) similar thresholds at low frequencies show that developmental differences are not due to differences in the sensitivity of the testing procedure at the two ages and (b) thresholds obtained from the 4-day-old birds are similar to estimates from mature birds. In conclusion, responsiveness to low frequencies develops before responsiveness to higher frequencies, showing that the development of absolute thresholds is correlated with other measures of functional maturation in the auditory system.

PACS numbers: 43.66.Gf, 43.66.Yw, 43.66.Cb [RDS]

INTRODUCTION

Psychophysical studies in young animals endeavor to track the early development of sensory function in order to discover general principles of perceptual development. Such studies must begin with a thorough understanding of how basic perceptual abilities, such as auditory thresholds, develop in normal neonates (Gottlieb, 1976).

Young chickens provide useful subjects for the study of auditory development. The chicken's auditory system is relatively mature at birth, though still developing in both structure and function (Gray and Rubel, 1981; Rubel, 1978; Smith, 1981). Furthermore, an unconditioned response in young chicks is particularly well suited for psychophysical studies. Young chicks, when in a well-lit, room-temperature (approximately 20°–25° C) enclosure, will peep nearly incessantly (Bermant, 1963). When an acoustic stimulus is presented, the birds will orient to that tone and momentarily suppress their ongoing series of peeps (Kerr et al., 1979). Longer suppressions are elicited by louder tones (Gray and Rubel, 1981). Such unconditioned responses that vary monotonically with stimulus intensity are often ideal for studies of perceptual development (Blough and Blough, 1977; Moody, 1970).

In this study three psychophysical procedures— a two-interval staircase, a method of constant stimuli, and a modified method of limits—were used to estimate auditory thresholds in two ages of young birds. Each procedure used peep suppressions to estimate thresholds in a slightly different way. Similarities among the results from these different procedures reveal general trends in the neonatal auditory development of this species; all of the experiments indicate that sensitivity to high frequencies develops after sensitivity to low frequencies. Differences among the procedures show some of the compromises involved in studying the thresholds of neonates: The two-interval staircase procedure is well controlled, but young animals habituate to repeated stimulus presentations; there is no opportunity for habituation in the method of constant stimuli, but the procedure is inefficient; the modified method of limits is efficient, but results at each frequency may be variable because of the limited number of trials; finally, the starting levels in any test can affect the validity of estimated thresholds.

I. EXPERIMENT 1: A TWO-INTERVAL STAIRCASE

The general strategy was to use a transformed up–down rule or staircase procedure (Wetherill and Levitt, 1965) to find the intensities of various tones which elicit suppressions that are just greater than those expected when there is no tone. A comparison of these estimated thresholds between two ages of young chickens is the crux of the analysis.

A. Methods

Many methodological details, such as incubating and brooding procedures, calibration of switching transients, and model numbers of equipment can be found in Gray and Rubel (1981).

1. Subjects

Subjects were 120 domestic chicks (Hubbard × Hubbard strain), normally hatched and brooded in our laboratory. Animals were tested between 6 and 24 h after hatching, termed "0-day-old chicks" (called 1-day-old in previous papers) or between 96 and 120 h after hatching, termed "4-day-old." The testing of 12 additional subjects was terminated, as discussed below, before useful data were collected.

---

[a] Present address: Department of Otolaryngology, University of Texas Medical School, Houston, TX 77025.

2. Apparatus

Subjects were individually tested in a small (1.25 l) cylindrical chamber inside a sound-attenuating room. Sound delivered through a speaker below the chamber varied no more than 6 dB (± 0.5 dB at frequencies lower than 2 kHz) around the chamber at the level of animals' ears.

The stimuli were pulsing pure tones (2 per s: 460 ms on, 40 ms off, with 32-ms linear rise and decay times, meaning 428 ms fully on and 8 ms fully off). Six frequencies were used, spaced at even octaves over the limits of chickens' hearing (Kerr *et al.*, 1979): 125, 250, 500, 1000, 2000, and 4000 Hz.

A microphone suspended above the chamber provided the input to a specially built peep discriminator set to trigger at the end of the bird's loud peeps or "distress calls" (Andrew, 1975; Collias and Joos, 1953). This circuit has been shown to agree closely with an experienced human observer in deciding what constitutes a peep and when the peep occurred. Ninety-five percent of discrepancies between human and electronic estimates of supression times were less than 160 ms (Severns *et al.*, in press, analyzed over 4000 peeps from chicks of both ages).

Stimulus production and response measurement were under the control of a computer (PDP 11–04). The computer also recorded the stimuli and responses from each trial as described below.

3. Response measurement

The dependent variable was the duration of peep suppression. Suppression was measured, to the nearest 60th of a second, as the length of time from shortly after the onset of a trial until the next peep. The timing of suppression did not start until 500 ms after the onset of a trial. As discussed below, this half-second of adjustment gives the chicks some time to react to the stimulus before the response is measured and minimizes the chances that the subjects' vocalizations will mask the tones.

Responses of young animals often depend on the "state" of the subject (Hutt *et al.*, 1968; Kearsley, 1973). Thus our trials were begun only when animals were peeping regularly. The computer started trials only after the subject had peeped twice in 2 s (as in Gray and Rubel, 1981).

Data were collected in the same way during stimulus trials and control trials (explained below). As soon as the bird peeped twice in 2 s, a stimulus was (or was not) turned on. Starting after 500 ms, the time until the next peep was recorded.

4. Procedure

An adaptive procedure was used to estimate auditory thresholds. That is, the intensities of the stimuli were varied according to the subject's previous responses.

Two suppression times from a pair of trials were compared to form a "response." One "stimulus" trial and one "control" trial were presented in random order within each pair. On stimulus trials a tone was presented and on control trials no tone was presented. Each response, therefore, consisted of one suppression on a stimulus trial and another suppression on a control trial. If the suppression time on the stimulus trial was greater than that on the control trial, the

response was considered correct. If the suppression time on the stimulus trial was less than or equal to that on the control trial, the response was considered incorrect.

A 2 down–1 up staircase rule was used to change intensities, as shown in Fig. 1. After each incorrect response, the intensity of the stimulus was increased. This increase in intensity constitutes part of an up run. The intensity of the stimulus was decreased after two successive correct responses. This decrease in intensity constitutes part of a down run. Changes of direction in the staircase, or "peaks," occur at stimulus levels where the subject's responses cause up runs to change to down runs or vice versa. An advantage of the double-interval staircase procedure is that each stimulus trial has its own control. This should correct for any differences in baseline peep rates over ages or trials.

Staircases were started at 25 dB SPL at all frequencies. This starting level was at the lowest threshold estimated from pilot data. Initially soft stimuli were used because suppressions to loud tones are known to habituate (Gray and Rubel, 1981; Kerr *et al.*, 1979).

