

BRIEF REPORT

## The Head-Turn Preference Procedure for Testing Auditory Perception

DEBORAH G. KEMLER NELSON  
*Swarthmore College*

PETER W. JUSZYK  
*State University of New York at Buffalo*

DENISE R. MANDEL, JAMES MYERS, ALICE TURK, AND LOUANN GERKEN  
*State University of New York at Buffalo*

The Head-Turn Preference Procedure (HPP) is valuable for testing perception of sustained auditory materials, particularly speech. This article presents a detailed description of the current version of HPP, new evidence of the objectivity of measurements within it, and an account of recent modifications.

---

speech perception    auditory stimuli    observer bias    head-turn    preferences

---

For almost a decade, the Head-Turn Preference Procedure (HPP) has been used in infant speech perception research. Over this period, the procedure has evolved in ways that render it more sensitive and also more flexible in its application to experimental questions and to infants of different ages. Because HPP is well-suited to testing responses to long samples of speech or other sustained auditory stimuli (i.e., unlike the high-amplitude sucking procedure), it is a methodology of considerable interest to infant researchers. Yet, because it has been evolving, no complete, up-to-date description of HPP is easily accessible. Even recent empirical papers that use HPP require the reader to piece together a complete picture of how the method works from previously published reports. Given that the current version of the procedure is proving so effective, the time is ripe for focusing on the methodology itself in this article. In addition, we report some new data that bolster the case for the objectivity of the measurements obtained in the procedure.

---

The preparation of this article was assisted by research grants from NICHD (#15795) to P.W.J. and from NSF to L.A.G. (#DBS91020952). In addition, D.R.M., J. M., and A.T. were supported on a training grant from NIDCD (DC #00036). We thank Ann Marie Jusczyk and Nancy Redanz for comments made on earlier versions of the manuscript.

Correspondence and requests for reprints should be sent to Peter W. Jusczyk, Department of Psychology and Center for Cognitive Sciences—Park Hall, State University of New York at Buffalo, Buffalo, NY 14260-4110.

The HPP takes advantage of certain facts about infants. First, they tend to orient visually to an attended sound source. Second, they learn to maintain a response (e.g., a head-turn) when motivating stimulation is contingent on their behavior. Accordingly, the extent of infants' attention to an auditory stimulus can be assessed by examining the length of time they turn their heads toward the sound. To increase the sensitivity of this measure for auditory preferences, the infants are permitted to continue to hear the sound only as long as the head-turn is maintained.

The questions to which HPP applies concern whether infants prefer one kind of auditory stimulus over another. The preferences themselves may be of interest, or the preferences may simply serve as an index of discriminative skills. Within the current version of HPP, preferences are measured by exposing each infant to one type of sound on half the trials and the other type of sound on the remainder. The sounds emanate from locations that require the infants to turn their heads to visually localize them. The index of preference is the difference in the average length of the infant's looking time to the two different kinds of stimuli over the test-trial series.

### THE UPDATED VERSION OF THE PROCEDURE

#### Details of the Procedure

The HPP is conducted in a testing booth which is depicted in Figure 1. A three-sided booth is

constructed from panels that are 120 cm  $\times$  180 cm. Curtains are hung from the ceiling to the top of the booth to block the infant's view of the rest of the room. Loudspeakers are mounted into the walls of the two side panels at about the level of the infant's head. A small red light is mounted on each of the side panels in the vicinity of the loudspeaker. The center panel which the infant faces has a small green light mounted at the infant's eye level.

The panels are made of pegboard, backed with cardboard, except for a small area just above the center light. The pegboard holes in this area allow an observer to monitor the infant's behaviors. Directly below the center light is a 5-cm hole to accommodate the lens of a video camera. During the experiment, the caregiver sits on a chair, facing the center panel, holding the infant. The distance of the chair from the front panel results in alignment of the caregiver's knees with the two side lights. The entire room containing the test booth is dimly illuminated.

A test trial begins by drawing the infant's attention to center by flashing the center light and, if necessary, moving a silent puppet just

above the panel. Once the infant's attention is at midline, the puppet disappears, the center light is turned off, and a flashing light at one (and only one) of the two sides signals the availability of an auditory stimulus on that side. Once the infant turns to that side, the stimulus begins to play. It continues (and the side light keeps flashing) until the infant turns away for at least a continuous period of 2 s (or until the entire stimulus for that trial has been played, e.g., about 2 s). The infant's looking time is the total time the infant orients to the sample.

