Published in final edited form as:

Cognition. 1992 June ; 43(3): 253–291.

Some consequences of stimulus variability on speech processing by 2-month-old infants*

Peter W. Jusczyk,

Department of Psychology, State University of New York at Buffalo, Buffalo, NY 14260, USA and
Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS, 54, Boulevard Raspail, 75270
Paris, France

David B. Pisoni, and

Department of Psychology, Indiana University, Bloomington, IN 47405, USA

John Mullennix

Department of Psychology, Wayne State University, Detroit, MI 48202, USA

Abstract

The present study explores how stimulus variability in speech production influences the 2-month-old infant's perception and memory for speech sounds. Experiment 1 focuses on the consequences of talker variability for the infant's ability to detect differences between speech sounds. When tested with high-amplitude sucking (HAS) procedure, infants who listened to versions of a syllable, such as /bʌg/, produced by 6 male and 6 female talkers, detected a change to another syllable, such as /dʌg/, uttered by the same group of talkers. In fact, infants exposed to multiple talkers performed as well as other infants who heard utterances produced by only a single talker. Moreover, other results showed that infants discriminate the voices of the individual talkers, although discriminating one mixed group of talkers (3 males and 3 females) from another is too difficult for them. Experiment 2 explored the consequences of talker variability on infants' memory for speech sounds. The HAS procedure was modified by introducing a 2-min delay period between the familiarization and test phases of the experiment. Talker variability impeded infants' encoding of speech sounds. Infants who heard versions of the same syllable produced by 12 different talkers did not detect a change to a new syllable produced by the same talkers after the delay period. However, infants who heard the same syllable produced by a single talker were able to detect the phonetic change after the delay. Finally, although infants who heard productions from a single talker retained information about the phonetic structure of the syllable during the delay, they apparently did not retain information about the identity of the talker. Experiment 3 reduced the range of variability across talkers and investigated whether variability interferes with retention of all speech information. Although reducing the range of variability did not lead to retention of phonetic details, infants did recognize a change in the gender of the talkers' voices (from male to female or vice versa) after a 2-min delay. Two additional experiments explored the consequences of limiting the variability to a single talker. In Experiment 4, with an immediate testing procedure, infants exposed to 12 different tokens of one syllable produced by the same talker discriminated

*Portions of the research reported here were presented in a paper at the *Workshop on Spoken Language* at the State University of New York at Buffalo, on May 22, 1989. The research reported here was supported by grants from NICHD (#15795) to P.W.J., and by grants NIDCD (#00012 and #00111) to D.B.P. Jacques Mehler, Josiane Bertoncini and Ranka Bijeljac-Babic made a number of useful suggestions about the design of some of the experiments reported here. In addition, the authors would like to extend their gratitude to Douglas Hintzman, Deborah Kemler Nelson, and Ann Marie Jusczyk and the reviewers for helpful comments made on earlier versions of the manuscript. Finally, we thank Eric Bylund, Tracy Schomberg, Nan Koenig, Ann Marie Jusczyk and Tara DeLeon for their help in running these experiments.

Correspondence to: P.W. Jusczyk, Department of Psychology, State University of New York at Buffalo, Buffalo, NY 14260, USA.

these from 12 tokens of another syllable. In addition, there was no evidence that infants detected differences between two different tokens of the same syllable. However, Experiment 5, using a delayed testing procedure, yielded different results. When multiple versions of the same syllable were present, infants did not discriminate the phonetic change that occurred after the delay period, even though they discriminated the same change when only single versions of the syllables were used in each phase of the experiment. These results are discussed in the context of the development of word recognition and the mental lexicon and the role of stimulus variability in speech perception.

Introduction

One important aspect of language acquisition that begins to unfold during the first year of life is the development of a lexicon in the native language. Just as children begin to produce their first words towards the end of the first year, so too do they begin to understand words from their native language (e.g., Huttenlocher, 1974). Comprehending spoken words from caretakers in the environment requires that infants store away some representation of the sound structure of the word so that they can retrieve the appropriate meaning. A number of the prerequisites necessary for the successful storage of words, and hence for the development of the mental lexicon, have been studied during the last 20 years of research on infant speech perception. For example, the capacities of infants to discriminate subtle phonetic distinctions have been well documented (e.g., Aslin, 1987; Aslin, Pisoni, & Jusczyk, 1983; Eimas, 1982; Jusczyk, 1981; Kuhl, 1987; Werker, 1990). In addition, a number of studies have shown that, by 6 months, infants are apparently able to ignore the variability in the speech signal introduced when the same item is uttered by different talkers (e.g., Kuhl, 1979, 1983). This latter ability is critical for the recognition of the same word spoken by different individuals. Some recent work also demonstrates that even newborn infants apparently have some minimal capacity to represent different speech sounds (Bertoncini, Bijeljac-Babic, Jusczyk, Kennedy, & Mehler, 1988; Jusczyk, Bertoncini, Bijeljac-Babic, Kennedy, & Mehler, 1990). Nevertheless, many other important factors related to lexical development have yet to be explored. For example, very little is known about the capacity of infants to retain information about the speech sounds that they hear, or, indeed, about factors that might affect the retention of information about the sound properties of words. Information about such issues is critical for understanding the way that the lexicon is structured and how it develops.

One factor known to affect the way that adults encode speech sounds is talker variability. Although adults are able to adjust to differences in talkers' voices in perceiving speech sounds (e.g., Bladon, Henton, & Pickering, 1984; Dechovitz, 1977; Disner, 1980; Fourcin, 1968; Gerstman, 1968; Nearey, 1978; Rand, 1971; Summerfield, 1975; Syrdal & Gopal, 1986; Verbrugge, Strange, Shankweiler, & Edman, 1976), these kinds of adjustments are not without consequences for perceptual processing. Thus, the accuracy with which items are identified suffers when the talker's voice varies as opposed to when it remains constant throughout a block of trials (Creelman, 1957; Fourcin, 1968; Verbrugge et al., 1976). Similarly, the latencies required to perform identification (e.g., Summerfield, 1975; Summerfield & Haggard, 1973) and matching tasks (Allard & Henderson, 1976; Cole, Coltheart, & Allard, 1974) have been shown to increase significantly when listeners are required to adjust to different talkers' voices. With respect to speech perception, the consequences associated with adjusting to different talkers' voices appear to be confined to early stages of acoustic-phonetic processing as opposed to higher-level word recognition stages (Mullennix, Pisoni, & Martin, 1989). Thus, Mullennix et al. showed that talker variability interacts with variables that affect acoustic-phonetic encoding (such as the

presence of envelope- shaped noise) but not with variables that affect higher-level word recognition processes (such as lexical density and word frequency).

There is also evidence that talker variability can affect cognitive processes other than those involved in perception. For instance, Martin, Mullennix, Pisoni, and Summers (1989) found that memory processes in *recall tasks* were adversely affected when listeners had to cope with talker variability. In particular, their results suggested that both encoding processes and the efficiency of rehearsal processes used to transfer items into long-term memory were disrupted by talker variability. At the same time, there is reason to believe that if the information gets into long-term memory talker variability can actually improve memory performance on certain tasks. Thus, Craik and Kirsner (1974) found that talker variability could actually have a beneficial effect on *recognition memory* of items presented in a list. They exposed subjects to lists of items spoken by different talkers. Their subjects were faster and more accurate in recognizing words from a previously heard list when the words were repeated in the same voice as in the original recording than in a different voice. In a more recent study, Goldinger, Pisoni, and Logan (1991) found that talker variability can either help or hinder recall performance - the outcome is dependent on rate of presentation. At fast presentation rates, talker variability interferes with rehearsal and encoding, preventing subjects from using elaboration techniques. At slow presentation rates, subjects have more time to encode the stimulus items and elaborate on them in rehearsal. The elaboration provides additional talker-specific cues in long-term memory that can be used at the time of retrieval to improve recall performance.

In summary, despite the fact that adult listeners are able to adjust fairly rapidly to talker variability, there are indications that some costs are associated with the process. These costs show up both with respect to the initial perceptual processing of the signal and its subsequent encoding into long-term memory. It is also clear that information about talker differences is detected in perception and may be retained in memory. Although costs may be incurred in terms of the amount of information that can be encoded when talker variability occurs, there are some benefits as well for subsequent recognition of the items that are encoded into long-term memory.

As noted earlier, there is evidence that, by 6 months of age, infants display some basic ability to cope with talker variability. Kuhl (1979, 1983) showed that 6-month-olds will continue to detect a phonetic contrast in the face of changes in speaking voices that range from children to adults, and include both males and females. Thus, infants trained to distinguish a contrast between two vowel tokens produced by a single talker successfully generalized this distinction to vowel tokens produced by different talkers, even when there was considerable acoustic overlap among the tokens of the distinctive vowel classes (Kuhl, 1983). Of course, if infants succeed in this task because they are simply unable to distinguish differences between talkers' voices, then their achievement would not be very remarkable. However, there is ample evidence to believe that this is not the case. Studies with newborn infants show them to be capable of recognizing their mothers' voices from those of other mothers (e.g., DeCasper & Fifer, 1980; Mehler, Bertoncini, Barriere, & Jassik-Gerschenfeld, 1978; Mills & Meluish, 1974). Moreover, 6-month-olds are able to perform a task that requires responding to tokens produced by a particular talker as opposed to another talker (Miller, Younger, & Morse, 1982).