The step size was decreased after each change of direction: from 8–6–4–3 dB. That is, ascending intensities were increased by 8 dB after every incorrect response until the
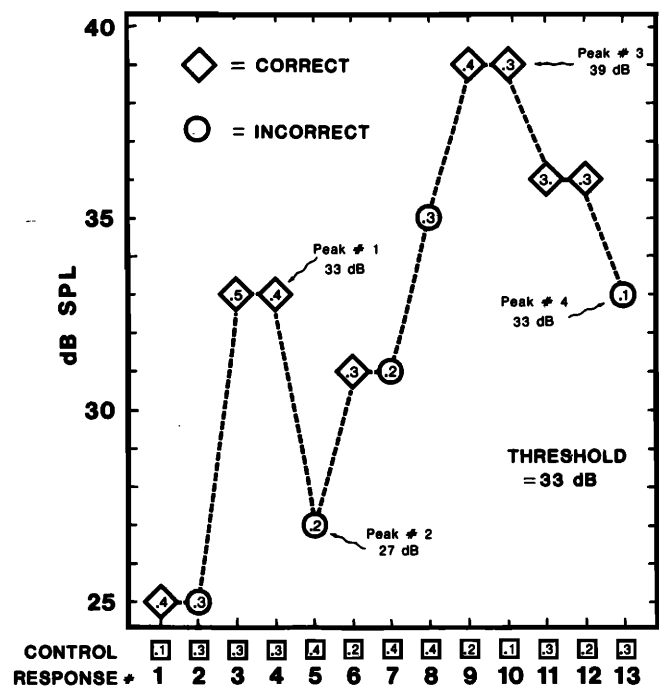

FIG. 1. An example of the two-interval staircase procedure. An adaptive tracking of absolute threshold is shown. These are data from a 0-day-old bird in experiment 1 tested at 250 Hz. The intensity of the stimulus is graphed in dB (SPL) over enough responses to reach the fourth peak. Each response is derived from two suppression times (shown in seconds): one each from a stimulus and a control trial. Suppression times on stimulus trials are shown in the body of the graph inside diamonds or circles; suppressions on control trials are shown along the bottom inside squares. Correct responses, shown by diamonds, occur only when suppression on the stimulus trial is longer than suppression on the control trial; otherwise responses are incorrect and are shown by circles. Intensity is increased after every incorrect response and decreased after two successive correct responses. Peaks occur where these responses cause increasing intensities to decrease and vice versa. Changes in intensity decrease after each peak, from + 8 to − 6 to + 4 to − 3 dB. Threshold is calculated from the mean of the first four peaks.

bird got two successive responses correct. This intensity defined the first peak and started the first down run. If the first two responses were correct, then the starting level was used as the first peak. Intensities were then decreased by 6 dB until the first incorrect response. This intensity defined the second peak and ended the first down run. Intensities were then increased by 4 dB until the third peak, then subsequently changed by 3 dB until the end of the test. Decreasing step sizes after each change of direction should provide both efficient and precise estimates of thresholds, as the algorithm progressively converges on the correct value.

Thresholds were calculated as the average of the middle of the first two down runs; that is, the mean of peaks 1–4. Data were thus retained for analysis, only if the subject completed four peaks in the staircase. The mean of midpoints from several down runs is thought to be the most robust and unbiased estimate of threshold (Levitt, 1971), and at least four changes of direction are required to calculate such a mean.

Testing began after 2–5 min of acclimation to the chamber and as soon as the birds were peeping regularly. Testing terminated after seven changes of direction or 34 trials, whichever came first. Testing was also terminated at any time after 2 min without a peep from the subject. At least 7.5 s elapsed between all trials. Tones were turned off shortly after the peep that ended the trial (a random period between 0.5–2.5 s).

There were ten animals in each of 12 groups. Two ages of subjects (0- and 4-day-old) were tested at six frequencies (125, 250, 500, 1000, 2000, and 4000 Hz).

5. Neutral model

A Monte-Carlo simulation was used to determine what to expect from random behavior. The computerized proce-
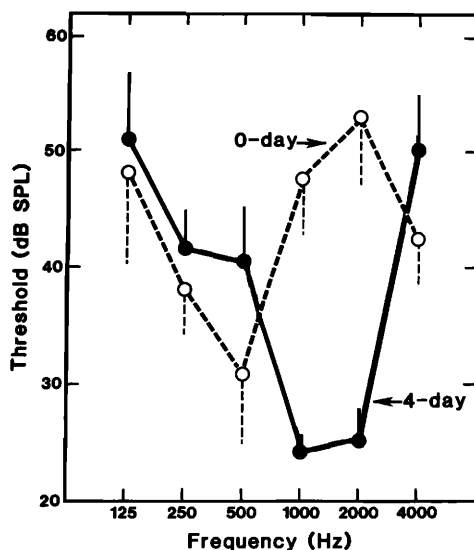

FIG. 2. Absolute thresholds in two ages of young chicks. Shown are means and standard errors of estimated thresholds for 10 birds each at two ages and six frequencies. The estimated threshold from each bird was the mean of the first four peaks in the staircase. Random expectation derived from a neutral model is 47.5 dB, indicating that the thresholds at 125 and 4000 Hz are not meaningful for comparison between the ages (as discussed in the text).

dure was similar to the one used to test chickens with the critical exception that suppressions on all trials were random numbers. Staircases were started at 25 dB and run for seven changes of direction or 34 responses. Correct responses occurred half the time, when the random number assigned to the "stimulus" trial was greater than the random number assigned to the "control" trial. Two successive correct responses decreased the intensity of the modeled stimulus. Otherwise, the intensity was increased. Step sizes were decreased from 8–6–4–3 dB after each peak. The mean of the first four peaks was recorded as one random estimate of threshold. The mean of 32 767 such estimates was used as random expectation.

B. Results

1. Absolute thresholds

Estimates of absolute thresholds are shown in Fig. 2 for each of 12 groups: two ages and six frequencies.

T tests showed significant differences between the two ages at 1 and 2 kHz [$4.3 < t(18) < 4.5$, $p < 0.0003$]; the ages were not different at any other frequency [$- 1.3 < t(18) < 0$, $p > 0.2$]. Therefore, 0- and 4-day-old chicks appear to have equivalent thresholds at 250 and 500 Hz, but they have different thresholds at 1 and 2 kHz. Comparisons at 125 and 4000 Hz are not meaningful as discussed below.

A two-way, cross-classified analysis of variance confirmed these trends. There was a highly significant age by frequency interaction [$F(5, 108) = 5.5$, $p < 0.0002$], no significant main effect of age [$F(1, 108) = 2.6$, $p > 0.1$], and a marginally significant effect of frequency [$F(5, 108) = 2.6$, $p < 0.05$].

2. Comparison with random expectation

Random expectation under these conditions was 47.5 dB, as derived from a neutral model. The probabilities that a randomly generated threshold would be as low as the estimated thresholds were evaluated with one-tailed t tests. For 0-day-old chicks these probabilities were 0.47, 0.02, 0.01, 0.48, 0.31, and 0.12 at 125 to 4000 Hz, respectively. For 4-day-old chicks these probabilities were 0.32, 0.052, 0.085, 0.0001, 0.0001, and 0.30 at 125 to 4000 Hz, respectively. That is, both ages gave thresholds no different from random behavior at 125 and 4000 Hz. When thresholds of *both* ages are random ($p > 0.05$), the data are useless for comparison because the data underestimate the real thresholds by an unknown amount. It is therefore not appropriate to compare thresholds between the two ages at 125 and 4000 Hz in these data. Frequencies such as 250, 500, 1000, and 2000 Hz where at least one group of subjects gave thresholds that were significantly different from random expectation ($p < 0.05$) are, however, valid for comparison; we know the ontogenetic trends are at least as large as those indicated.

3. Derived psychometric function

The psychometric function depicts performance over a fairly wide range of stimuli. This is usually obtained from a method of constant stimuli as in experiment 2, where many responses are observed to preselected stimuli (Kling and

Riggs, 1971; Martin et al., 1980). Although the staircase procedure concentrates stimuli at intensities close to the estimated threshold, the responses from all of the stimulus trials can be reorganized to show suppressions as a function of intensity.

