Published in final edited form as:

Child Dev. 1980 March ; 51(1): 107–112.

Effects of Early Linguistic Experience on Speech Discrimination by Infants: A Critique of Eilers, Gavin, and Wilson (1979)

Richard N. Aslin and David B. Pisoni
Indiana University

Abstract

In a recent report in this journal, Eilers, Gavin, and Wilson (1979) presented discrimination data obtained from 2 groups of infants exposed to different language-learning environments. The results showed differences in voice onset time (VOT) discrimination between Spanish and English infants, suggesting an effect of early linguistic experience. A critique of this study indicates that such conclusions about the effects of early experience on speech perception are unwarranted on both methodological and conceptual grounds. Methodological flaws include the absence of reliable statistical analyses and the failure to guard against experimenter bias effects. Conceptual flaws involve the erroneous interpretation of failures to discriminate certain selected speech contrasts. Inferences concerning the developmental course of speech perception in young infants based on the results of the Eilers et al. study need to be interpreted cautiously in light of these serious criticisms.

A precise description of the processes and mechanisms used by young infants to perceive the phonetic distinctions of their native language is a difficult challenge that demands careful application of reliable methods of measurement and a reasoned interpretation of empirical findings. Eilers, Gavin, and Wilson (1979) recently reported a study of speech discrimination by infants employing a cross-language design. These investigators applied a relatively new and potentially powerful operant head-turning technique to infants selected from two different language-learning environments in order to assess the effects of early linguistic experience on the discrimination of voicing in stop consonants. Although the goals and general rationale of this study were quite sound, several serious methodological and interpretative deficiencies led us to question the major conclusions and their implications for the role of early linguistic experience in speech discrimination. The purpose of this report is to point out these problems and also caution other investigators who might follow a similar strategy in using this methodology to study perceptual development in young infants.

Our criticisms of this paper focus on several methodological and interpretative problems, including (a) the number of test trials used to measure discrimination and the statistical analysis of the data, (b) the criterion selected to assess the “state” or degree of “attentiveness” of the infant, (c) the potential presence of strong experimenter bias effects in collecting the data, and (d) the elimination of potentially important within-subject discrimination data by presenting only grouped means.

The major purpose of the Eilers et al. study was to measure voice onset time (VOT) discrimination for several selected contrasts that are phonologically distinctive in either Spanish or English but not both languages. Based on perceptual data obtained from adult subjects, Eilers et al. selected two critical test pairs for use in their infant study. One

stimulus contrast, called the “lag” pair (+ 10 vs. +40 msec VOT), was distinctive for adult English subjects, whereas the other contrast, the “lead” pair (+ 10 vs. – 20 msec VOT), was distinctive for adult Spanish subjects. The test stimuli, which differed on VOT, were originally produced on a speech synthesizer at Haskins Laboratories and have been used in earlier investigations of voicing perception in adults and infants (see Eimas, Siqueland, Jusczyk, & Vigorito 1971; Lasky, Syrdal-Lasky, & Klein 1975; Lisker & Abramson 1970; Streeter 1976).

In the Eilers et al. study, discrimination of each pair of test stimuli was measured for each infant in an operant head-turning procedure (Eilers, Wilson, & Moore 1977; Moore, Thompson, & Thompson 1975; Moore, Wilson, & Thompson 1977). In this procedure the infant is reinforced for a head-turn response toward a loudspeaker when a change occurs in a sequence of repetitive background stimuli. An experimental trial consists of a change from the background stimulus (S1) to a target stimulus (S2). A control trial, on the other hand, consists of no change from the background to target stimulus. A correct response is defined as the presence of a head-turn response on experimental trials and the absence of a head-turn response on control trials. A visual reinforcement is only presented for correct head-turn responses on experimental trials. According to Eilers et al., a measure of discrimination performance can be obtained by simply computing the overall percent correct responses for both experimental and control trials.