Nevertheless, whether the infant's success at coping with talker variability also affects perceptual processing and memory has not been directly addressed in previous research. In fact, aside from the two studies by Kuhl (1979, 1983), the only attempt to focus on the way that infants handle irrelevant variation in speech sounds was a study by Kuhl and Miller (1982) with infants 1–4 months of age. Kuhl and Miller used synthetic vowel stimuli and

examined the capacity of infants to detect a change in one dimension when a second dimension varied irrelevantly. The two dimensions were pitch and vowel quality. Their results indicated that when pitch varied irrelevantly, the infants were able to detect a vowel change. However, the converse did not hold. Namely, when infants were exposed to a series of randomly alternating vowels, /a/ and /i/, they were not able to detect a subsequent change in pitch contour. Consequently, Kuhl and Miller interpreted this as an indication that the vowel quality dimension was more salient for the infants and distracted them from detecting the change in pitch quality. As further support of their interpretation, Kuhl and Miller noted that infants took significantly longer to habituate to the stimuli when vowel quality varied irrelevantly than when pitch quality varied, suggesting that infants attended more to the vowel variation than to the pitch variation (however, see Carrell, Smith, & Pisoni, 1981).

Kuhl and Miller's findings are an interesting demonstration that irrelevant variation along some dimension may hinder infants from detecting a change along another less salient dimension. Still, it is hard to predict whether the type of variation introduced by alternating different vowels is of the same order of magnitude as one stemming from the presence of different talkers (Carrell et al., 1981). In fact, the studies with older infants indicate that coping with talker variability does not prevent infants from discriminating a contrast between different vowels (Kuhl, 1979, 1983). Nevertheless, what is not known at present is the extent to which infants may incur the kinds of subtle costs in processing and encoding speech that have been reported for adult listeners when dealing with talker variability (e.g., Martin et al., 1989; Mullennix et al., 1989). Information about the way that talker variability influences speech processing by infants is important not only for determining how the lexicon develops, but also for understanding the mechanisms that underlie the process of perceptual normalization. For instance, it has been suggested that normalization may operate at early stages of speech processing in a mandatory fashion, independently of higher-level cognitive processes (Miller, 1987; Mullennix et al., 1989). If so, then in line with other features that are associated with modular systems, one might expect to find that the characteristics of the normalization system are innately wired and fixed. Hence, strong parallels would be predicted for the way that talker variability influences perceptual processing in infants and adults.

It was with these broad issues in mind that the present investigation was undertaken. Accordingly, we designed a series of experiments to evaluate the way that talker variability affects the processing and retention of speech sounds by 2-month-old infants. Our first experiment focused on potential effects of talker variability on the perception of a speech contrast.

EXPERIMENT 1

Previous investigations of infants' capacity for dealing with talker variability (e.g., Kuhl, 1979, 1983) have focused on infants 6 months of age. Hence, not much is known about the capacity of younger infants to cope with talker variability.¹ Consequently, we decided to use a modified version of the high-amplitude sucking (HAS) procedure to explore this capacity in 2-month-olds. The speech contrast that was selected involved a change in the initial consonant of two CVC syllables, /bʌg/ and /dʌg/ (corresponding to the English words "bug" and "dug"). The stimulus tokens for the study were chosen from 6 adult male and 6 adult female talkers who produced them originally for use in the Mullennix et al. (1989) study. It should be noted that previous studies (Kuhl, 1979, 1983) have typically used contrasting vowel stimuli, such as [a] and [i]. We decided to use a more subtle contrast between [b] and

¹Marean, Werner, and Kuhl (in press) have recently reported research with a new paradigm suggesting that 2-month-olds are capable of generalizing from the vowels of one talker to those of other talkers.

[d], in part because the acoustic properties of these segments are influenced by their surrounding context (e.g., Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Hence, if talker variability disrupts speech processing in 2-month-olds, it might be more readily observed with subtle contrasts, such as the one between [b] and [d].

Determining the consequences of talker variability requires comparisons between conditions involving single-talkers versus conditions involving multiple- talkers. For this reason, we decided to examine the same contrast in both the single- and multiple-talker comparisons. Hence, one experimental and one control group were tested with tokens from a single-talker. The experimental group was habituated to one of the syllables, either /bʌg/ or /dʌg/, and was presented with the remaining syllable during the test phase. The control group was habituated to one of the two syllables and continued to hear the same syllable during the test phase. There were two comparable multiple-talker conditions. The only difference was that tokens from all 12 talkers were used during both the habituation and test phases of the experiment. By comparing the performance of the infants in the multiple-talker conditions with that of the infants in the single-talker conditions, we could evaluate the consequences of talker variability on the 2-month-old's capacity to detect a phonetic change. We hypothesized that any increase in processing load associated with the multiple-talker condition might show itself either in discrimination performance or in the time that it took for infants to habituate to the syllable(s) during the first phase of the procedure (much as Kuhl and Miller reported for their study).

Two additional conditions were included in the study. First, to determine whether the tokens from the different talkers were discriminable for the infants, we tested another group of infants on a contrast not involving a phonetic change, but rather a difference between two talkers. Several previous studies have examined the ability of infants to detect differences in talkers' voices, but these studies used speech samples longer than a single syllable (e.g., DeCasper & Fifer, 1980; Kaplan, 1969; Turnure, 1971). In the present case, the syllable type (e.g., /dʌg/) was the same for both phases of the experiment, but the identity of the talker was changed after habituation to the first syllable. The final test condition was one that involved habituating the infants to tokens of a particular syllable type (e.g., /bʌg/) spoken by a set of 6 different talkers (3 males and 3 females). Then, following habituation to these tokens, the infants were switched to an entirely new set of talkers (3 males and 3 females) uttering the same syllable. The purpose of this last condition was to observe possible limitations on infants' abilities to encode information about talker identity. Thus, to discriminate the contrast in this last condition successfully, infants would have to encode the syllables according to the identity of the talker and retain this information for comparison with the new tokens presented after habituation. Previous work with infants at this age indicates that they are capable of representing information about the *phonetic* content of syllables (e.g., /bi/, /ba/, /bʌ/) so as to detect the presence of new syllable types (e.g., /bu/) presented after habituation (e.g., Bertoncini et al., 1988; Jusczyk & Derrah, 1987; Jusczyk et al., 1990). However, little is known about whether information about talker identity is also included in their representations of these syllables in memory.

Method

Procedure—Each 2-month-old was tested individually in a small laboratory room. The infant was placed in a reclining chair facing a blank wall approximately 1 m away. An image of flowers was projected on the wall for the entire test session. The picture was situated just above a loudspeaker through which the test stimuli were played. Each infant sucked on a blind nipple held in place by an experimenter who wore headphones and listened to recorded music throughout the test session. A second experimenter in an adjacent room monitored the test apparatus.

The experimental procedure was a modification of the high-amplitude sucking technique (Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Jusczyk, 1985b; Siqueland & DeLucia, 1969). For each infant, the high-amplitude sucking criterion and the baseline rate of high-amplitude sucking were established prior to the presentation of any test stimuli. The criterion for high-amplitude sucking was adjusted to produce rates of 15–35 sucks/min. After a baseline rate was established, the presentation of stimuli was made contingent on the rate of high-amplitude sucking. Criterion sucks resulted in the presentation of one speech syllable. For infants in the single-talker conditions, the same syllable was presented throughout the preshift phase of the experiment. For infants in the multiple-talker conditions, the syllable token that the infant heard at a given moment was unpredictable because the syllables were selected at random from a set stored on computer disk. Thus, it was possible that an infant in a multiple-talker condition might hear the same syllable or a different one for successive criterion sucks. The maximum stimulus presentation rate was one syllable per second. If the infant produced a burst of sucking with inter-response times less than 1 s, then each response did not produce one presentation of a stimulus. Instead, the timing apparatus was reset so as to provide continuous auditory feedback for 1 s after the last response of the sucking burst. In any case, if the 1-s period would have terminated in the middle of a syllable, it was delayed until the syllable was completed.

The criterion for habituation during the preshift phase of the experiment was a decrement in sucking rate of 25% or more over two consecutive minutes compared with the rate in the immediately preceding minute. At this point, the auditory stimulation was changed to that which was appropriate for the postshift phase of a given condition. For infants in the experimental conditions, this resulted in a change in the stimuli presented. Infants in the control conditions continued to hear the same stimuli as before. The postshift phase began with the presentation of the first stimulus after the habituation criterion had been achieved. The infants' sensitivity to changes in auditory stimulation was inferred from comparisons of response rates of subjects in the experimental and control conditions during the postshift period. The postshift period lasted for at least 4 min or until the infant showed a 25% decrease in sucking for two consecutive minutes.

Stimuli—The stimuli consisted of natural tokens of the syllables /bʌg/ and /dʌg/ produced by 6 male and 6 female talkers. The stimulus items were selected from a larger set of pre-recorded words obtained from 15 talkers with a midwestern dialect. During the recordings, each stimulus item appeared on a CRT screen in front of the talker, embedded in the carrier sentence, “Say the word _____ for me”, where the blank corresponded to a particular target word. The talker was instructed to read the entire sentence in a normal voice at a constant speaking rate. The stimuli were recorded on audio tape in a sound-attenuated booth using an Electro-Voice Model D054 microphone and a Crown 800 series tape recorder. The utterances were subsequently digitized via a 12-bit analog-to-digital converter and stored on a PDP 11/34 computer at the Speech Research Laboratory at Indiana University. The digitized versions of these stimuli were copied on floppy disk and transferred to a PDP 11/73 computer at the Speech Perception Laboratory at the University of Oregon. The stimuli were converted to analog form in real-time via a 12-bit digital-to-analog converter. They were accessed directly during the course of the experiment and played out through a 4.8-kHz low-pass filter. All of the stimuli used in the experiment had been previously tested for intelligibility using a group of adult listeners at Indiana University. The items received identification scores of 95% correct or above when presented in isolation.