Figure 3 shows average suppressions as a function of stimulus intensity, considering only the stimulus trial of each pair. The duration of suppression in seconds is plotted in reference to the estimated threshold of each age and frequency combination (dB relative to threshold). The average suppression over all control trials is shown at the far left. Notice first in Fig. 3 that suppressions increase with increasing intensities, demonstrating that the response was under stimulus control. Figure 3 also indicates that the estimated threshold (0 dB relative to threshold) was the lowest intensity where suppression times were significantly greater than those on control trials. This shows that the midpoint of the first two down runs occurs at an intensity which elicits suppressions that are just significantly longer than baseline.

4. The problem of habituation

Young chicks habituate to acoustic stimulation (Gray and Rubel, 1981; Kerr et al., 1979; Rubel and Rosenthal, 1975), as do young mammals (Haroutunian and Campbell, 1981). Suppressions decrease after repeated presentations of the same stimulus; thus psychometric functions in these young subjects probably shift toward the right over time as estimates of thresholds increase. The staircase procedure can, however, be used to track changes in threshold due to habituation over trials. Figure 4 shows the average intensity of all stimuli at each peak of the staircases in an attempt to depict the process of habituation. Intensities clearly rise over peaks, showing that progressively louder tones are required to elicit reliable suppression.

FIG. 4. The effect of habituation over trials. The mean intensities of the stimuli at each peak of the staircase in experiment 1 are shown.

A multiple regression of suppression times with trials and intensities can evaluate the ability of the staircase to successfully track habituation. A reliable decrease in suppression times over trials would indicate that the algorithm failed to increase intensities sufficiently to counteract for this habituation. There was no significant effect of trials [$F(1,3087) = 0.45, p > 0.5$], indicating that the adaptive procedure kept up with habituation. There was a strong linear effect of intensity on suppression times [$F(1,3087) = 23$, $p < 2 \times 10^{-6}$], indicating again that the response is related to the stimulus.

One problem in this developmental study was to include enough trials to get an accurate estimate of threshold, yet few enough trials to avoid increasing the estimate due to habituation. Although it appears from Fig. 4 that the 0-day-old chicks habituated more than the 4-day-old chicks, the amount of habituation depends on frequency. For example, between the 1st and 4th peaks at 1 and 2 kHz the mean intensity increased in 0-day-old chicks and decreased in 4-day-old chicks. At 250 and 500 Hz, in contrast, mean intensity decreased in 0-day-old chicks and increased in 4-day-old chicks. Thus the inclusion of data up to the 4th peak in the calculation of thresholds makes the age by frequency interaction clearer.

The important age by frequency interaction in thresholds, however, does not depend on habituation. A re-examination of the data shows that all of the significant developmental differences can be seen from using only the first peak from the staircase as an estimate of threshold, and there is no significant difference between the ages in habituation at the first peak ($p > 0.5$). Using the first peak as the dependent variable, the two ages differed at 1 and 2 kHz [$2.9 < t(18) < 3.1, p < 0.01$] and not at any other frequency [$-1.5 < t(18) < -0.5, p > 0.1$]. Analysis of variance showed a strong age by frequency interaction [$F(5, 108) = 3.5, p < 0.005$] and no main effect of age or frequency.

Figure 5 shows the stimulus levels where chicks first made two correct responses. Collection of data up to only the first peak is thus all that is required to show a significant age
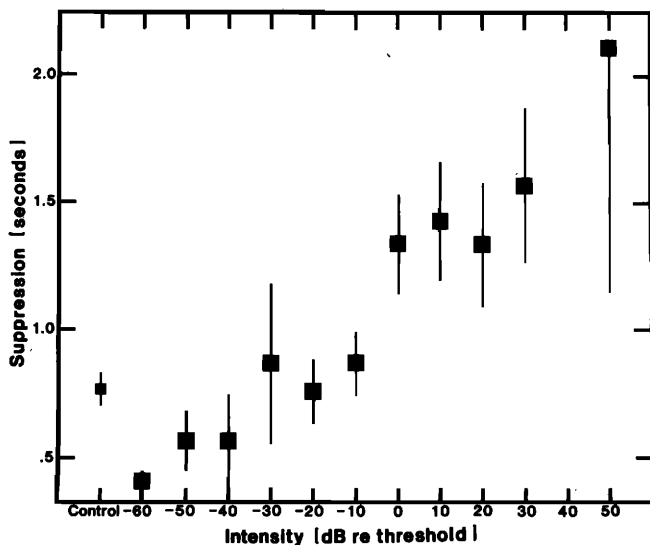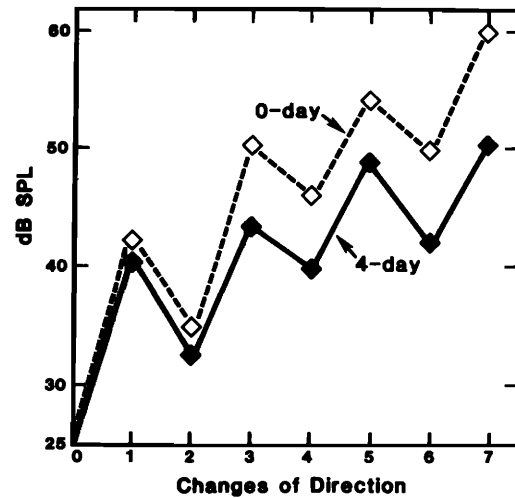

FIG. 3. Psychometric function. Shown are means and standard errors of suppression times from 6180 trials in experiment 1. Intensities are shown relative to each groups' estimated thresholds, from Fig. 2. The average suppression time on all control trials is shown at the far left. The point at 40 dB is off the scale at 4.7 s. Sample sizes are much larger around the estimated threshold than at higher or lower intensities.
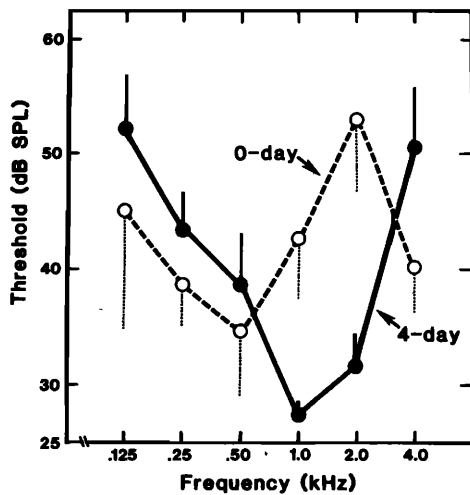
FIG. 5. Absolute thresholds as defined by the 1st change of direction. Means and standard errors of intensities in experiment 1 where subjects first got two successive correct responses are shown. Sample sizes are ten in each group. Note the similarity with Fig. 2, indicating that a single ascending method of limits gives as much information as a longer staircase. Random expectation derived from a neutral model is 47.2 dB.

by frequency interaction. Because a limited number of trials give less opportunity for the algorithm to converge on threshold, these more rapid estimates of threshold may be less precise. (For example, notice that the thresholds at 1 kHz are less different in Fig. 5 than in Fig. 2.) This problem will recur in experiment 3 as only two changes of direction will be used to estimate thresholds.