Important features of the current version of HPP are that, over the series of test trials (usually 12), both kinds of stimuli are heard by each infant from both sides of the booth, and the order of the stimuli and the order of locations are independent of the infant's behavior. In the evolution of HPP, this constitutes a significant set of changes. In some of our earlier studies and others that preceded them (Fernald, 1985), type of stimulus and location of presentation (or response location; Colombo & Bundy, 1981) were perfectly correlated within subjects. Moreover, in Fernald (1985), the sole measure of preference was based on the direction of

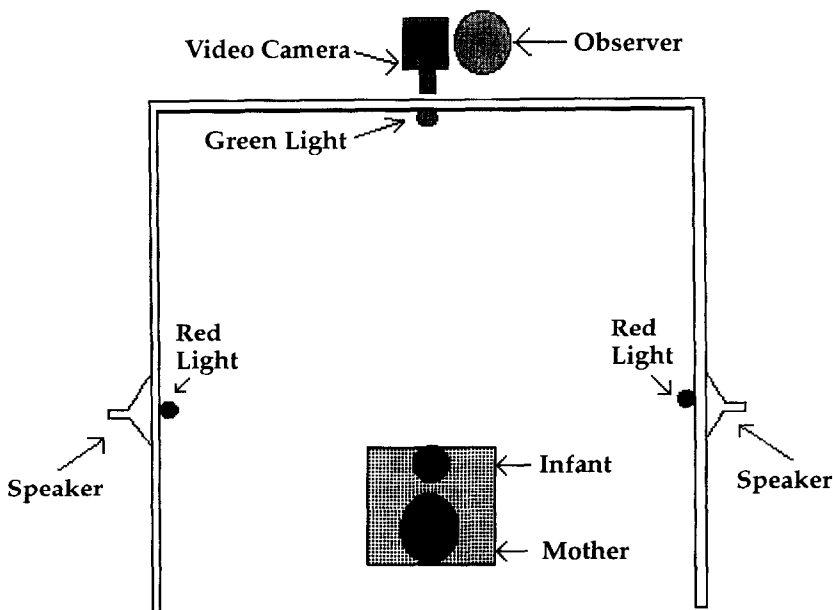


Figure 1. Layout of the testing booth for use in the Head-Turn Preference Procedure. The infant is seated on the mother's lap and facing the center panel. The observer behind the panel looks through the holes in the pegboard to judge the head-turns.

head-turns. We have found this measure to be far less sensitive to preferences than the durational measure of preference (e. g., Hirsh-Pasek et al., 1987; Kemler Nelson, Hirsh-Pasek, Jusczyk, & Wright-Cassidy, 1989), which now serves as our only dependent variable.

In HPP, the test period is always preceded by a training period. Training trials are designed to acquaint the infants with the stimuli and "inform" the infants of the contingency between their head-turns and auditory stimulation. Usually there are four training trials, one occurrence of each type of auditory stimulus at each of the two locations. During training trials, unlike the test trials, the side light is extinguished as soon as the infant looks to that location, and only the sound continues throughout the head-turn. This procedure was adopted because a previous practice of keeping the side light on during training appeared to lead to shorter lengths of orientation overall on the test trials to follow. In all other respects, the training trials are the same as the test trials.

#### The Role of the Observer

A single observer sits behind the center panel. The observer has control of a response box, tethered to a computer, which controls the presentation of the stimuli and registers the timing of the responses. The observer signals with one button when attention is centered at the beginning of the trial, and then with another button whenever the infant orients by at least 30° in the direction of stimulus location (as indicated to the observer by the side of the flashing light). An additional button is pressed when the infant deviates from the critical orientation. On test trials, these observer responses initiate a flashing side light, begin an auditory stimulus, and terminate the sound and the light when 2 consecutive seconds have passed without (at least 30° of) orientation to them.

Aside from the increased sensitivity of a live observer to the infant's shifts of gaze, the presence of a live observer in the testing room has the advantage of allowing this person to control a puppet in the between-trial intervals. This serves to bring infants' attention to center as trials begin and generally maintains interest and minimizes restlessness over the entire experimental session (between 5 and 10 min in length).