Design—Each infant was seen for one experimental session. Twelve subjects were randomly assigned to each of 6 test conditions (see Table 1 for examples of what infants might typically hear in each condition). During the preshift phase of the experiment, infants in the three single-talker conditions were exposed to repetitions of an utterance of either /

bΛg/ or /dΛg/ selected from one of the 12 different talkers. Half of the infants in each condition heard /dΛg/ and the other half, /bΛg/. To ensure that we had not selected the most discriminable pairs from one of our talkers, each infant in each control and experimental group was tested with tokens from a different talker. The infants in the single-talker control group heard the same token during both the habituation and postshift phase of the experiment. Infants in the single-talker phonetic change condition heard one of the two syllables from a particular talker during the habituation phase (e.g., /bΛg/) and the other one (e.g., /dΛg/) during the postshift phase. Infants in the single-talker talker change condition heard one syllable (e.g., /dΛg/ from male 1) from a particular talker during the habituation phase and a phonetically identical syllable (i.e., /dΛg/ from male 5) taken from a *different talker* of the same gender during the postshift phase. For half of the infants, the syllable type was /bΛg/ and for the other half it was /dΛg/. Similarly, for half the infants, the tokens were produced by female talkers, and for the other half they were produced by males.

The multiple-talker conditions were roughly parallel to the single-talker conditions. The multiple-talker control condition was identical to the single-talker control except that the tokens of a particular syllable type (e.g., /bΛg/) from all 12 talkers were presented in random order during each phase of the experiment. The multiple-talker phonetic change condition heard tokens of a particular syllable (e.g., /dΛg/) from all 12 talkers during the habituation phase and were switched to the multiple tokens of the other syllable type (e.g., /bΛg/) during the postshift phase. Half of these infants heard /bΛg/ and the others heard /dΛg/. Finally, the multiple-talker talker change condition consisted of a habituation phase in which infants heard tokens of a particular syllable type (e.g., /bΛg/) spoken by a set of 6 different talkers (3 males and 3 females). Then, during the test phase, the infants were switched to an entirely new set of talkers uttering the same syllable. The particular talkers included in the habituation and postshift sets was varied randomly from infant to infant but always included 3 males and 3 females. Half of the infants heard utterances of /bΛg/ and the remainder, /dΛg/.

Apparatus—A blind nipple was connected to a Grass PT5 volumetric pressure transducer, which in turn was coupled to a Grass (Model 7) polygraph. A Schmitt trigger provided a digital output of the criterial high-amplitude sucking responses. This output was relayed to a PDP 11/73 computer which recorded and saved the number of criterion responses on a minute-by-minute basis. In addition, the computer accessed the digitized syllables and controlled the presentation of the auditory stimuli at a level of 72 ± 2 dB (C) SPL in response to criterion-level sucking. The sounds were played out over a Kenwood (KA-3500) amplifier to a JBL (4310) loudspeaker. The computer was programmed to record the level of baseline responding, detect the attainment of the criterion for habituation, select the appropriate set of postshift stimuli, and terminate the experiment in the event that the criterion for habituation was achieved after 4 min during the postshift period.

Subjects—The subjects were 72 full-term infants (36 males and 36 females) from the Eugene, Oregon area with a mean age of 9.2 weeks (range 6.1–11.2 weeks). To obtain the 72 infants for this study, it was necessary to test 136. Subjects were excluded for the following reasons: crying (45%), falling asleep prior to shift (11%), repeatedly rejecting the pacifier (15.5%), ceasing to suck during the course of the experiment (e.g., two consecutive minutes of zero-level responding) (11%), failure to achieve the habituation criterion within 24 min (12.5%) and miscellaneous (e.g., equipment failure, bowel movement, etc.) (5%).

Results

For purposes of statistical comparison, subjects' sucking rates were examined for four intervals: baseline minute, third minute before shift, average of minutes 1 and 2 before shift,

and average of the first two minutes after shift. These data were then used to calculate difference scores for each of the following rate comparisons: (a) acquisition of the sucking response: third minute before shift - baseline; (b) habituation: third minute before shift - average of the last two minutes before shift; (c) release from habituation: average of the first two minutes after shift - average of the last two minutes before shift.²

As is typically the case in studies employing the HAS procedure, subjects in all groups acquired the conditioned high-amplitude sucking response and attained the habituation criterion. To assess possible group differences during the preshift period, the sucking rates of each infant for the baseline minute and each of the last three minutes before shift were entered into an analysis of variance. The ANOVA revealed only the expected significant effect of minutes, $F(3, 264) = 113.95, p = .000$. There was no evidence of any significant main effect for groups, $F(5, 264) < 1.00$ or interaction of this variable with minutes, $F(15, 264) < 1.00$.

The data for release from habituation during the postshift period are displayed in Figure 1. Randomization tests for independent samples (Siegel, 1956) were used to assess postshift sucking performance. The release from habituation scores of each experimental group were compared to its appropriate control group (i.e., the single-talker groups with the single-talker control and the multiple-talker groups with the multiple-talker control). The results indicated that the infants showed significant ($p < .001$ or better) increases in sucking to the phonetic change in both the single- and multiple-talker conditions, $t(22) = 4.09$ and 2.62 , respectively. Thus, both groups perceived the difference between /bʌg/ and /dʌg/.

The results from the talker change conditions presented a different pattern. Infants in the single-talker talker change condition readily detected the change in the talker's voice during the postshift period, $t(22) = 8.28 (p < .001)$. This is an indication that even within the same gender the differences in talkers' voices were highly discriminable for the infants. Nevertheless, there are apparently some limits on the ability of infants this age to encode information about talker identity, because infants in the multiple-talker talker change condition did not display evidence of discriminating the difference, $t(22) = 0.27$. This finding replicates one reported for 7-month-old infants by Miller et al. (1982) in which they tried to train infants using a conditioned head-turning procedure to respond to mixed groups of male and female talkers.³

Thus far, the results indicate that, like their older counterparts (Kuhl, 1979, 1983), 2-month-olds are able to adjust to talker variability in detecting a phonetic contrast. However, to evaluate the consequences that this adjustment to variability might have on infants' processing of speech, several additional tests were conducted. First, we compared the postshift levels of responding by the infants in the single- and multiple-talker phonetic change conditions using randomization tests for independent samples. The two groups did not differ significantly with respect to this measure, $t(22) = 0.29$. Next, we sought to determine whether a difference between infants in the single and multiple-talker conditions might show up in the time to habituation measure used by Kuhl and Miller' (1982). We collapsed across the three groups in both the single- and multiple-talker conditions since the treatment in each of the groups was similar for the preshift period.⁴ Subjects in the combined multiple-talker conditions took significantly longer to attain the habituation

²In addition, we also calculated a measure of the release from satiation for the full four minutes after shift (i.e., average of all four minutes after shift minus average of last two minutes before shift). However, since the pattern of results with this measure was identical to that observed with the two minutes measure for all the experiments in the paper, we report only the two-minute measure since it is recognized as the more sensitive of the two.

³For the record, informal testing with adults in our laboratory indicated that they could detect the shift from one group of talkers to the other, although most reported that they did so on the basis of the presence or absence of some distinctive voices.

criterion than those in the combined single-talker conditions (9.61 and 7.58 min, respectively, $t(70) = 3.05$, $p > .005$). Thus, the greater stimulus variability apparently sustains the interest of infants longer during the preshift phase of the experiment.

A third analysis examined whether infants in the multiple-talker conditions might take longer to rehabilitate to the stimuli during the postshift phase of the experiment. We used the measure employed by Bertoni et al. (1988) and calculated, for each experimental group, the amount of time it took in the postshift period before the habituation criterion was achieved and/or the experiment was terminated.⁵ For the phonetic change conditions, no significant differences were observed between the single- and multiple-talker groups, $t(22) = 0.22$. Thus, any changes in processing load attributable to talker variability are not evident in the postshift performance of the infants.

Discussion

What conclusions can be drawn from these results concerning the effects of talker variability on the perception of speech by 2-month-olds? First, it is clear that by this age infants already display some rudimentary form of perceptual normalization. Infants can detect a subtle phonetic change between two stop consonants when as many as 12 talkers' voices vary irrelevantly. This result confirms and extends the findings reported for vowel and fricative contrasts reported by Kuhl (1979, 1983; Holmberg, Morgan, & Kuhl, 1977) with 6-month-olds. Second, 2-month-olds are capable of perceiving differences between two different talkers' utterances of the same syllable, although there do seem to be limits to what they can encode about talker differences. Thus, when listening to a set of talkers uttering a particular syllable, they do not detect a change to a new set of talkers uttering the same syllable. Third, discrimination performance does not suffer significantly when tokens from multiple talkers are used as opposed to when only tokens from a single talker are used. Hence, like Kuhl and Miller's (1982) finding that irrelevant pitch variation did not significantly interfere with infants' ability to detect a pitch change, the present study found no evidence that talker variability interferes with infants' detection of a phonetic contrast. However, this is not to say that the infants' processing of speech is unaffected by talker variability. In fact, infants exposed to tokens from a variety of different talkers took significantly longer to habituate to syllables than did those who were exposed to tokens from a single talker. Thus, the infants clearly display some sensitivity to stimulus variability.⁶

Perhaps the most striking finding is that infants listening to syllables produced by many different talkers did so well in detecting the subtle phonetic change between [b] and [d]. Can it then be assumed that, in contrast to adults, lower-level perceptual processes in infants are unaffected by talker variability? Such an assumption would be premature for a variety of reasons. First, note that talker variability did have an effect on the time it took infants to habituate to the sounds. Hence, stimulus variability may have affected the development of

⁴As one of the reviewers pointed out to us, the conditions in multiple-talker conditions are not identical because there are only 6 tokens in the talker change condition, but 12 in the other two. However, because the comparison of interest was how the existence of multiple tokens affects habituation, we felt that it was legitimate to collapse across the groups in the way we report here. For the record, the pattern of significant results does not change, even if the data from the talker change condition is omitted from the multiple-talker conditions.