C. DISCUSSION

1. The age by frequency interaction

At 250 and 500 Hz, absolute thresholds in 0-day-old chicks are no higher than those in 4-day-old chicks. These low-frequency thresholds from the younger subjects are significantly different from random expectation, as discussed above. There is thus no reason to believe these estimates are biased. At 4 days of age these low-frequency thresholds approach a reliable difference from random expectation. This suggests that at low frequencies the thresholds of both ages are likely to be the same.

Because the low-frequency thresholds from the youngest subjects are at least as sensitive as the thresholds from the older subjects, there can be no confounding factors unrelated to auditory perceptions that cause differences in estimated thresholds at these two ages. There is additional evidence that this procedure is appropriate for estimating neonatal thresholds because the data are comparable to results from operant conditioning in adult birds (compare Figs. 2 and 8 at 250 and 500 Hz).

Thresholds at 1 and 2 kHz are clearly lower at 4 days after hatching than on the first day. These data indicate that sensitivity to high frequencies develops after sensitivity to lower frequencies.

In conclusion, the age by frequency interaction shows that the observed differences have something to do with auditory perception. Had there been a consistent difference with age across all frequencies, the appropriateness of the technique for the youngest subjects would remain in doubt.

2. The procedure

These data demonstrate that the 2 down–1 up staircase works well with this unconditioned response to quickly estimate auditory thresholds in young animals. The psychometric function in Fig. 3 shows that the response is under stimulus control. The adaptive algorithm seems to have compensated for all age-related differences that may have affected peep suppression but are not under stimulus control.

Estimated thresholds that are not different from the neutral model indicate that the procedure failed to increase intensities to a level where the tones would elicit reliable suppressions. The real threshold thus lies somewhere above the estimate. Increasing the intensity of the first stimulus is probably the easiest way to correct these ambiguous data. Larger step sizes and more trials could also create nonrandom behaviors on subsequent staircases, but changing these parameters may introduce other problems such as increased habituation.

Habituation to the stimuli over repeated trials is a potential problem, but as indicated by a multiple regression discussed above, the staircase was able to successfully compensate for this confounding variable. To further assess the effects of this habituation, thresholds at one frequency (1 kHz) were re-estimated in experiment 2 with a method of constant stimuli.

II. EXPERIMENT 2: METHOD OF CONSTANT STIMULI

In this experiment absolute auditory thresholds at 1000 Hz were estimated in two ages of young chickens with a method of constant stimuli. This technique requires a large number of birds to estimate a single threshold but has four advantages. First, since each bird responds only once to a pure tone, there is no opportunity for habituation. Second, the method of constant stimuli provides detailed information about underlying psychometric functions (Kling and Riggs, 1971). Third, the sensitivity of one measure of responsiveness to the tones, the duration of peep suppression, can be compared with that of another dependent variable, the number of peeps in each 10-s trial. Fourth, the test is rapid, taking approximately 3 min per subject.

The five intensities used in this experiment were selected to bracket the range around the previously estimated thresholds at 1000 Hz. The experiment was primarily designed to determine if 4-day-old chicks had lower thresholds than 0-day-old chicks at 1000 Hz, but also to provide a comparison between the staircase procedure and the method of constant stimuli.

A. Method

1. Subjects

Subjects were fifty 0-day-old chicks and fifty 4-day-old chicks that had not been previously tested. The testing apparatus was identical to that in experiment 1.

2. Procedure

Five intensities of pulsing 1000-Hz pure tones were preselected to be the stimuli: 0, 20, 26, 32, and 38 dB SPL. Each
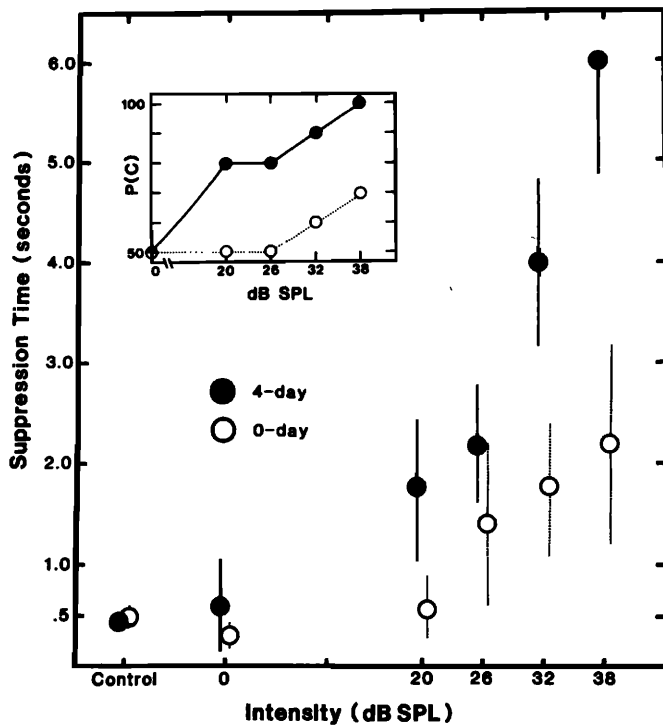
FIG. 6. Data from the method of constant stimuli. Means and standard errors of suppression times from experiment 2 are shown in the main part of this figure. Ten birds at each age heard 1000-Hz tones at 0, 20, 26, 32, or 38 dB. The control trials are shown at the far left. Increasing suppression is shown to louder tones. The inset shows these same data replotted as percent correct. A response, one per subject, was counted as correct if the suppression on the stimulus trial was greater than that on the control trial. Percentages less than 50% are graphed as 50%.

subject was randomly assigned to receive one of these five stimuli. There were thus ten animals in each group.

Animals were acclimated to the chamber for 2–5 min. There was then one control trial, with no stimulus, and then a stimulus trial. As in experiment 1, each trial began 500 ms after 2 peeps in 2 s or less. Unlike experiment 1, each trial lasted 10 s. Both the duration of suppression to the first peep and the number of peeps in each 10-s trial were measured.

B. Results

Mean suppression times are shown for each group in the main part of Fig. 6. The duration of peep suppression rose monotonically with stimulus intensity in both ages. Suppression times increased more rapidly with intensity in 4-day-old chicks than in 0-day-old chicks. An analysis of variance confirmed this developmental trend; there is a significant effect of age $[F(1,90) = 12.9, p < 0.001]$ and intensity $[F(4,90) = 7.9, p < 0.001]$ and no interaction $[F(4,90) = 1.9, p > 0.1]$. The suppression at 0 dB SPL was the same as during control trials, showing that the sound delivery system had no significant source of sound other than the oscillator, and that baseline peep rate did not change over the duration of this brief test.

One way to estimate thresholds from these data is to determine the lowest intensity that elicits suppressions that are greater than those on control trials. Since there was no

difference between the two ages in the duration of suppressions on control trials $[t (98) = 0.3, p > 0.7]$, the 100 control trials were pooled and tested against each group of ten suppressions on stimulus trials. The conservative Mann-Whitney U-test first shows a significant difference between stimulus and control trials at 38 dB for 0-day-old chicks $(U = 304, p < 0.04)$ and at 20 dB for 4-day-old chicks $(U = 230, p < 0.005)$.