To guard against observer bias, the observer is blind to the stimulus and location order, both

determined by the computer. Also, adjustment of sound levels before the session is carried out by a different person (a safeguard, no longer strictly necessary, which was instituted in earlier versions of the procedure when the stimulus types were perfectly correlated with the speaker locations). When the session begins, the observer and the infant's caregiver wear tight-fitting earphones (SONY MDR-V600) through which a loud masker is played. In most studies, we have used taped instrumental music as the masker. The selections are recorded without breaks between pieces and chosen to be rather loud when the volume is set at a standard level. On occasion, we have used a speech masker instead, in which the materials are similar to the speech materials the infant is hearing, produced by the same talker (e.g., Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993).

#### THE OBJECTIVITY OF THE MEASUREMENTS

##### Reliability Checks

In recent versions of HPP, a video camera has been used to provide a permanent record of the test sessions. Videotapes allow us to do reliability checks on the response measures provided by the live observer. A reliability checker views and scores the videotapes with the sound turned off, using the same type of response box as the live observer. Reliability in judging the timing of the head-turns has proven to be quite high. For the 22 infants they videotaped, Jusczyk, Cutler, and Redanz (1993) compared the average differences in listening time to the two types of samples as judged by the live and videotape observers. They found a correlation of .94. On 72% of the individual trials, the discrepancies in the timing of trials between the live and videotape observers was less than 0.50 s. On 15%, there was a discrepancy of 1.00 s or more. Inspection of these discrepant trials showed that relative to the videotape observer, the live observer was no more likely to make timing decisions consistent with the hypothesis that was being tested. The live observer recorded times that were on average 0.88 s longer than the videotape observer on trials hypothesized to be of the favored type (i.e., consistent with preferences predicted by the working hypothesis), and 1.47 s longer on trials that were hypothesized to be of the disfavored type

(i.e., contrary to the hypothesis).<sup>1</sup> The difference was not significant,  $t(39) = 0.56$ ,  $p = .60$ , and was in the direction opposite to live observer bias.

Gerken, Jusczyk, and Mandel (1994) reported comparable results. Interobserver agreement was .95 in one experiment and .99 in the other. The discrepancy between the live and videotape observers was less than 0.50 s on 65% and 70% of the trials in the respective experiments. Discrepancies of greater than 1.00 s occurred on 17% and 12% of the trials in the two experiments, again with no systematic differences between observers.

#### New Tests for the Possibility of Observer Bias

Recently, Fernald (1993) questioned whether the conditions in our studies serve as sufficient safeguards against observer bias. She noted that correlations in response times between live and videotape scorers may be inflated by variation in stimulus durations across trials, which are under the control of the live observer (who terminates the stimulus when the infant looks away for 2 s). That is why, in addition to reporting correlations, we analyze the nature of the discrepancies for those trials on which the two observers disagree. Our analyses of the discrepant trials go a long way toward allaying the concerns raised by Fernald. Now, we report for the first time an even stronger test for the possibility of observer bias.

To test the effectiveness of music maskers in our studies, we ran some experiments with college-aged adults. In one experiment, we investigated the effectiveness of maskers for earlier versions of HPP in which all the samples of a particular type are played to the same side throughout the test trials (e.g., Jusczyk et al., 1992). We employed the stimuli from the two test conditions that produced significant results in the infant studies by Gerken et al. (1994).

Eight normal-hearing adults were tested on each set of stimuli. The listener was seated in

the chair (normally occupied by the caregiver and infant) in the middle of the test booth. The nature of the stimulus materials was carefully explained to each participant. Participants were told that some of the speech samples contained pauses just prior to the verbs in the sentences and that other samples had pauses immediately after the verbs. They were also told that all the samples of a given type (e.g., pause before verb) would be played on the same side of the room. For each trial, their task was to circle the correct response (i.e., "before verb" or "after verb") on an answer sheet. Then, without earphones or masker, each listener heard a different order of 12 test trials. Not surprisingly, performance was virtually perfect (2 of the 16 listeners missed only the first trial and the other 14 listeners were perfect). Thus, it was clear that participants understood the nature of the task.

During the next phase, listeners were told they would be presented with the same materials, but in a new test order. They were informed that this time they would wear headphones over which loud masking music would be played (at the same loudness level used by the observers in our experiments). Their task was still to identify the type of sample on each trial. Participants were told that they might have great difficulty in hearing the stimuli and to use the flashing side lights as an indication of when a sample was playing. They were told to respond on each trial even if they had to guess. This listening situation was meant to mimic that of the experienced observer in the infant experiments.