⁵Only the data from the two phonetic change groups were used in this calculation because responding in the two control groups is already near floor level and no new stimulation is introduced during the postshift period for these groups. Similarly, the multiple-talker talker change group did not show significant evidence of dishabituating during the postshift period, so a comparison of its performance with the single-talker talker change group was not meaningful.

⁶One of the reviewers questioned whether we could attribute the slower habituation to more difficulty in learning the tokens or because the variation in the tokens was simply more interesting to the infants. We prefer the former explanation because in our previous studies using multiple stimulus sets (e.g., Bertoni et al., 1988; Jusczyk et al., 1990; Jusczyk & Derrah, 1987) we did not find evidence of significantly longer habituation times as compared to studies in which a single stimulus was used. However, we agree that it is not possible to resolve this issue in the present study.

the perceptual representation of these speech sounds. Second, the lower-level perceptual effects caused by talker variability that have been reported for adults (Mullennix et al., 1989) are most evident when the stimulus conditions are less than optimal. For example, the largest effects in the Mullennix et al. experiments occurred when the stimuli were degraded with noise. It is possible that under similar circumstances infants might also show deficits in discrimination performance in the presence of talker variability. Moreover, as noted earlier, encoding processes in memory also are affected in adults when talker variability is present (Martin et al., 1989). Given our finding of longer habituation times for the multiple-talker conditions, we wondered whether infants might also be affected by talker variability when they are required to encode and subsequently remember speech signals. To study this issue, we conducted the following experiment.

EXPERIMENT 2

The subject of how speech is encoded in memory and remembered by infants has been discussed at times in conjunction with previous studies. For example, in their study of vowel perception by infants, Swoboda, Morse, and Leavitt (1976) noted that the likelihood that infants discriminated certain vowel contrasts appeared to be inversely related to the length of the interval between the last occurrence of the preshift stimulus and the first occurrence of the postshift stimulus when the HAS procedure was employed. Morse (1978) later suggested that manipulations of the preshift-postshift interval duration could provide a way of assessing memory effects in the HAS paradigm. In addition, discussions of the role that memory may play in discrimination performance have been raised in conjunction with studies that have used varied stimulus sets in place of a single stimulus during the habituation phase of the HAS procedure (e.g., Bertoncini et al., 1988; Jusczyk & Derrah, 1987; Kuhl & Miller, 1982; Miller & Eimas, 1979). Nevertheless, until very recently direct attempts to manipulate memory factors in the HAS procedure have not been reported.

Clearly, information about the encoding processes that infants use for speech is critical to understanding the growth and development of the mental lexicon of the native language. The infant must ultimately store away some sort of acoustic-phonetic representation that will allow him or her to access the meanings of spoken words (see Jusczyk, 1985a, 1986 see Jusczyk, in press-a, in press-b, for further discussion of this point). In the present context, one can ask about the way in which talker variability might influence the encoding of speech sounds by infants. On the one hand, talker variability might be expected to interfere with encoding processes, as Martin et al. (1989) observed for adults in several recall tasks. On the other hand, the classic study of Posner and Keele (1968) on prototypes indicates that category formation is enhanced by exposure to a diverse set of instances as opposed to repeated exposure to a single instance. Given this finding, one might argue that tokens from multiple talkers would permit infants to form a robust prototype that would actually facilitate later recognition of a syllable or word type. Interestingly enough, Grieser and Kuhl (1989; see also Kuhl, 1991) have recently reported evidence consistent with the view that 6-month-olds may form prototypes for some speech sound categories and that “this may contribute to their seemingly efficient processing of speech information ...” (p. 577). Indeed, one way of interpreting the lack of discrimination by infants in the multiple-talker talker change condition in the previous experiment is that the infants formed a prototype for the syllable category that they were exposed to during the habituation phase and that the new instances that they heard during the postshift phase were simply treated as members of a familiar category.

The first step toward understanding the consequences of talker variability on encoding by infants is to devise a means of tapping their representation and memory of speech sounds. In addition to modifying the traditional HAS procedure by presenting a randomized set of

sounds as in the previous experiment (see also Bertoncini et al., 1988; Jusczyk & Derrah, 1987; Kuhl & Miller, 1982), we also introduced another modification first employed by Jusczyk, Kennedy, and Jusczyk (in preparation). Specifically, a 2-min delay period filled with a slide presentation is introduced between the habituation and postshift phases of the HAS procedure. No auditory stimulation is present during this period. When the slide presentation is completed, the postshift period begins and the auditory stimulation resumes with novel or familiar stimuli depending on whether an experimental or control condition is involved.

The basic issue here is to determine whether talker variability affects infants' encoding of speech in long-term memory. Consequently, we decided to compare performance by 2-month-olds under both single- and multiple-talker conditions. Four groups of infants were tested in conditions that paralleled the phonetic change and control conditions of Experiment 1 (single-talker phonetic change and control conditions, and multiple-talker phonetic change and control conditions). If talker variability affects encoding, then discrimination performance in the multiple-talker condition should be worse than for the single-talker condition. If, instead, talker variability promotes the formation of prototypes, then performance may actually be better in the multiple-talker condition. Finally, in addition to these four groups, a fifth group, single-talker talker change condition, was included in order to see whether infants might encode information about talker identity into their representations of syllables. To the extent that information about talker identity is stored in memory, one would expect that infants should respond to the talker change after the delay interval.⁷

Method

Procedure—A modified version of the high-amplitude sucking procedure described in the previous experiment was used. The modification consisted of the insertion of a 2-min delay interval between the habituation and postshift phases of the experiment. Upon the attainment of the habituation criterion, the computer beeped, signaling to the experimenter in the control room to initiate the slide show. The fixation slide was extinguished and in its place new slides were projected. The slides consisted of a series of 24 colorful family vacation slides that were projected on the wall facing the infant in the test room. Each slide was shown for 5 s. During the slide presentation, the experimenter in the test room continued to hold the pacifier in the infant's mouth although no auditory stimulation was presented. Following the 24th slide, the fixation slide was projected once again and auditory stimulation was available in response to criterion sucking. In all other respects, the procedure was identical to that used in Experiment 1. Extensive pilot testing conducted by Jusczyk et al. (in preparation) was used to determine the parameters for the memory delay interval. For example, the decision to keep the pacifier in place during the delay was made when it was determined that the removal and reinsertion of the pacifier during the delay interval led to spurious increases in sucking in the control and experimental groups. Similarly, the number of slides employed and their projection durations were optimal for maintaining the infants' attention.

Apparatus—The apparatus used was identical to that described for the previous experiment.

Stimuli—The same stimulus materials were used as in the previous experiment.

⁷The parallel talker change condition for multiple-talkers was not tested because of the failure of the infants in Experiment 1 to discriminate the difference even when no delay interval was employed

Design—Each infant was seen for one experimental session. Twelve subjects were assigned randomly to each of five test groups. Two of these groups employed tokens of /bʌg/ and /dʌg/ from all 12 talkers. For the multiple-talker phonetic change condition, randomly ordered tokens of one syllable type (/bʌg/) for half the infants, /dʌg/ for the other half) were presented during the habituation phase, and tokens of the other syllable type were played during the postshift phase. For the multiple-talker control condition, one of the two syllable types spoken by all 12 talkers was presented for both phases of the experiment. Two other groups heard tokens produced by a single talker for the entire test session (although the identity of the talker varied for each infant). For the single-talker phonetic change condition, one syllable (/bʌg/ for half the infants, /dʌg/ for the other half) was played during the habituation phase and the other one during the postshift phase. For the single-talker control condition, one of these two syllables was presented for both phases. Finally, the single-talker talker change condition heard tokens of the same syllable spoken by two different talkers of the same gender. During the habituation phase, the token from one talker was played, and during the postshift phase the token from the other talker was played. Once again, each infant heard a different pair of talkers. Half heard a female pair and half heard a male pair. Similarly, half of the subjects listened to versions of /bʌg/ and the other half to versions of /dʌg/.

Subjects—The subjects were 60 infants (32 males and 28 females) from the Eugene, Oregon area with a mean age of 7.4 weeks (range 5.6–10.6 weeks). To obtain the 60 infants for this study, it was necessary to test 121. Subjects were excluded for the following reasons: crying (51%), falling asleep prior to shift (16%), repeatedly rejecting the pacifier (18%), failure to achieve the habituation criterion within 24 min (11.5%), miscellaneous (experimenter error, parental interference) (3.5%).

Results

The data were analyzed as in the previous experiment. Difference scores were calculated for each subject to assess (a) acquisition of the sucking response, (b) habituation to the preshift stimuli, and (c) release from habituation during the first 2 min of the postshift period. As in the previous experiment, all groups acquired the conditioned response and habituated to the preshift stimuli. Moreover, an ANOVA used to assess possible group differences during the baseline minute and each of the last 3 min prior to the shift revealed only the expected significant effect of minutes, $F(3, 220) = 194.93, p = .000$. Neither the main effect for groups, $F(4, 220) = 2.036, p = .09$, nor the interaction of this variable with minutes, $F(12, 220) = 0.648, p = .80$, was statistically significant.

The data on release from habituation are shown in Figure 2. Randomization tests for independent samples were again used to assess postshift sucking performance. In contrast to the previous experiment, a difference emerged in the way in which infants in the single- and multiple-talker conditions responded to the phonetic change after the delay period. In particular, only in the single-talker condition did the phonetic change group show a significant increase in sucking relative to the control group during the postshift period, $t(22) = 2.11, p = .046$. Not only was the difference between the phonetic change and control groups not significant for multiple-talker conditions, $t(22) = -0.29$, but it was even in the wrong direction. Thus, the presence of talker variability apparently does hamper infants' encoding of speech sounds in memory.

Performance in the single-talker talker change group also proved to be different from the results observed in Experiment 1. When compared to the single-talker control group, infants in the talker change group did not exhibit a significant increase in postshift sucking, $t(22) =$

0.89, $p = .38$. This suggests that talker identity may not be salient with respect to the kind of information that infants encode and/or retrieve about speech sounds.