A more traditional way to plot the results from the method of constant stimuli is the percentage correct, $p(c)$. As in experiment 1, a "correct response" in this experiment was defined as a pair of suppressions where there was a longer interval before the first peep on the stimulus trial than on the control trial for that animal. The inset in Fig. 6 shows the data replotted as percent correct. These psychometric functions are parallel, with the older birds showing greater sensitivity at 1000 Hz than younger birds. Extrapolating the curves to an intensity where 71% of the trials might be correct suggests a threshold less than 20 dB SPL in 4-day-old chicks and greater than 38 dB SPL in 0-day-old chicks.

The alternate dependent variable, number of peeps per trial, showed no significant difference between the two ages $[F(1, 90) = 0.03, p = 0.86]$. Brief suppressions at the beginning of relatively long trials (10 s) explain the insensitivity of the alternate dependent variable. If animals stop peeping, long suppressions and few peeps are expected. Accordingly, durations of suppression and counts of peeps per trial are expected to be negatively correlated $(r = -0.81$ in Gray and Rubel, 1981; $r = -0.78$ in Kerr et al., 1979). The stimuli in both of these previous experiments, however, were well above threshold and the trials were shorter (5 and 8 s). In the present experiment the duration of suppression was not as well correlated with counts of peeps in 10 s $(r = -0.60)$. This poor correlation indicates that soft tones elicit more transitory changes in the rate of peeping than do louder tones, as expected.

C. Discussion

The age-related difference in absolute threshold at 1000 Hz seen in experiment 1 was repeated with a method of constant stimuli. The method of constant stimuli was, however, slightly more sensitive than the staircase procedure. The estimated thresholds in 4-day-old birds are less than 20 dB SPL compared to 24 dB from the staircase procedure. Thus the prediction of Teller (1984) that staircases can give no better results than the method of constant stimuli with optimally placed stimuli, seems to be supported by these data. The absence of any opportunity to habituate to acoustic stimulation probably also contributed to enhanced sensitivity.

A major disadvantage of the method of constant stimuli is that a large number of subjects are required to estimate a single threshold. A contrasting algorithm, where several thresholds are estimated from only a few subjects, is explored in the next experiment.

III. EXPERIMENT 3: A MODIFIED METHOD OF LIMITS

This experiment is basically a replication of experiment 1, with several changes in the procedure to make it more efficient. These changes include stopping each staircase after

1167   J. Acoust. Soc. Am., Vol. 77, No. 3, March 1985

L. Gray and E. W. Rubel: Development of hearing in chickens   1167

only two changes of direction, testing each subject at several frequencies, and starting the staircases at different intensities.

A statistically reliable difference between 0- and 4-day-old chickens was obtained in experiment 1 after only one change of direction in an ascending s aircase (see Fig. 5). That is, to demonstrate a significant age by frequency interaction, it was sufficient to increase intensities until the first two successive correct responses were obtained. Reaching the first peak takes only a few trials. Reaching a second peak would add only a few more trials. Using a mean of these two peaks might improve the estimates of threshold as well as normalize their distribution (Mood and Graybill, 1963).

If only a few trials were run at each frequency, then it would be possible to test each subject at several frequencies. A relatively complete audiogram could thus be measured in each bird.

The thresholds shown in Fig. 2 are likely to provide efficient starting intensities for future staircases. The "2-down–1-up" algorithm tends to drive intensities upward as shown in Fig. 4. An ideal starting intensity would thus be slightly below the lower mean threshold at each frequency. Birds that could hear these tones would be likely to get the first few trials correct, while less sensitive subjects would drive the intensities upward. The ambiguous results at 4000 Hz would hopefully be avoided after starting at a louder intensity, as discussed above. Step sizes were also enlarged to enhance possible age-related differences in absolute thresholds.

A. Method

The general strategy was to use the two-interval procedure of experiment 1, with only two changes of direction in each staircase, attempting to test each bird at all six frequencies. The starting intensities and step sizes were also slightly different.

1. Subjects

Subjects were 41 0-day-old chicks and 47 4-day-old chicks that had not been tested previously. The unequal sample sizes are explained below. An additional 13 subjects failed to give any usable data.

2. Procedure

Staircases were started at different intensities for each of six frequencies: 47, 40, 35, 23, 26, and 47 dB SPL for 125, 250, 500, 1000, 2000, and 4000 Hz, respectively. Staircases to estimate absolute thresholds of these six frequencies were presented in random order; one staircase was finished before the next began. As in the previous experiments the stimuli were pulsing pure tones. The testing apparatus was identical.

As in experiment 1, stimulus and control trials were presented in randomly ordered pairs. The intensity of the acoustic stimulus was increased by 10 dB after every incorrect response until the first peak; that is, where the bird made two successive correct responses. Intensities were then decreased by 5 dB after every two correct responses until the second peak, at the next incorrect response. Threshold was
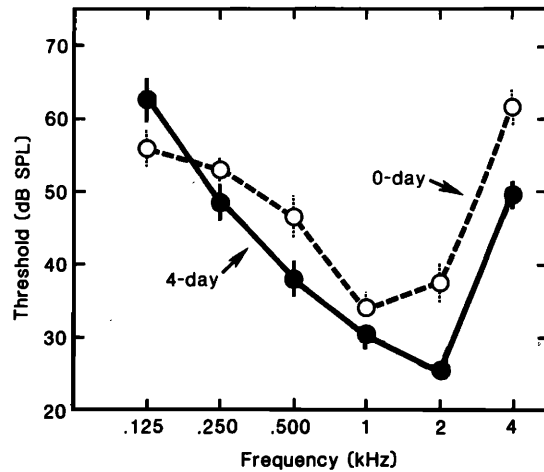

FIG. 7. Data from the modified method of limits. Means and standard errors of all estimated thresholds available from experiment 3 are shown. Sample sizes from 0-day-old chicks are 31, 35, 32, 34, 35, and 29 at 125–4000 Hz, respectively. Sample sizes from 4-day-old chicks are 30, 27, 28, 31, 33, and 23 at 125–4000 Hz.

calculated as the mean of the two peaks. Each staircase terminated after two changes of direction or 11 trials, whichever came first.

A staircase with a different frequency began without pause after the termination of the previous staircase until all six frequencies were tested or the bird failed to peep for 2 min. Data were retained for analysis if there was at least one completed staircase at any frequency. A test in which a bird completed all six frequencies took about 25 min.

3. Neutral model

Simulation of this procedure revealed that a calculated threshold 18 dB higher than the starting intensity is expected from random behavior.

B. Results

Average absolute thresholds from the modified method of limits are shown in Fig. 7 for two ages and six different frequencies. All available data are pooled to make this figure; forty-one 0-day-old and forty-seven 4-day-old subjects contributed from one to six thresholds each. There are approximately 30 estimates in each group.

There is a clear developmental gradient over frequencies with age evident in these data. Six t tests comparing the two ages at each frequency, keeping the overall alpha below 0.05, show the only reliable differences are at 2000 and 4000 Hz. Thus thresholds at low frequencies were not reliably different at both ages whereas 4-day-old chicks had lower thresholds at high frequencies than newly hatched birds. With different starting levels there is now a clear difference in thresholds at 4000 Hz, not seen in experiment 1.

Normalized thresholds were used to look for (1) an effect of the order, 1st through 6th, in which the frequencies were tested, (2) consistently high or low thresholds across all six frequencies within given subjects, and (3) a relationship between thresholds and the number of staircases completed. Z scores were calculated for each estimated threshold based on the 12 distributions for each of two ages and six frequencies shown in Fig. 7.