Upon completion of the experiment, all 16 listeners reported that they were unable to hear the stimuli over the masker. These impressions were reflected in their performance: There was no evidence that they discriminated the two types of samples when masking music was present,  $t(7) = 0.23$ ,  $p > .80$ , for the stimuli from Experiment 1 of Gerken et al. (in press) and  $t(7) = 0.33$ ,  $p > .70$ , for the stimuli from Experiment 2. Across all 16 listeners, there were 97 correct answers and 95 incorrect answers. Four listeners reported using a guessing strategy, deciding to choose one side throughout to associate with "before verb" pauses, and the other side with "after verb" pauses. Of these listeners, two guessed correctly and achieved perfect scores; the other two guessed incorrectly and had no correct answers. Thus, these results strongly suggest that the live

<sup>1</sup> The fact that the looking times recorded by the live observer tend to be longer overall than those of the videotape observer may have to do with the fact that it is easier to pick up eye movements from the live observations. Such movements are harder to pick up on the videotapes because of the low light levels under which these recordings are made. Therefore, the live observer may be able to react more rapidly to changes in the infant's movements.

observers in the infant experiments could not discriminate the two versions of the stimuli over the masking music and thereby unconsciously influence the results.

The current, random-assignment version of HPP seems even less subject-to-observer biased than the fixed-assignment version. By uncoupling sample type and sample side within sessions, even a momentary hearing by an exquisitely sensitive observer can provide no useful information to transfer across trials. Nevertheless, we ran an additional experiment using the music maskers and this newer procedure. The stimuli were chosen from an ongoing study investigating infant sensitivity to word boundaries. Sixteen normal-hearing adults were tested. They were informed that there were two different types of samples, one which had pauses within (multisyllabic) words and one which had pauses between words. The participants' task was to circle which type occurred on a given trial. They were also told that samples of a given type (e.g., pause within a word) could occur on either side. As in the previous study, listeners were first tested without headphones and masking music, and then with headphones and the masker.

Performance without the maskers was again virtually perfect (one listener missed the first trial) and dropped to chance levels with the masker. There was no evidence that the participants discriminated the two types of samples under this condition,  $t(15) = 0.10, p > .90$ . The numbers of correct responses (out of a possible 12) achieved by each of the listeners were distributed as follows: 3, 4, 4, 4, 4, 5, 5, 5, 6, 6, 6, 6, 8, 9, 10, 12.<sup>2</sup> Overall, there were 97 correct and 95 incorrect answers. These results reaffirm that the music masker was effective in preventing listeners from discriminating the two types of samples, even when they were exclusively and intentionally focussed on this goal.

#### FURTHER EXTENSIONS OF THE HPP

The HPP has proved to be a valuable tool in investigations of auditory perception in infants between 4 1/2 and 12 months old. The recent practice by which stimulus types are inconsistently assigned to locations within sessions and

which no longer allows infants to select the target location on a trial has enhanced the applicability of the procedure. Participant loss due to position preferences has been eliminated. The drop-out rate in normal-hearing and otherwise healthy infants is typically only 15 to 20%, and never higher than 40%.

The current practice of counterbalancing the assignment of samples to sides within subjects also provides more valid measures of individual infants' stimulus preferences. Looking times to the two stimulus types are no longer confounded with the sides on which the types occur. In addition, sensitive within-subjects comparisons can be arranged by testing individual infants on two different versions of the same stimulus materials. For these reasons, the recent version of HPP may serve as a means of getting at individual differences, a direction in infant speech perception research that has been hardly explored. For example, we have shown that, as a group, infants of 4 1/2 months are sensitive to prosodic cues that correlate with clause boundaries (Jusczyk, 1989), and that by 9 months, they are sensitive to prosodic marking of still smaller grammatical units (Jusczyk et al., 1992). We have argued that these perceptual abilities may provide building blocks for language acquisition (Jusczyk & Kemler Nelson, in press). If so, is there a relation between individual differences in these prosodic sensitivities and the later acquisition of grammatical competence? Additional questions about individual differences in infant auditory cognition can be easily tested with HPP. For instance, is later musical ability (or early musical experience) related to infant performance in Krumhansl and Jusczyk's (1990) test of sensitivity to musical structure?