As in the previous experiment, we also examined the effects of talker variability on the time to achieve the habituation criterion in both the preshift and postshift phases of the experiment. For the preshift phase, we collapsed across all the single-talker groups and across both multiple-talker groups since the stimulus presentation was the same for this period. Once again, there was evidence of significantly longer times to habituation, $t(58) = 2.53$, $p < .02$, for the multiple-talker group (11.74 min) than for the single-talker group (9.22 min). To evaluate rehabilitation during the postshift period, the comparable groups were the single-talker phonetic change (5.08 min) and multiple-talker phonetic change (5.58 min) groups. There was no evidence that these groups differed significantly on this measure, $t(22) = 0.76$. Hence, talker variability appears to have affected the time to habituation only during the preshift phase of the experiment.

Discussion

Two-month-old infants are able to retain some information about speech sounds for a delay period of 2 min. This is evident in the ability of infants in the single-talker phonetic change group to detect the difference between the stimuli played during the preshift and postshift periods. Nevertheless, it is also clear that talker variability affects encoding and/or retrieval processes in infants this age. Thus, infants in the multiple-talker phonetic change group did not discriminate the difference between the preshift and postshift stimuli. The locus of this effect appears to be in the encoding processes associated with transfer to long-term memory. In the previous experiment without the delay interval, infants in the multiple-talker group were able to perceive the very same phonetic change. Hence, as is the case for adults (Martin et al., 1989), we find evidence that when infants are exposed to different talkers, stimulus variability can adversely affect encoding processes that are critical for subsequent retrieval of this information.

Evidently, given the kind of experience that infants received in the present experiment, exposure to different talkers uttering the same syllable did not encourage the formation of a prototype - at least not a prototype that could be used to pick up the phonetic contrast between the syllables. Rather, coping with talker variability appeared to interfere with the way in which infants encoded the stimulus information. This is shown not only in the failure of the infants in the multiple-talker phonetic change group to discriminate the contrast, but also by the fact that they took longer to habituate to the syllables in the first place. One possible explanation of the difficulty that these infants had is that they were trying to encode the syllables individually according to talker identity. However, this explanation seems unlikely in view of the fact that infants in the single-talker talker change group gave no evidence of retaining information about talker identity over the delay period, despite the fact that their counterparts in Experiment 1 did detect such a change in the absence of any delay interval.

An alternative explanation of the present results is that the pattern observed here is not the result of talker variability per se but simply the presence of multiple tokens in the familiarization phase combined with the delay in testing. Were this explanation correct, then one would expect that whenever multiple tokens are used during the preshift phase and testing is delayed, then infants should fail to detect the presence of new items in the test phase. However, Jusczyk et al. (in preparation) used a series of phonetically distinct syllables (e.g., [bi], [ba], [bu]) in the preshift phase of their experiment and found that 2-month-olds *did* detect phonetic changes after a 2-min delay in testing. Therefore, the decrements in performance observed in the present experiment probably had more to do

with the kind of information that was varying (talkers' voices) than the mere fact that something was varying.

To gain a better understanding of the effects that talker variability has on speech processing by infants, we conducted a series of further experiments in which we restricted the kind of stimulus variability that was present. We also sought to determine whether talker variability disrupts infants' memory for all types of contrasts between speech sounds. For instance, one possible consequence of hearing a number of different talkers uttering the same syllable is that infants might begin to focus their attention on voice quality differences. If this is so, then information relating to voice quality may be more distinctive in memory than phonetic details in any representations that infants have about the speech sounds they heard previously during the familiarization phase.

EXPERIMENT 3

Recent work by Jusczyk et al. (1990) on the role of attention in infant speech perception suggests that the composition of the set of instances to which infants are exposed in the familiarization phase of the HAS procedure can make them more or less likely to pick up certain phonetic contrasts. Specifically, the inclusion of syllables containing consonants that were perceptually very similar enabled infants to detect the addition of a new item with a different consonant during the test phase. Moreover, if the familiarization set was composed of items that were perceptually very dissimilar, the infants failed to pick up contrasts that they had previously discriminated. These results are very similar to the findings that Nosofsky (e.g., 1986 e.g., 1987 e.g., 1988) has observed about the role of attention in categorization processes. Jusczyk et al. have argued that the composition of the familiarization set helps direct the attentional focus of infants to certain perceptual dimensions, rendering them more salient, while reducing sensitivity to unattended dimensions. In the present situation, a comparable case could be made that changes in talkers' voices lead infants to focus on aspects of speech relating to voice quality rather than to phonetic information. Thus, changes relating to voice quality may be detectable even with delayed testing. In principle, there are many such changes that infants might detect (e.g., spoken vs. whispered speech; familiar vs. unfamiliar voices; male vs. female voices; speaking rate, dialect differences, etc.). We chose to examine male versus female voices because it also moved us closer to a second goal - namely, investigating whether the detection of the phonetic contrast after the delay occurs when the overall variability among talkers is reduced (by using either just the males or just the females).

Previous work by Miller et al. (1982) showed that 7-month-old infants could perform a categorization task requiring them to respond differentially to male versus female voices. Moreover, they showed that infants' success on the task was not simply a result of using the fundamental frequency of the voices (males have generally lower fundamental frequencies than females) to distinguish the categories. Thus, over and above any differences in fundamental frequencies, there are some aspects of voice quality differentiating male and female voices to which infants are able to respond.

Accordingly, we examined whether infants exposed to syllables produced by different talkers from one gender would retain speech information (either about the phonetic details or voice quality) over a 2-min delay interval. Because the main objective of the study was to determine whether talker variability disrupts the memory for any sort of speech contrast, we elected to test infants only on multiple-talker stimuli. Hence, the present experiment had only three test groups. One of these groups was assigned to the multiple-talker phonetic change condition. This group was similar to the ones used in the two previous experiments, with the exception that only tokens from either the male or the female talkers were used

throughout the experiment. A second group, the talker gender change condition, was exposed to utterances of a particular syllable by 6 talkers of one gender during the preshift period, and to utterances of the same syllable by 6 talkers of the opposite gender in the postshift period after a 2-min delay. The purpose of this group was to provide a contrast based on voice quality differences. If the exposure to different voices causes infants to focus on voice quality differences, then infants may be more attentive to changes along these dimensions. The remaining group, gender control condition, heard utterances of a particular syllable by 6 talkers of the same gender throughout the entire test session.

Method

Procedure and apparatus—The procedure and apparatus were identical to that of Experiment 2.

Stimuli—The same stimulus materials were used as in the previous two experiments.

Design—Each infant was seen for one experimental session. Twelve subjects were assigned to each of three test groups. Infants in the multiple-talker phonetic change and the talker gender change groups both heard randomly ordered tokens of one syllable type (half heard /bʌg/, half heard /dʌg/) produced by either 6 male or 6 female talkers during the preshift phase of the experiment. These two groups differed during the postshift period which began after a 2-min delay interval. Infants in the multiple-talker phonetic change group heard utterances of the other syllable produced by *the same 6 talkers*. Infants in the talker gender change group heard utterances of the same syllable type produced by *6 talkers of the opposite gender*. Infants in the no-shift control group were treated in the same way as the other groups for the preshift period, but during the postshift period they continued to hear the same utterances that they had heard prior to the 2-min delay. For the infants in this group, half of them heard utterances from females and half heard utterances from males. Similarly, half heard /bʌg/ and half heard /dʌg/.

Subjects—The subjects were 36 infants (15 males and 21 females) from the Eugene, Oregon area with a mean age of 8.3 weeks (range 6.0–12.0 weeks). In order to obtain the 36 infants for the study, it was necessary to test 69. Subjects were excluded for the following reasons: crying (45.4%), falling asleep prior to shift (15.2%), repeatedly rejecting the pacifier (27.3%), and failure to attain the habituation criterion within 24 min (12.1%).

Results

The data were analyzed as in the previous two experiments. Difference scores were calculated for each subject to assess (a) acquisition of the sucking response, (b) habituation to the preshift stimuli, and (c) release from habituation during the first 2 min of the postshift period. As in the previous experiments, all the groups acquired the conditioned response and habituated to the preshift stimulus. Moreover, an ANOVA used to assess possible group differences during the baseline minute and each of the last three preshift minutes revealed only the anticipated significant effect of minutes $F(3, 132) = 81.823, p = .000$. Neither the main effect for groups, $F(2, 132) = 2.17, p = .118$, nor the interaction of this variable with minutes, $F(6, 132) = 1.017, p = .417$, was statistically significant.

The data on release from habituation are shown in Figure 3. Randomization tests for independent samples, used to assess postshift sucking performance, indicated a significant difference between the talker gender change group and the no-shift control group, $t(22) = 3.923, p = .001$. In the presence of talker variability, 2-month-olds were able to retain information about the gender of the talkers over a 2-min delay interval. However, the pattern was different for infants in the multiple-talker phonetic change group. There was no

indication that subjects in this group differed significantly from their counterparts in the no-shift control group during the test phase, $t(22) = 0.291$, $p = .774$. Hence, even though the variability among talkers was limited to a single gender, detection of the phonetic contrast over the delay interval was impaired because of the stimuli subjects were exposed to during habituation.

Discussion

Evidently coping with talker variability does not entirely disrupt the retention of all information about speech by 2-month-olds. In the present case, infants did not appear to experience a difficulty in detecting a change related to voice quality (i.e., from talkers of one gender to talkers of the opposite gender). In this respect, the results also replicate the basic findings of Miller et al. (1982) and extend them to a younger age group. Hence, 2-month-olds are able to distinguish between male and female voices. Moreover, they can, in some circumstances, retain information about the gender of the talker for at least a 2-min interval.