In their study, infants received both experimental and control trials for each of the two VOT contrasts, the lag pair and the lead pair. The main hypothesis tested in the study was whether early linguistic experience would differentially affect the likelihood of discriminating the two VOT contrasts. If early linguistic experience is a necessary prerequisite for VOT discrimination, then infants should discriminate only the stimulus pair that is distinctive in their language-learning environment. That is, the Spanish infants should discriminate only the lead VOT contrast (i.e., +10 vs. –20 msec) and the English infants should discriminate only the lag VOT contrast (i.e., + 10 vs. +40 msec). Eilers et al. reported that the English infants discriminated only the lag contrast, while the Spanish infants discriminated both the lead and the lag contrasts. They attributed these results to an effect of linguistic experience on speech perception. That is, experience in the language-learning environment was presumably responsible for the differences observed in discrimination between the two groups of infants.

The first criticism of this study concerns the number of test trials that were collected for both control and experimental conditions and the subsequent statistical analysis that was carried out on these data. The probability of observing any positive evidence of discrimination is dependent, to a large degree, on the minimum number of test trials employed in the experiment. Measures of discrimination performance, whether based on traditional threshold procedures or more sophisticated methods involving signal detection analysis, require that a sufficient number of trials be collected in order to rule out the possibility that chance factors may have inadvertently affected the observations. In the study reported by Eilers et al., only three experimental and three control trials were collected for each stimulus contrast with each infant tested in the experiment. Estimates of the likelihood of discrimination of either the lag or lead pair were therefore based on only six responses per subject. A criterion of five out of six correct responses was employed in deciding whether each infant discriminated a particular VOT contrast. Unfortunately, this five-out-of-six criterion is not statistically significant at the .05 level according to either the binomial expansion, Fisher’s exact test, McNemar’s test of correlated proportions, or a χ^2 test based on a 2×2 contingency table.¹ Six out of six correct responses would be a statistically acceptable criterion according to these procedures, but undoubtedly several infants included in the Eilers et al. results did not meet this more stringent criterion. Furthermore, infants who

failed to meet the five-out-of-six criterion were retested on an “easier” /bit-/bIt/ contrast. If the infant passed the five-out-of-six criterion on this vowel contrast, then the below-criterion performance on the VOT contrast was accepted as a “real” failure to discriminate. In our opinion, the logic of this procedure is clearly flawed. Not only are some infants providing false evidence of discriminating the VOT contrast and thus inflating their percent correct scores, but other infants are being falsely judged as having discriminated the vowel contrast and thus deflating their low percent correct scores on the VOT contrast. At the very least, infants who passed the criterion on the VOT contrast should also have been posttested on the vowel contrast. Without that posttest, we must conclude that both the evidence for VOT discrimination and the evidence for a failure to show VOT discrimination are inconclusive at best.

The interpretation of negative results is particularly important in a study of this kind because the authors’ argument concerning the role of early experience in speech perception rests on establishing that English infants fail to discriminate a contrast in voicing that is non-distinctive to adult speakers of English. Empirical investigations that attempt to “prove” the null hypothesis need extra precautions to insure that their conclusions can be justified on both methodological and statistical grounds. Unfortunately, the results reported by Eilers et al., because they were based on such a small number of test trials, do not provide convincing evidence that English infants cannot discriminate a particular voicing contrast. For the same reason, one could also argue that the evidence cited in support of infants’ discriminating the other VOT contrasts is also unreliable.

Furthermore, there are additional problems with the particular statistical procedures that were applied to the data. Eilers et al. applied an analysis of variance to their data. But analysis of variance, a parametric statistical test, cannot be used with the data collected in this study because the scores simply do not meet the acceptable criteria required for interval data. With such a small number of trials, the underlying scale could hardly be considered continuous, varying over the entire range of values in the scale. The data only meet the criteria for an ordinal scale at best. Moreover, it is highly unlikely that either the normality assumption or the requirement of homogeneity of variances could be met.