Another interesting adaptation of HPP has already been implemented. This new application moves beyond measurements of naturally occurring preferences to questions concerning the retention of spoken materials. In the new application, the training phase also becomes a familiarization phase, in which the infants are exposed to certain speech materials. Then, in the test phase, one asks whether preferences are affected by the particular materials that were heard previously.

For example, Jusczyk and Aslin (in press) investigated whether infants who heard words produced in isolation would recognize them when they occurred later in the context of sen-

<sup>2</sup> We tested the listener who had obtained a perfect score over again to determine whether she had exceptional hearing or simply made a number of lucky guesses. On the retest, she obtained a score of 4 correct.

tences. During the training phase, 7 1/2-month-olds were exposed to repetitions of two different words on alternating trials. Each sample consisted of 15 different tokens of the same word. The training phase ended when the infants had listened to each word for at least 30 s. Four different passages of fluent speech were played on the test trials. Within two passages, one of the words heard in training occurred frequently, whereas in the other two passages, no familiarized words were present. During the test phase, infants listened significantly longer to the passages containing the familiar words. This suggests that the infants remembered the words and were able to recognize them when they occurred in sentential contexts.

Such a study shows how HPP can be adapted to investigate the ability of infants to recognize units (e.g., words) which are embedded in larger objects (e.g., sentences). A logical extension is to use HPP to tap other kinds of part-whole relationships, such as the appearance of a particular phrase in different sentences. This sort of investigation, which we are currently pursuing, can reveal the nature of the units that infants use in processing fluent speech.

More generally, HPP is a viable tool to investigate infants' memory for and categorization of many kinds of speech materials or other auditory information. Essentially, HPP can come to play the same role with regard to processing of sustained auditory passages that visual preference and habituation paradigms have come to play in investigations of visual information processing in infants.

#### REFERENCES

- Colombo, J., & Bundy, R.S. (1981). A method for the measurement of infant auditory selectivity. *Infant Behavior and Development*, 4, 219-223.
- Fernald, A. (1985). Four-month-olds prefer to listen to motherese. *Infant Behavior and Development*, 8, 181-195.
- Fernald, A. (1993, February). *Effects of prosody and word position on lexical comprehension in infants*. Paper presented at the Signal to Syntax Workshop, Brown University, Providence, RI.
- Gerken, L.A., Jusczyk, P.W., & Mandel, D.R. (1994). When prosody fails to cue syntactic structure: Nine-month-olds sensitivity to phonological vs. syntactic phrases. *Cognition*, 51, 237-265.
- Hirsh-Pasek, K., Kemler Nelson, D.G., Jusczyk, P.W., Wright Cassidy, K., Druss, B., & Kennedy, L. (1987). Clauses are perceptual units for young infants. *Cognition*, 26, 269-286.
- Jusczyk, P.W. (1989, April). *Perception of cues to clausal units in native and non-native languages*. Paper presented at the meeting of the Society for Research in Child Development, Kansas City, MO.
- Jusczyk, P.W., & Aslin, R.N. (in press). Recognition of familiar patterns in fluent speech by 7 1/2 month old infants. *Cognitive Psychology*.
- Jusczyk, P.W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development*, 64, 675-687.
- Jusczyk, P.W., Friederici, A.D., Wessels, J., Svenkerud, V.Y., & Jusczyk, A.M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, 32, 402-420.
- Jusczyk, P.W., & Kemler Nelson, D.G. (in press). Syntactic units, prosody, and psychological reality during infancy. In J.L. Morgan & K. Demuth (Eds.), *Signal to syntax*. Hillsdale, NJ: Erlbaum.
- Jusczyk, P.W., Kemler Nelson, D.G., Hirsh-Pasek, K., Kennedy, L., Woodward, A., & Piwoz, J. (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology*, 24, 252-293.
- Kemler Nelson, D.G., Hirsh-Pasek, K., Jusczyk, P.W., & Wright-Cassidy, K. (1989). How prosodic cues in motherese might assist language learning. *Journal of Child Language*, 16, 55-68.
- Krumhansl, C.L., & Jusczyk, P.W. (1990). Infants' perception of phrase structure in music. *Psychological Science*, 1, 70-73.

13 July 1993; Revised 19 January 1994 ■