We argued earlier that familiarizing the infants with a set of utterances from different talkers may encourage them to selectively attend more to voice quality differences than phonetic details. The pattern of results obtained here are consistent with this view.⁸ Thus, even though talker variability was limited to a single gender, infants in the phonetic change group were no more successful in discriminating the syllable types than were subjects in the previous experiment where tokens from both genders were included.

The present experiment demonstrates that the presence of talker variability need not disrupt the encoding of all information about speech by 2-month-olds. Some information related to the dimensions that are being varied does appear to be encoded and retained by infants. However, the encoding of phonetic details does appear to be disrupted by the variability that exists across different talkers' voices, even from the same gender. Does this problem arise because of difficulties that infants have in normalizing for speech produced by different talkers? Or, would similar effects occur if infants were listening to multiple tokens produced by the same talker? In other words, is it talker variability, or simply variability among stimulus tokens that causes the difficulties for the infants? The next two experiments were undertaken to address these issues. The first investigates the consequences of token variability under immediate test conditions; the second does the same for delayed test conditions.

EXPERIMENT 4

To gain a clearer understanding of the way that stimulus variability affects the encoding and retrieval of speech sounds by infants, we decided to limit the variability among the tokens to ones produced by a single talker who uttered the tokens in the same sentential context on a number of different occasions. Would exposure to the different tokens spoken by the same talker lead infants to focus on subtle voice quality differences and neglect perceptual dimensions relevant to distinguishing among phonetic contrasts? Based on previous work in the field such an outcome seems unlikely for situations involving immediate testing. For instance, Trehub (1976) used multiple tokens of particular syllables in her study which showed that infants are sensitive to certain non-native language contrasts. There was no

⁸There is a question that arises in conjunction with our hypothesis that focusing on voice quality differences during the familiarization period may have made the subjects more apt to encode information of this kind into memory rather than phonetic details. If familiarizing infants with different talkers causes them to focus on voice quality differences, why did infants in the multiple- talker talker change group of Experiment 1 fail to pick up a change from one set of 6 talkers to an entirely different set? The most plausible answer is that the grouping according to gender in the present experiment may be a perceptually salient dimension for infants, whereas no such salient dimension was provided by the essentially random grouping of talkers in Experiment 1. Clearly, this issue bears further investigation.

indication that having multiple tokens of the syllables affected discrimination performance. Still, in order to interpret any findings concerning the delayed testing paradigm with multiple tokens from a single talker, it is necessary to know how infants respond to these stimuli under immediate testing.

Our basic concern in this experiment was to determine whether stimulus variability among tokens produced by the same talker would affect the discrimination of the contrast between [b^g] and [d^g] by infants. For this reason, two groups of infants were tested in conditions that involved the presentation of multiple tokens. The multiple-token phonetic change group was habituated to 12 different tokens of one of the two syllables during the familiarization phase (e.g., [b^g]) and was presented with 12 different tokens of the other syllable during the test phase ([d^g]). The multiple-token no-shift control group was treated in the same manner during the familiarization phase but continued to hear the same items during the test phase. In addition to the multiple-token groups, two single-token groups were tested in order to determine whether infants discriminated among the different tokens of the same syllable. The single-token change group heard one token (e.g., [b^g1]) during the familiarization phase and a different token of the same syllable type (e.g., [b^g2]) during the test phase. The performance of this group was compared to a single-token no-shift control group.⁹

Method

Procedure and apparatus—The procedure and apparatus were identical to that of Experiment 1.

Stimuli—Tokens from two of the original talkers (one male and one female) in the previous experiments were used. The utterances from each talker consisted of 12 utterances each of the English words “bug” and “dug.” The talkers were not asked to vary their production of the syllables in any way. The syllables were recorded, digitized, and stored as described in Experiment 1.

Design—Each infant was seen for one experimental session. Twelve subjects were assigned to each of four test groups. Infants in the multiple-token phonetic change and the multiple-token control groups were presented with 12 different tokens of one of the two syllables (e.g., [b^g]) during the familiarization phase (half heard tokens from the male talker; the other half heard tokens from the female talker). During the test phase, the phonetic change group heard the 12 tokens of the other syllable (e.g., [d^g]) produced by the same talker, whereas the control group continued to hear the 12 tokens of the original syllable. The remaining two groups heard only one token in each phase of the experiment. Half of the subjects in each group heard the male talker’s tokens, and the other half heard the female talker’s tokens. In addition, half of the subjects heard tokens of [b^g], and the others [d^g]. Infants in the single-token change group heard a token of one syllable (e.g., [d^g1]) during the familiarization phase and another token of the same syllable (e.g., [d^g2]) during the test phase. Infants in the single-token no-shift control group heard the same token throughout both phases of the experiment.

Subjects—The subjects were 48 infants (27 males and 21 females) from the Eugene, Oregon area with a mean age of 8.3 weeks (range 6.0–10.4 weeks). In order to obtain the 48 infants for the study, it was necessary to test 98. Subjects were excluded for the following reasons: crying (58%), falling asleep prior to shift (12%), repeatedly rejecting the pacifier (14%), and failure to attain the habituation criterion within 24 min (16%).

⁹We decided not to test a single-token phonetic change group in the present experiment since it is essentially equivalent to the single-talker phonetic change group of Experiment 1 (see also Experiment 5).

Results and discussion

The data were analyzed as in the previous experiments. Once again, all groups acquired the conditioned response and habituated to the preshift stimulus. An ANOVA used to assess possible group differences during the baseline minute and each of the last three preshift minutes revealed only the anticipated significant effect of minutes, $F(3,176) = 101.693$, $p = .000$. Neither the main effect for groups, $F(3, 176) = 2.018$, $p = .113$, nor the interaction of this variable with minutes, $F(9, 176) = 0.678$, $p = .728$, was statistically significant.

The data on release from habituation are shown in Figure 4. Randomization tests for independent samples, used to assess postshift sucking performance, indicated a significant difference in the multiple-token conditions between the phonetic change and no-shift control groups, $t(22) = 5.589$, $p = .000$. This indicates that the infants were able to perceive the contrast between [b^g] and [d^g] even though multiple instances of each syllable were used. By comparison, although in the right direction, the difference between the two single-token conditions, the token change and no-shift control groups is not statistically significant, $t(22) = 1.634$, $p = .116$. Thus we cannot conclude that the infants detected any difference between two tokens of the same syllable.

Because both single- and multiple-token conditions were tested in the present experiment, we were able to conduct the analysis on time to habituation that we had carried out in the first two experiments. Once again, we collapsed across the two groups in each condition since they received the same treatment during this period. Subjects in the multiple-token conditions took an average of 8.67 min to habituate, whereas those in the single-token conditions took 8.00 min. This difference was not statistically significant, $t(22) = 1.337$, $p = .175$.

The present results demonstrate that, under immediate test conditions, 2- month-olds can compensate for the variability that exists among the tokens of a particular talker in perceiving a contrast between two syllables. This finding is not surprising, given that they were able to cope with a greater range of variability among tokens of different talkers in detecting the same phonetic contrast in Experiment 1. That the task in the earlier experiment may have been more formidable is suggested by two other findings from the present study. First, unlike Experiment 1 where there was evidence that infants discriminated differences between tokens of the same syllable produced by different talkers, this did not appear to be true for differences between two tokens of the same syllable produced by the same talker. Second, infants in the multiple-talker conditions of the first two experiments took significantly longer than their counterparts in the single-talker conditions. Although in the same direction, the habituation times for the multiple-token conditions were not significantly longer than those of the single-token conditions. Hence, it seems reasonable to assume that, in settings like the present one, coping with the variability from a single talker taxes the perceptual processing capacities of the infant less than does coping with variability from multiple talkers. Can we further assume that token variability will not disrupt encoding processes under delayed testing conditions? This question is addressed in the next experiment.

EXPERIMENT 5

The previous experiment demonstrates that infants are able to compensate for within-talker variability, as well as for between-talker variability, in discriminating a phonetic contrast. In fact, unlike the situation with variability among talkers, there were no obvious signs that the variability among tokens affected infants' performance. Still, it is possible that infants were focusing their attention on voice quality differences in a way that was not measured by the procedure used in the previous experiment. Thus, the deficits in discrimination performance

with talker variability were only evident when delayed testing was used. An indication that within-talker variability also disrupts discrimination performance under delayed testing would be evidence that it is variability among tokens in general, and not just variability among tokens of different talkers, that disrupts encoding of phonetic information. Alternatively, a finding that discrimination performance under delayed testing does not suffer would suggest either that the *amount* of variability is responsible for the disruption or else that the perceptual mechanisms responsible for normalization across different talkers are somehow interfering with infants' capacity for encoding speech into memory.

We decided to include both single- and multiple-token conditions in this experiment. A single-token phonetic change condition was tested against a single-token no-shift control to determine whether our talkers' tokens of [b^g] and [d^g] were discriminable for infants over a 2-min delay interval. The remaining two groups of infants were exposed to multiple tokens of these syllables. One group, multiple-token phonetic change, heard 12 tokens of one of the syllables during the familiarization phase, and 12 tokens of the other syllable during the test phase. The remaining group, multiple-token no-shift control, heard the 12 tokens of one of the syllables throughout the experiment.

Method

Procedure and apparatus—The procedure and apparatus were identical to that of Experiment 2.

Stimuli—The stimuli used were identical to those in Experiment 4.

Design—Each infant was seen for one experimental session. Twelve subjects were assigned to each of four test groups. Infants in the multiple-token phonetic change and the multiple-token control groups were presented with 12 different tokens of one of the two syllables (e.g., [b^g]) during the familiarization phase (half heard tokens from the male talker; the other half heard tokens from the female talker). During the test phase, the phonetic change group heard the 12 tokens of the other syllable (e.g., [d^g]) produced by the same talker, whereas the control group continued to hear the 12 tokens of the original syllable. The two groups in the single-token conditions heard only one token in each phase of the experiment. Half of the subjects in each group heard the male talker's tokens, and the other half heard the female talker's tokens. Infants in the single-token phonetic change group heard a token of one syllable (e.g., [d^g1]) during the familiarization phase and a token of the other syllable (e.g., [b^g2]) during the test phase. Infants in the single-token no-shift control group heard the same token throughout both phases of the experiment.