A second criticism of this study deals with the procedures used by Eilers et al. to assess the state or attentiveness of an infant during the course of an experimental session. As the authors note, failure to discriminate a difference between two test stimuli, particularly stimulus pairs that may be nondiscriminable, could be a “real” effect due to limitations of the infants’ perceptual abilities. However, the result could also be due to the attentional state of the infant. To check on an infant’s responsiveness in this experiment, Eilers et al. introduced what they believed to be a more easily discriminable contrast, a distinction based on vowel quality between “bit” and “beat.” Eilers et al. reasoned that by using a more discriminable contrast such as this, they could determine whether the infant was still “on task” or whether there was a momentary or possibly more lasting change in the state of the infant during the testing session. However, the critical comparisons of interest in this experiment involved whether “fine” discriminative abilities involving the detection of VOT are present in young infants, particularly infants selected from different language-learning environments. Unfortunately, the specific stimulus contrast selected to check on task responsiveness in this study focused the infant’s attention on an entirely new and irrelevant stimulus dimension, a dimension involving primarily the detection of spectral (i.e.,

¹Hays (1973) describes each of these four tests (pp. 735–742; Table II, pp. 880–884). Although a simple χ^2 test on five out of six trials is significant at the .05 level, the low expected cell frequency (<5) in a 2×2 table containing only six trials demands a correction for continuity. This correction (Hays 1973, p. 735) results in the failure of five out of six correct to reach the .05 level of significance. In addition, the trials must also be independent, which in this case they are not. According to the binomial expansion tables, a minimal criterion is eight out of 10 ($p = .055$), although clearly 15 out of 20 trials would be advisable.

frequency specific) rather than temporal (i.e., VOT) differences between stimuli. Although there is no universal measure of how well the infant is “on task,” one certainly should attempt to direct the infant’s attention to the relevant dimension under consideration. The simplest way to accomplish this is to increase the difference in VOT between the background and test stimuli. By introducing a contrast that required the detection of spectral differences between vowels, the infants may have shifted their attention from the particular dimension under investigation (i.e., VOT) to another dimension (e.g., vowel color, stimulus duration, etc.) during the remainder of the test session. Such shifts in attention could easily result in a modification of the infant’s criterion for initiating a head-turn response on experimental trials during the course of an experiment, and could well have affected the likelihood of correctly detecting a contrast involving small differences in VOT on subsequent trials. It is interesting to note here that Spanish does not have a contrast between the vowels [i] and [I]. Thus, if early experience does influence the infants’ discriminative capacities at this age, a systematic bias would have been introduced by using these particular vowel contrasts to assess attentional state.

Taken together with the earlier criticism concerning the small number of test trials, the procedures used to assess the “state” of the infant could have biased the outcome of the experiment toward finding negative evidence of VOT discrimination in these infants. Both of these criticisms, however, cannot account for the finding that the infants from the Spanish-speaking environment discriminated both of the VOT contrasts, whereas infants from the English-speaking environment discriminated only the lag contrast. This result could simply be due to chance or experimenter bias and may have nothing whatsoever to do with early environmental experience.

Our third criticism of this study relates to the possibility of strong experimenter bias effects operating in the version of the head-turning procedure used by Eilers et al. According to the description provided in the methods section, two independent scorers, one located in the testing booth with the infant (E2), and one controlling the stimulus presentations from outside the booth (E1), recorded whether a head-turn response occurred during each of the six test trials. The authors state that E2 was “blind” to the particular type of trial presented to the infant, since this experimenter, as well as the parent, listened to “masking music” over headphones while the experiment was in progress. However, E1, the experimenter located outside the experimental booth, apparently was not blind to the type of stimulus contrast presented on each trial, since no information was provided in this research report about the procedures taken to insure that experimenter bias was eliminated. Since E1 initiated all trials and therefore knew before each trial began whether an experimental or control trial would be presented, it is possible that E1 could have biased the outcome in the following manner. If E1 initiated an experimental trial when the infant was inattentive to E2 (i.e., the person in the booth playing with toys to attract the infant’s gaze), then the infant would be more likely to turn away, and consequently that head turn would be scored as a correct response. Similarly, if E1 initiated a control trial while the infant was highly attentive to E2, then the likelihood of a head-turn response on control trials would be reduced substantially. As a result of the procedure that Eilers et al. used, the possibility of strong scorer bias effects appears quite high, especially since no provision was made, as far as we know, to insure that both scorers were “blind” to the language-learning environment of the individual infants tested or the type of contrast presented on each trial.