1. Order effect

There was no order effect. Six separate pools of Z scores were obtained for tests that occurred 1st through 6th in the randomized testing sequence. There was no significant trend in thresholds based on the position of the staircase within the testing sequence ($r = -0.15$). Thus habituation occurring from one frequency to another does not appear to confound these data. A threshold measured last is likely to be as sensitive as one measured first.

2. Subject effect

Some animals are more sensitive than others. Eighty-eight pools of Z scores were obtained from all thresholds from each subject. Ten of these, or 16%, were significantly different from zero at the 0.05 level. This is more difference than expected by chance ($p < 0.0005$). This means that a bird with a higher than average threshold at one frequency is more likely to have higher thresholds at other frequencies. The repeated measures on each subject are thus somewhat correlated.

3. Number effect

Of all the subjects tested, only sixteen 0-day-old and ten 4-day-old birds completed all six staircases. Only this subset of the data is appropriate for a repeated measures analysis of variance (Winer, 1971). This evaluation showed a significant age by frequency interaction [$F(5, 90) = 3.14, p = 0.01$], as expected from experiment 1.

Examination of the normalized thresholds showed no difference between data from birds that finished all six tests and birds that finished less then six tests [$t(365) = 0.002$]. Thus there appears to be no bias in doing some statistical analyses on only the birds that completed all six tests, yet pooling all available data to obtain the best estimates of thresholds. The statistical problems resulting from partially dependent and partially independent measures are a disadvantage of the modified method of limits.

4. Comparison with random expectation

Repeated simulation of this procedure with a neutral model revealed an average calculated threshold 18 dB higher than the starting intensity. Random expectation is thus 65, 58, 53, 41, 44, and 65 dB SPL at 125–4000, respectively. One-tailed t tests showed that for the 0-day-old chicks the probabilities that the estimated thresholds were below the chance level were less than 0.005, 0.002, 0.02, 0.002, 0.01 and 0.12 at 125–4000 Hz, respectively. For the 4-day-old subjects these probabilities were less than 0.25 at 125 Hz and 0.0001 at the other frequencies. Remember that results which are not different from the neutral model may be underestimates of the true threshold. It is thus appropriate to conclude that thresholds are similar for both ages at 250, 500, and 1000 Hz and different at 2000 and 4000 Hz. At 125 Hz we can conclude only that the younger subjects are not less sensitive than the older subjects.

C. Discussion

This experiment showed the same developmental gradient across frequencies and age as shown in experiment 1.

There is an age by frequency interaction in the development of absolute thresholds. Low-frequency thresholds were similar at 0 and 4 days of age, while high-frequency thresholds were different. Thus thresholds mature at low frequencies before high.

While the modified method of limits replicates the age by frequency interaction, differences at each frequency may be less consistent because of the limited number of trials in each staircase. There appears to be such an inconsistency at 1 kHz in Figs. 2 and 7. Statistically, however, the 95% confidence intervals of the differences between the ages at both 1 kHz and 500 Hz in experiments 1 and 3 overlap. That is, observed differences in estimated thresholds are within the range expected by chance. The location of 1 kHz at a "transition zone"—within that range of frequencies where thresholds are in the process of improving at 0 days of age—probably contributes to this variability between experiments.

This version of the modified method of limits provides an efficient algorithm for obtaining an audiogram on each neonatal subject. There seems to be little or no habituation from one frequency to another, as indicated by normalized thresholds; the first estimate is as sensitive as the last. More than one threshold from each subject allows a test of intra-subject consistency but causes some statistical difficulties.

Compared to experiment 1, the louder starting intensities and larger initial step size have caused less random behavior, again demonstrating that peep suppression is strongly influenced by stimulus intensity. Because of the overall departure from randomness, the thresholds from the 4-day-old birds are probably the best available estimate of a true audiogram for these young subjects.

IV. GENERAL DISCUSSION

A. The age by frequency interaction

Inferences about the perceptions of nonverbal subjects are admittedly speculative, even under the best circumstances. The age by frequency interaction in these data, however, provides strong evidence that the developmental differences are the result of changing reactions to the onset of pure tones. In each experiment we showed that thresholds of younger and older chicks are different at high frequencies but the same at low frequencies. This trend was consistent across all three experiments, and thus supports a conclusion that thresholds to low-frequency tones mature before thresholds to high-frequency tones.

The double-interval staircase procedures (experiments 1 and 3) successfully compensated for any age-related differences in baseline peep rate, attention span, or proclivity to suppress vocalizations. The start of trials only when birds were peeping at a predetermined rate as well as the control trial for each stimulus trial was probably responsible for being able to detect the age by frequency interaction. For these reasons age-related differences observed with these procedures likely indicate some change in perceptual development and not some nonsensory artifact. Estimates of absolute thresholds may be lowered slightly by future methodological improvements, but the age by frequency interaction should not change with a more sensitive algorithm.
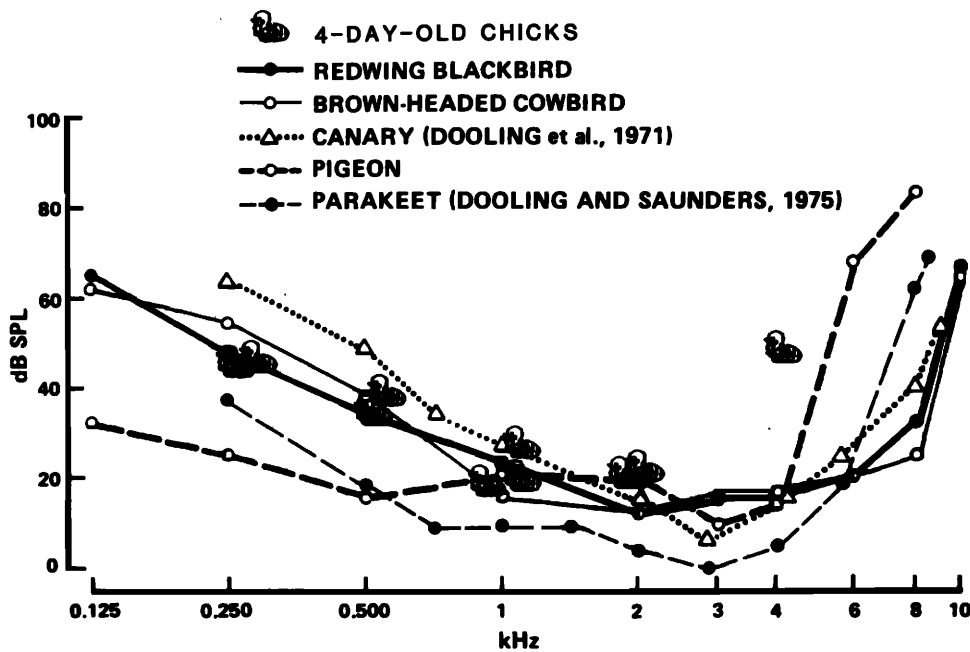
FIG. 8. Comparison with other data. The data from the 4-day-old subjects in experiments 1, 2, and 3 are shown with absolute thresholds from mature birds. Note the similarity in thresholds between young chickens and other mature birds, especially at low frequencies. Note also the similarity among the various estimates of chickens' thresholds. A figure from R. D. Hienz, J. M. Sinnott, and M. B. Sachs, copyright 1977 by the American Psychological Association, is reprinted with permission of the authors and publisher. All nonrandom estimates ($p < 0.10$) from experiments 1 and 3 are shown, and the threshold at 1 kHz from experiment 2 is shown at 20 dB. Some symbols are spread horizontally to accommodate similar data.