Subjects—The subjects were 48 infants (24 males and 24 females) from the Eugene, Oregon area with a mean age of 8.5 weeks (range 6.1–10.5 weeks). In order to obtain the 48 infants for the study, it was necessary to test 108. Subjects were excluded for the following reasons: crying (65%), falling asleep prior to shift (10%), repeatedly rejecting the pacifier (14%), failure to attain the habituation criterion within 24 min (14%), and experimenter error (2%).

Results and discussion

The data were analyzed as in the previous experiments. Once again, all groups acquired the conditioned response and habituated to the preshift stimulus. An ANOVA used to assess possible group differences during the baseline minute and each of the last three preshift minutes revealed only the anticipated significant effect of minutes, $F(3,176) = 78.40$, $p = .000$. Neither the main effect for groups, $F(3, 176) = 2.24$, $p = .09$, nor the interaction of this variable with minutes, $F(9,176) = 0.59$, $p = .81$, was statistically significant.

The data on release from habituation are shown in Figure 5. Randomization tests for independent samples were used to assess postshift sucking performance. In the single-token conditions, the phonetic change group was found to differ significantly from its no-shift control group, $t(22) = 4.76$, $p = .000$. However, no significant difference was found in the postshift sucking behavior of the multiple-token phonetic change group and its no-shift control, $t(22) = 1.13$, $p = .269$. Thus, the variability among the tokens seems to have once again affected the encoding of the syllables into memory.

As in the previous experiments, we also examined whether the time to habituation measure indicated any differences between the single- and multiple-token conditions. Infants in the single-token conditions took an average of 8.13 min to attain the habituation criterion, while those in the multiple-token conditions took 8.58 min. This difference was not significant, $t(46) = 0.789$, $p = .434$. Thus, any differences in the way that infants processed the speech sounds during the familiarization phase of the experiment were not evident with this measure.

There are several similarities and differences in the data between the present results using different tokens from the same talker and Experiments 1 and 2 using different tokens from different talkers. In both instances, variability did not disrupt discrimination of the phonetic contrast under immediate testing conditions. Only when delayed testing was used did variability among tokens result in a failure to discriminate the phonetic contrast. By comparison, infants' ability to detect the phonetic contrast remained good for the delayed testing period when only a single token of each syllable was used. These results suggest that it is variability among tokens in general, and not just variability among tokens from different talkers, that affects the infants' encoding of phonetic information into memory. In both types of situations, the variability that exists among tokens in the familiarization set may draw infants to focus their attention on voice quality as opposed to phonetic details. Finally, in the experiment involving multiple tokens from a single talker, time to habituation was not significantly longer than in the single-token condition, although it was in the earlier experiments with tokens from multiple talkers. We have no explanation for this difference. We can only speculate that the greater variability that exists among the tokens of the multiple talkers somehow sustains infants' interest longer during the familiarization phase of the test procedure.

One remaining issue deserves comment. In the present experiment, as in Experiment 2, we found that infants in the single-token phonetic change group were able to discriminate the difference between the syllables even with a delay interval, whereas the multiple-token phonetic change group could not. We attributed this difference to the variability among the tokens affecting the encoding of phonetic information in the latter case, but not the former. Is there some other explanation for these results? Perhaps the difference between these groups lies not in the way that they were affected in the familiarization period but has to do with differences in the nature of the control groups that they are compared with during the test period. In the present experiment, both control groups were no-shift controls (i.e., the same stimuli were used in both phases of the experiment). Yet, examination of Figure 5 suggests that the sucking rate in the single-token no-shift control group dropped considerably more in the test period than did that of the multiple-token no-shift control. One reason for this may have been the fact that there is absolutely no variability among tokens in the single-token group. Thus, one might argue that a better control group for this condition would be one along the lines of the single-token change group of Experiment 4 (i.e., present a token like [dʌg1] during the familiarization phase, and a different token of the same syllable, [dʌg2], during the test phase). To explore this possibility, we tested an additional group of 12 infants on this type of pairing and compared the performance of this group to each of the single-token groups in Experiment 5. Subjects in this new group behaved much

like the single-token no-shift control subjects. Specifically, the new group did not differ significantly from the single-token no-shift control group, $t(22) = 1.481$, $p = .153$, but it did have significantly lower sucking rates than the single-token phonetic change group, $t(22) = 2.56$, $p = .034$. So, even if different tokens had been used in the familiarization and test phases for the control group in the single-token conditions, the overall pattern of results would not have changed in any substantial way. Hence, we believe that the most plausible explanation for the differences in the pattern of results for the single- and multiple-token conditions lies in the way that encoding processes in memory are affected by the presence of stimulus variability among the tokens.

General discussion

The present study demonstrates that infants as young as 2 months of age have the basic capacities to cope with talker variability in speech perception. This finding replicates the earlier results reported by Kuhl (1979, 1983) with 6-month-old infants. In addition, the study shows that infants are also able to cope with the kind of variability that is present in utterances of the same syllable produced by a single talker. However, dealing with stimulus variability, both within a talker and among different talkers, also appears to carry some costs with respect to the way that speech is processed. For instance, when stimulus variability from different talkers is present, infants take longer to habituate to repetitions of a particular syllable. More importantly, the presence of variability can hamper infants' encoding of speech sounds so that they fail to detect a phonetic change after a short delay interval. Consequently, variability among different tokens of a given syllable primarily affects the way that infants remember information in the speech signal.

In certain respects, our results are similar to ones previously reported for adults. Specifically, in the absence of any noise-induced degradation of the speech signal, there is little evidence that perceptual processes related to the identification of items are disrupted significantly in adults (Mullennix et al., 1989), whereas the mere presence of talker variability is sufficient to adversely affect processes associated with the retention of speech information (Martin et al., 1989). These parallels that we have observed in the way in which infants and adults respond to talker variability are certainly at least consistent with Miller's (1987) contention that the basic mechanisms for perceptual normalization may be innately prewired. Of course, there may be other aspects of perceptual normalization that are either incomplete at this age or require further experience with a native language. Nevertheless, infants as young as 2 months of age do show somewhat remarkable perceptual abilities to deal with variability in speech perception.

An important part of understanding how the infant develops a lexicon for words in the native language is to determine the kind of information that they retain about speech sounds. By comparing how infants perform on the same contrasts under conditions of immediate and delayed testing, we were able to gain some rudimentary appreciation of the kind of information that is most likely to be retained from the speech signal. As noted earlier, previous research by Jusczyk et al. (in preparation) demonstrated that infants are able to retain information about the phonetic features of syllables for a short delay period. The performance of the infants in the single-talker phonetic change group of Experiment 2 and the single-token phonetic change group of Experiment 5 replicated this basic finding. Hence, information relevant to the phonetic coding of speech sounds is one type of information that infants can retain over a short delay interval. However, by exploring how variability affects infants' retention of phonetic information, the present study moves in the direction of obtaining a more precise specification of infants' representations of speech sounds. Thus, the present study shows that retention of phonetic information about syllables may be adversely affected in certain situations, such as when there is variability present among

tokens produced by either a single talker or many different talkers (e.g., the multiple-talker phonetic change group in Experiment 2 and the multiple-token phonetic change group in Experiment 5).

Why should variability across different tokens affect the retention of phonetic information by infants? To answer this, let us consider the overall pattern of results obtained in this investigation. First, note that information about phonetic differences is available under immediate testing conditions, regardless of whether stimulus variability is present from either the same or different talkers. Second, infants apparently do retain information relevant to phonetic contrasts under delayed testing conditions when there is no variability present. Third, information relating to other aspects of speech sounds, such as voice quality, is also retained for delay periods, even when variability from different talkers is present. Our explanation of this pattern of results is that familiarizing infants with a series of different tokens of the same syllable causes them to focus their attention on the way the tokens differ - in this case, by aspects of voice quality.¹⁰ Under conditions of immediate testing, information relating to the phonetic features of the familiar syllable is still available in short-term memory when the new items are introduced. Hence, infants are still able to discriminate the phonetic differences between the syllables. However, the delayed testing situation requires that infants encode attributes about the items into a more permanent form in long-term memory. Because the infants have focused on those aspects of speech relating to voice quality, this information has the highest priority for encoding into the memory representations. Delayed testing requires infants to compare what they are currently hearing to their long-term memory representations of the stimuli heard during the familiarization period. Because voice quality differences figure prominently in those representations, infants will be more apt to detect differences in voice quality dimensions than those relating to phonetic dimensions.

One might argue that the explanation just offered has trouble explaining the results of the single-talker talker change group of Experiment 2. Recall that infants listening to a token of a particular syllable produced by one talker before the delay did not pick up the change to an utterance of the same syllable produced by a new talker. Moreover, this failure to retain information about talker identity was not due to an inability to detect the difference in talkers' voices because infants who heard the same pairs of syllables without the delays *did* discriminate them. However, note that our explanation claims that infants attend to voice quality differences in the multiple-talker and multiple-token situations because these are the dimensions along which the stimuli are varying. When only a single token is played, then voice quality dimensions may not be the ones which attract and hold infants' attention. Instead, there may be a bias to attend to phonetic aspects of the signal. In fact, the pattern observed in the single-talker conditions of Experiment 2 where the phonetic change was picked up, but the talker change was not, is consistent with just such a bias. Obviously, more research is necessary to determine just which features of the speech signal hold infants' attention and under what circumstances. For instance, as noted earlier, the phonetic contrast investigated here is a subtle one; perhaps a different pattern of results would obtain if a more salient phonetic contrast (e.g., [a] vs. [i]) were used. Thus the present study is only a first step in the direction of understanding some factors that affect infants' attention to speech and their consequences for what attributes are encoded into long-term memory.