The fourth and final methodological criticism of this study concerns the presentation of only group data in each of the conditions of the experiment. The operant head-turning procedure is unique among currently available methods for assessing speech discrimination in young infants because highly reliable within-subject data on specific stimulus contrasts can be obtained, provided, of course, that a sufficient number of test trials are obtained from each

infant. The absence of individual subject data, taken together with the small number of data points, is especially critical in light of the claims made by these investigators about the potential role that early experience plays in perceptual development. We turn to these considerations in the final section to show how the conclusions of the Eilers et al. study are flawed on both logical and conceptual grounds.

The major interpretive problem with the study is the claim that “the present results suggest that the Spanish adult lead boundary *is* precisely the boundary to which infants’ perception is tuned through experience with the Spanish language” (p. 17). On close inspection this statement is, however, filled with contradictions and misunderstandings. First, the fact that pre-linguistic infants from a Spanish-speaking environment also discriminated the VOT contrast that spanned the adult English lag boundary is simply ignored in the discussion of the results. Why should Spanish infants discriminate both VOT contrasts and English infants discriminate only one contrast that was tested in the study? No consideration was given to this rather curious asymmetry in the discrimination data.

Second, the authors claim that “linguistic listening experience may be a necessary pre-requisite for the acquisition of lead VOT contrasts in infants” (p. 17). But this argument is clearly at variance with the previous cross-language investigations of voicing discrimination reported by Lasky et al. (1975) and Streeter (1976), who found evidence for discrimination of a lead contrast in VOT by both Spanish and Kikuyu infants despite the fact that the specific distinction was nondistinctive in the adult phonological system that the infants were exposed to in the language-learning environment. Apparently, Eilers et al. seem to believe that all infants can innately discriminate the lag boundary without any early linguistic experience, but only those infants who actually “hear” the lead contrast used in the postnatal period will acquire the ability to discriminate these VOT contrasts. This conceptualization is especially curious in light of the observation made in the introduction to their paper that “English-learning infants are normally exposed to prevoiced (or lead) stop consonants by English speakers...” (p. 14).

How do infants from English-speaking environments learn to selectively attend to precisely those acoustic attributes that are distinctive in the adult phonological system? Eilers et al. propose an explanation for this anomaly by suggesting that infants “must be engaged in an active analysis of the frequency of occurrence of speech sounds at various points along phonetic continua long before these sounds are associated with arbitrary linguistic meanings” (p. 17). But if frequency of occurrence were the mechanism responsible for improving discrimination of the lead contrasts, then why do Spanish infants discriminate stimuli in the voicing lag region—stimuli that are not contrastive in the production of stops by Spanish-speaking adults? An alternative account of the Eilers et al. results is possible based on the fact that stimuli in the voicing lead region of the VOT continuum are known to be less discriminable than stimuli in the short lag region.² Data supporting the salience of the lag over the lead region are now quite overwhelming (see Aslin, Pisoni, Hennessy, & Perey 1979; Lisker & Abramson 1967; Stevens & Klatt 1974; Williams, Note 1; Aslin & Pisoni, Note 2). In essence, Eilers et al. have overlooked several serious methodological deficiencies in their attempts to provide support for the null hypothesis that English infants cannot discriminate one particular VOT contrast. Their entire study rests on the validity and replicability of their failure to show that English infants discriminate the prevoiced-voiced contrast. A single case of an English infant who could discriminate that contrast would offer

²This asymmetry may result from the fact that in stops with voicing lead, the first formant has a relatively low amplitude and the primary cue for discrimination is the duration of the prevoicing. In contrast, there are several additional and potentially more salient cues for the discrimination of voicing in the lag region, including the presence or absence of an F1 transition, the onset of F1 relative to higher formants, and the presence of aspiration noise after the release from stop closure.