A developmental gradient in responsiveness to different frequencies over age is seen in the vast majority of studies on auditory development in both mammals and birds (Gottlieb, 1971; Rubel, 1978). Behavioral responsiveness to low frequencies develops before responsiveness to high frequencies in chicks (Gray and Rubel, 1981; Jackson and Rubel, 1978), mice (Hack, 1968; Shnerson and Willot, 1980), cats, rabbits, dogs, and mink (Foss and Flottorp, 1974). Additionally, cochlear responses to low frequencies develop before responses to high frequencies in chicks (Rebillard and Rubel, 1981; Saunders et al., 1973), rats (Crowley and Hepp-Reymond, 1966), and gerbils (Finck et al., 1972) as do responses in central auditory neurons in bats (Brown et al., 1978) and mice (Mikaelian et al., 1965). The observations of ontogenetic changes in absolute thresholds reported in this paper are therefore in close agreement with other measures of functional development in the auditory system. The major conclusion of this study, that behavioral sensitivity develops later for high frequencies than for low frequencies, is also consistent with recent evidence for a changing tonotopic organization along the developing cochlea (Harris and Dallos, 1984; Lippe and Rubel, 1983; Rubel and Ryals, 1983).

B. Comparison with other data

The estimated thresholds from this procedure compare well with data from other studies of birds. Figure 8 shows the absolute thresholds of the 4-day-old birds from experiments 1–3 compared with estimates from other species of birds. The earlier data are all from mature subjects that were tested after relatively elaborate training (Hienz et al., 1977).

Thresholds are equivalent, especially at low frequencies. The sensitivity of the staircase procedure using an unconditioned response thus approaches the sensitivity of other psychophysical techniques involving operant or classical conditioning. This similarity applies to grouped data, the mean of estimated thresholds from several subjects, and may apply to individual audiograms as well.

Since the peep suppression technique takes only a few minutes per bird and requires no previous training, it allows an investigation of developmental psychophysics. Tests that take days to complete are clearly inappropriate for measuring rapid neonatal changes in auditory perception. In addition, training introduces a potential source of confounding because observed developmental differences could be due to age-related differences in abilities to learn the testing procedure.

Auditory thresholds in other neonatal animals are, by present estimates, relatively high (50–110 dB SPL in 2- to 6-day-old cats and mice; Ehret, 1976, 1977; Ehret and Romand, 1981; Shnerson and Willott, 1980). Unfortunately, these high thresholds are somewhat difficult to interpret because lower sensitivity could be due to either the animals or the testing procedure. Since the auditory systems of cats and mice are relatively immature at birth (Pujol and Marty, 1968; Rubel, 1978), their high neonatal thresholds probably reflect an early stage in the development of auditory perception. Nevertheless, confounding and less interesting effects from the reactions of young animals to psychophysical tests are difficult to rule out (Stebbins, 1970). That is, without an example of psychophysical performance that approaches adult levels, the young animals' reactions to the *procedure*

rather than the *stimuli* remain a possible explanation for developmental differences. Neonatal thresholds that approach adult levels under some conditions ease these interpretive and methodological concerns that psychophysical tests in young animals uniformly underestimate optimal performance. A procedure is likely to be appropriate for measuring neonates' thresholds if the estimates approach adult levels under some conditions.

C. Discussion of the procedures

There are differences among the estimates of thresholds from these three experiments. The differences are probably due to (1) habituation, (2) starting levels, (3) number of trials, and (4) natural variability. (1) The method of constant stimuli is the most sensitive procedure. This is expected since subjects cannot habituate to a single trial. The method of constant stimuli is also the least efficient procedure because each subject responds to only one stimulus. (2) The starting level at 4 kHz used in experiment 3 reveals a significant difference between the ages. The lack of difference at 4 kHz between the ages in experiment 1 is thus likely due to an inappropriately low starting level. Starting levels at 125 Hz in both experiments 1 and 3 are too low and lead to random thresholds. (3) The accuracy of estimated thresholds depends heavily on the number of trials per staircase (Teller, 1984). With fewer trials there is less opportunity for convergence and a greater opportunity for random responses to misdirect the staircase. Such expected inconsistency is seen in a comparison of thresholds estimated from only one or two changes of direction (Figs. 5 and 7) with thresholds estimated from more trials (Fig. 2). Large numbers of trials would be ideal, but newborn subjects do not tolerate long testing sessions and repeated trials cause habituation. Thus these procedures were designed as different approaches to the trade-off between number of trials, attention span, and habituation. Developmental trends can be evaluated with statistically acceptable reliability using staircases that contain relatively few trials as seen in experiment 3, but differences at each frequency may be inconsistent. More consistency requires more trials, time, and subjects as seen in experiments 1 and 2. (4) Relatively large variability may be an unavoidable disadvantage of doing psychophysical research on neonates. Before these experiments were begun, multiple pilot studies were conducted on various modifications of the procedure. Starting trials only when the birds were actively vocalizing, pairing each stimulus trial with a control trial, and not counting peeps within 500 ms of each trial's start, seemed to decrease variability in estimated thresholds. Nevertheless, considerable variability remains and contributes to inconsistencies between the experiments.

A possible disadvantage of using peep suppression as a response for psychoacoustical experiments is that thresholds were measured in actively vocalizing subjects. It is possible that the peeps could mask the stimuli, or that an acoustic reflex during vocalization (Counter and Borg, 1979) could decrease sensitivity. There are, however, three reasons why this self-masking is unlikely to have affected the data: (1) the onset of tones at the end of peeps; (2) the use of a 500-ms adjustment period to pulsing tones; and (3) the spectral composition of the peeps. (1) The specially built discriminator triggers at the end of peeps. With the relatively long rise times of these stimuli, full power of the stimulus was thus reached after most vocalizations had ceased. (2) The period between pulsing tones was $\frac{1}{2}$ s, and the average interpeep interval was 0.4 s. Thus the onset of the second tone pulse is likely to occur when the chicken is silent. It is at this time, $\frac{1}{2}$ s after the start of each trial, that the measurement of peep suppression begins. There is one onset, almost $\frac{1}{2}$ s of steady tone, a brief pause, and then a second onset before the timing of suppression begins. Therefore, birds have ample opportunity between their peeps to react to the stimulus before the response is recorded. (3) Even if thresholds do not return to normal between peeps and some masking occurs, it is unlikely to be the reason for the age by frequency interaction shown by these data. The lowering of voices over these ages occurs between 3 and 2 kHz. Spectrographic analyses of peeps reveal no measurable power below 1500 Hz, equivalent power at both ages around 4 kHz, and more power around 2 kHz in the peeps of 4-day-old birds than in the 0-day-old birds. Because there is more energy around 2 kHz in the peeps of 4-day-old chicks than 0-day-olds, masking by the subjects' vocalizations would be expected to raise the thresholds at 2 kHz in 4-day-old chicks relative to 0-day-olds. The opposite result was obtained. Thus, the lower thresholds at high frequencies obtained from the older birds argues against a major contribution of masking to these results.

Andrew, R. J. (1975). "Effects of testosterone on the calling of the domestic chick in a strange environment," Anim. Behav. 23, 169–178.

Bermant, G. (1963). "Intensity and rate of distress calling in chicks as a function of social contact," Anim. Behav. 11, 514–517.

Blough D., and Blough, P. (1977). "Animal psychophysics," in *Handbook of Operant Behavior*, edited by W. K. Honig and J. E. R. Staddon (Prentice-Hall, Englewood Cliffs, NJ), pp. 514–539.

Brown, P. E., Grinell, A. D., and Harrison, J. B. (1978). "The development of hearing in the pallid bat, *Antrozous pallidus*," J. Comp. Physiol. 126, 169–182.

Collias, N., and Joos, M. (1953). "The spectrographic analysis of sound signals of the domestic fowl," Behaviour 5, 175–187.

Counter, S. A., and Borg, E. (1979). "Physiological activation of the stapedius muscle in *Gallus gallus*," Acta Otolaryngol. 403, 13–19.

Crowley, D. E., and Hepp-Reymond, M. D. (1966). "Development of cochlear function in the ear of the infant rat," J. Comp. Physiol. Psychol. 62, 427–432.

Ehret, G. (1976). "Development of absolute thresholds in the house mouse (*Mus musculus*)," J. Am. Audiol. Soc. 1, 179–184.

Ehret, G. (1977). "Postnatal development in the acoustic system of the

house mouse in the light of developing masked thresholds," J. Acoust. Soc. Am. 62, 143–148.

Ehret, G., and Romand, R. (1981). "Postnatal development of absolute auditory thresholds in kittens," J. Comp. Physiol. Psychol. 95, 304–311.

Finck, F., Schneck, C. D., and Hartman, A. F. (1972). "Development of cochlear function in the neonate mongolian gerbil (Meriones unguiculatus)," J. Comp. Physiol. Psychol. 78, 375–380.

Foss, I., and Flottorp, G. (1974). "A comparative study of the development of hearing and vision in various species commonly used in experiments," Acta Otolaryngol. 77, 202–214.

Gottlieb, G. (1971). "Ontogenesis of sensory function in birds and mammals," in The Biopsychology of Development, edited by E. Tobach, L. R. Aronson, and E. Shaw (Academic, New York), pp. 67–128.

Gottlieb, G. (1976). "The roles of experience in the development of behavior and the nervous system," in Studies on the Development of Behavior and the Nervous System, Vol. 3, Neural and Behavioral Specificity, edited by G. Gottlieb (Academic, New York), pp. 25–54.

Gray, L., and Rubel, E. W (1981). "Development of responsiveness to suprathreshold acoustic stimulation in chickens," J. Comp. Physiol. Psychol. 95, 188–198.

Hack, M. H. (1968). "The development of the preyer reflex in the sh-1 mouse," J. Aud. Res. 8, 449–457.

Harris, D., and Dallos, P. (1984). "Ontogenetic changes in frequency mapping of a mammalian ear," Science 225, 741–743.

Haroutunian, V., and Campbell, B. A. (1981). "Development and habituation of the heart rate orienting response to auditory and visual stimuli in the rat," J. Comp. Physiol. Psychol. 95, 166–174.

Hienz, R. D., Sinnott, J. M., and Sachs, M. B. (1977). "Auditory sensitivity of the redwing blackbird (Agelaius phoeniceus) and brown-headed cowbird (Molothrus ater)," J. Comp. Physiol. Psychol. 91, 1365–1376.

Hutt, S. J., Hutt, C., Lenard, H. G., Bernuth, H., and Muntjewerff, W. J. (1968). "Auditory responsivity in the human neonate," Nature 218, 888–890.

Jackson, H., and Rubel E. W (1978). "Ontogeny of behavioral responsiveness to sound in the chick embryo as indicated by electrical recording of motility," J. Comp. Physiol. Psychol. 92, 682–696.

Kearsley, R. B. (1973). "The newborn's response to auditory stimulation: a demonstration of orienting and defensive behavior," Child Dev. 44, 582–590.

Kerr, L. M., Ostapoff, E. M., and Rubel, E. W (1979). "Influence of acoustic experience on the ontogeny of frequency generalization gradients in the chicken," J. Exp. Psychol.: Anim. Behav. Proc. 5, 97–115.

Kling, J. W., and Riggs, L. A. (1971). Experimental Psychology (Holt, Rinehart and Winston, New York), 3rd ed.

Lippe, W., and Rubel, E. W (1983). "Development of the place principle: tonotopic organization," Science 219, 514–516.

Levitt, H. (1971). "Transformed up–down methods in psychoacoustics," J.

Acoust. Soc. Am. 49, 467–477.

Martin, G. K., Lonsbury-Martin, B. L., and Kimm, J. (1980). "A rabbit preparation for neuro-behavioral auditory research," Hear. Res. 2, 65–78.

Mikaelian, D., Alford, B. R., and Ruben, R. J. (1965). "Cochlear potentials and VIII nerve action potentials in normal and genetically deaf mice," Ann. Otol. Rhinol. Laryngol. 74, 146–157.

Mood, A. M., and Graybill, F. A. (1963). Introduction to the Theory of Statistics (McGraw-Hill, New York), 2nd ed.

Moody, D. (1970). "Reaction time as an index of sensory function," in Animal Psychophysics, edited by W. C. Stebbins (Appleton-Century Crofts, New York).

Pujol, R., and Marty, R. (1968). "Structural and physiological relationships of the maturing auditory system," in Ontogenesis of the Brain, edited by L. Jilek and S. Trojan (Charles U. P., Prague), pp. 377–385.

Rebillard, G., and Rubel, E. W (1981). "Electrophysiological study of the maturation of auditory responses from the inner ear of the chick," Brain Res. 229, 15–23.

Rubel, E. W (1978). "Ontogeny of structure and function in the vertebrate auditory system," in Handbook of Sensory Physiology, Vol. IX: Development of Sensory Systems, edited by M. Jacobson (Springer-Verlag, New York), pp. 135–237.

Rubel, E. W, and Rosenthal, M. H. (1975). "The ontogeny of auditory frequency generalization in the chicken," J. Exp. Psychol.: Anim. Behav. Proc. 1, 287–297.

Rubel, E. W, and Ryals, B. M. (1983). "Development of the place principle: acoustic trauma," Science 219, 512–514.

Saunders, J. C., Coles, R. B., and Gates, G. R. (1973). "The development of auditory evoked responses in the cochlea and cochlear nuclei of the chick," Brain Res. 63, 59–74.

Severns, M., Gray, L., and Rubel, E. W (in press). "An avian vocalization detector," Physiol. Behav.

Shnerson, A., and Willott, J. F. (1980). "Ontogeny of the acoustic startle response in c57b1/6j mouse pups," J. Comp. Physiol. Psychol. 94, 36–40.

Smith, Z. D. J. (1981). "Organization and development of brain stem auditory nuclei of the chicken: dendritic development in n. laminaris," J. Comp. Neurol. 203, 309–333.

Stebbins, W. C. (1970). Animal Psychophysics (Appleton-Century Crofts, New York).

Teller, D. Y. (1984). "Psychophysics of infant vision: Definitions and limitations," in Measurement of Vision and Hearing During the First Year of Postnatal Life, edited by G. Gottlieb and N. Krasnegor (Ablex, New York).

Wetherill, G. B., and Levitt, H. (1965). "Sequential estimation of points on a psychometric function," British J. Math. Stat. Psychol. 18, 1–10.

Winer, B. J. (1971). Statistical Principles in Experimental Design (McGraw-Hill, New York), 2nd ed.