strong counterevidence for their claim that contrastive use of the voicing lead contrast during early life is a necessary requirement for the acquisition of that discriminative ability. The report of only group means leads one to ask whether some of the infants from the English-speaking environment did, in fact, show evidence of discriminating the lead contrast. In our laboratory, using a similar operant head-turning procedure, but with rigorous controls for experimenter bias effects, we have obtained highly reliable within-subject data indicating that infants from an English-speaking environment can, in fact, discriminate lead contrasts in VOT (see Aslin, Hennessy Pisoni, &, Perey, Note 3). These results, along with our earlier methodological criticisms, indicate that the effects of early experience on speech discrimination have been grossly overestimated by Eilers et al.

In summary, we believe the findings of their cross-linguistic study are seriously flawed on methodological and procedural grounds and that their interpretation of the results is unwarranted. Any inferences or conclusions concerning the developmental course of speech perception in young infants based on the results of the Eilers et al. study should be interpreted cautiously by investigators interested in the effects of early experience on perceptual development.

Acknowledgments

The preparation of this paper was supported, in part, by NICHD grant HD-11915-01, NINCDS grant NS-12179-04, and NIMH grant MH-24027-05 to Indiana University, Bloomington. The paper was written while the second author was a Guggenheim fellow at the Research Laboratory of Electronics, M.I.T. We thank Susan Dumais, Peter Jusczyk, Pat Kuhl, and, Louis Goldstein for helpful comments on an earlier version of the paper.

Reference Notes

1. Williams, CL. Unpublished doctoral dissertation. Harvard University; 1974. Speech perception and production as a function of exposure to a second language.
2. Aslin, RN.; Pisoni, DB. Some developmental processes in speech perception. Paper presented at the NICHD conference on Child Phonology: Perception, Production and Deviation; Bethesda, Maryland. May 28–31, 1978;
3. Aslin, RN.; Hennessy, B.; Pisoni, DB.; Perey, AJ. Individual infants' discrimination of voice onset time: evidence for three modes of voicing. Paper presented at the Biennial Meetings of the Society for Research in Child Development; San Francisco. March 1979;

References

- Aslin, RN.; Pisoni, DB.; Hennessy, B.; Perey, AJ. Identification and discrimination of a new linguistic contrast. In: Wolf, JJ.; Klatt, DH., editors. Speech communication papers presented at the 97th meeting of the Acoustical Society of America; New York: Acoustical Society of America; 1979.
- Eilers RE, Gavin W, Wilson WR. Linguistic experience and phonemic perception in infancy: a cross-linguistic study. *Child Development*. 1979; 50:14–18. [PubMed: 446199]
- Eilers RE, Wilson WR, Moore JM. Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Disorders*. 1977; 20:766–780.
- Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. Speech perception in infants. *Science*. 1971; 171:303–306. [PubMed: 5538846]
- Hays, WL. Statistics for the social sciences. 2. New York: Holt, Rinehart & Winston; 1973.
- Lasky RE, Syrdal-Lasky A, Klein RE. VOT discrimination by four to six and a half month old infants from Spanish environments. *Journal of Experimental Child Psychology*. 1975; 20:215–225. [PubMed: 1194843]
- Lisker, LL.; Abramson, AS. The voicing dimension: some experiments in comparative phonetics. *Proceedings of the 6th International Congress of Phonetic Sciences*; Prague: Academia; 1970.

- Moore JM, Thompson G, Thompson M. Auditory localization of infants as a function of reinforcement conditions. *Journal of Speech and Hearing Disorders*. 1975; 40:29–34. [PubMed: 1123925]
- Moore JM, Wilson WR, Thompson G. Visual reinforcement of head-turn responses in infants under 12 months of age. *Journal of Speech and Hearing Disorders*. 1977; 42:328–334. [PubMed: 881813]
- Stevens KN, Klatt DH. Role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*. 1974; 55:653–659. [PubMed: 4819867]
- Streeter LA. Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*. 1976; 259:39–41. [PubMed: 1256541]

\$watermark-text

\$watermark-text

\$watermark-text