Knowledge about what information infants retain from speech sounds is certainly critical in understanding how a lexicon develops that serves speech recognition in a native language. Recognizing a word in fluent speech requires that elements in the sound stream activate the

¹⁰As noted earlier, there is a precedent for believing that the nature of the familiarization phase can set the attentional focus of infants and affect the kinds of contrasts they will detect (Jusczyk et al., 1990).

correct stored meaning. It is not obvious how this could be accomplished in the absence of some stored representation of the sound pattern of the word. One of the long-term goals of research on infant speech perception as it relates to the development of the lexicon is to determine the kind of information that goes into the infant's representation of the acoustic-phonetic characteristics of words (see Jusczyk, in press-a, in press-b, for further discussion of this point). If information particular to the actual tokens that are heard figures in the representation then this has certain consequences for models of word recognition. In fact, such a result would be difficult to handle for models that postulate the storage of some prototypical representation of the acoustic-phonetic characteristics of lexical items, because differences among pronunciations of the same word by different talkers is precisely the kind of information that a prototype might be expected to exclude. Instead, exemplar-based models, ones that postulate that listeners store traces of particular utterances that they hear, would be favored by results suggesting that talker characteristics (or even characteristics of individual tokens from a particular talker) are retained in memory (Hintzman, 1986; Nosofsky, 1987, 1988). There are studies with adults suggesting that they may retain information about specific characteristics of utterances they have heard. One example is the Craik and Kirsner (1974) study showing that subjects are faster and more accurate at recognizing items when they are repeated in the same voice as the original item than when they are repeated in a different voice. More recently, Lightfoot (1989) has reported improved recall performance in adults for spoken words produced by familiar talkers. Because exemplar models claim that listeners store traces of previously heard utterances, they provide a straightforward account of why recognition of previously encountered instances from a category tends to be better. At retrieval, all traces are contacted simultaneously, activating each according to its similarity to the input. The information retrieved reflects the summed content of all activated traces responding in parallel. In this way, exemplar-based models are able to account for effects specific to a particular trace and also handle the same range of facts about categories as prototype models (see Hintzman, 1986 for an interesting discussion of this point).

As noted earlier, Grieser and Kuhl (1989) have examined whether 6-month-old infants form prototypes for certain vowel categories. In their study, they compared generalization performance to novel instances from a category after exposure to good (prototypical) and poor exemplars from the category. Performance was significantly better in the case of exposure to the good exemplars. Grieser and Kuhl concluded that their results are consistent with a view that "holds that human infants organize vowel categories around prototypes" (p. 577). While we agree with their general conclusions, we also believe that some of the present findings may have to be taken into account in thinking about this issue. For example, just how specific is the representation for a vowel category? Is it specific to a talker, or even to a gender? Normally, the position that a prototype or general template provides access to lexical entries implies that information relating to a particular talker's voice is excluded from the representation. However, infants in the present study did have access to information relating to the gender of the talkers' voice for at least short intervals. Also, under normal circumstances one would expect that repeated exposure to a diverse set of exemplars from a category would make the detection of a change to a new category more likely than repeated exposure to a single instance from the category (Posner & Keele, 1968). Yet, precisely the opposite occurred in the present experiment. Clearly, it is premature to take a firm stand as to whether an exemplar-based or a prototype model best describes the way in which infants encode speech sounds. Moreover, it is certainly not our intention to criticize Grieser and Kuhl's idea that a prototype description may provide the best account of the way speech sounds are represented by infants. Our point is that any decision in favor of one or the other type of model will be possible only after considering a broad range of facts including the effects that several kinds of stimulus variability has on the way in which speech sounds are recognized. The present study is only a first step in this direction.

In summary, the present study with 2-month-old infants indicates that they are able to compensate for differences in tokens of the same utterance type, regardless of whether the tokens come from the same talker or a variety of different talkers. However, it is also clear that variability among tokens of the same type can affect the ability of infants to encode and retrieve selected attributes of the speech signal. Specifically, phonetic information appears to be less well retained when variability from the same talker or across different talkers is present in the stimulus array. Presenting infants with a series of tokens of a particular syllable with a great deal of stimulus variability may cause them to focus on attributes such as voice quality, as well as phonetic properties, when encoding the items into long-term memory. The demonstration that stimulus variability can affect how infants encode and remember speech sounds has implications for the development of the mental lexicon. In particular, the effects that stimulus variability has on encoding and retrieval of speech may influence the underlying organization of the lexicon.

References

- Allard F, Henderson D. Physical and name codes in auditory memory: The pursuit of an analogy. *Quarterly Journal of Experimental Psychology*. 1976; 28:475–482. [PubMed: 1005652]
- Aslin, RN. Visual and auditory development in infancy. In: Osofsky, JD., editor. *Handbook of infant development*. 2. New York: Wiley; 1987. p. 5-97.
- Aslin, RN.; Pisoni, DB.; Jusczyk, PW. Auditory development and speech perception in infancy. In: Haith, M.; Campos, J., editors. *Handbook of child psychology: Vol. 2. Infancy and developmental psychobiology*. New York: Wiley; 1983. p. 573-687.
- Bertoncini J, Bijeljac-Babic R, Jusczyk PW, Kennedy LJ, Mehler J. An investigation of young infants' perceptual representations of speech sounds. *Journal of Experimental Psychology: General*. 1988; 117:21–33. [PubMed: 2966228]
- Bladon RA, Henton CG, Pickering JB. Towards an auditory theory of speaker normalization. *Language and Communication*. 1984; 4:59–69.
- Carrell TD, Smith LB, Pisoni DB. Some perceptual dependencies in speeded classification of vowel color and pitch. *Perception and Psychophysics*. 1981; 29:1–10. [PubMed: 7243524]
- Cole RA, Coltheart M, Allard F. Memory for a speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*. 1974; 26:1–7. [PubMed: 4814860]
- Craik FIM, Kirsner K. The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology*. 1974; 26:274–284.
- Creelman CD. Case of the unknown talker. *Journal of the Acoustical Society of America*. 1957; 29:655.
- DeCasper AJ, Fifer WP. Of human bonding; Newborns prefer their mothers' voices. *Science*. 1980; 208:1174–1176. [PubMed: 7375928]
- Dechovitz D. Information conveyed by vowels: A confirmation. *Haskins Laboratories Status Report on Speech Research*. 1977; SR-51/52:213–219.
- Disner SF. Evaluation of vowel normalization procedures. *Journal of the Acoustical Society of America*. 1980; 67:253–261. [PubMed: 7354193]
- Eimas, PD. Speech perception: A view of the initial state and perceptual mechanisms. In: Mehler, J.; Garrett, M.; Walker, ECT., editors. *Perspectives on mental representation: Experimental and theoretical studies of cognitive processes and capacities*. Hillsdale, NJ: Erlbaum; 1982. p. 339-360.
- Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. Speech perception in infants. *Science*. 1971; 171:303–306. [PubMed: 5538846]
- Fourcin AJ. Speech-source interference. *IEEE Transactions on Audio and Electroacoustics*. 1968; ACC-16:65–67.
- Gerstman L. Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics*. 1968; ACC-16:78–80.

- Goldinger SD, Pisoni DB, Logan JS. On the locus of talker variability effects on the recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1991; 17:152–162.
- Grieser D, Kuhl PK. The categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*. 1989; 25:577–588.
- Hintzman DL. “Schema Abstraction” in a multiple-trace memory model. *Psychological Review*. 1986; 93:411–428.
- Holmberg, TL.; Morgan, KA.; Kuhl, PK. Speech perception in early infancy: Discrimination of fricative consonants. Paper presented at the meeting of the Acoustical Society of America; Miami Beach, FL. December; 1977.
- Huttenlocher, J. The origins of language comprehension. In: Solso, RL., editor. *Theories in cognitive psychology*. New York: Wiley; 1974. p. 331-368.
- Jusczyk, PW. Infant speech perception: A critical appraisal. In: Eimas, PD.; Miller, JL., editors. *Perspectives on the study of speech*. Hillsdale, NJ: Erlbaum; 1981. p. 113-164.
- Jusczyk, PW. On characterizing the development of speech perception. In: Mehler, J.; Fox, R., editors. *Neonate cognition: Beyond the blooming, buzzing confusion*. Hillsdale, NJ: Erlbaum; 1985a. p. 199-229.
- Jusczyk, PW. The high-amplitude sucking procedure as a methodological tool in speech perception research. In: Gottlieb, G.; Krasnegor, NA., editors. *Infant methodology*. Norwood, NJ: Ablex; 1985b. p. 195-222.
- Jusczyk, PW. Towards a model for the development of speech perception. In: Perkell, J.; Klatt, DH., editors. *Invariance and variability in speech processes*. Hillsdale, NJ: Erlbaum; 1986. p. 1-19.
- Jusczyk, PW. Developing phonological categories from the speech signal. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. *Phonological development: Models, research, implications*. Baltimore, MD: York Press; in press-a
- Jusczyk, PW. How word recognition may evolve from infant speech perception capacities. In: Altmann, G.; Shillcock, R., editors. *Cognitive models of speech perception*. Cambridge, MA: MIT Press; in press-b
- Jusczyk PW, Bertoncini J, Bijeljac-Babic R, Kennedy LJ, Mehler J. The role of attention in speech perception by infants. *Cognitive Development*. 1990; 5:265–286.
- Jusczyk PW, Derrah C. Representation of speech sounds by young infants. *Developmental Psychology*. 1987; 23:648–654.
- Jusczyk PW, Kennedy LJ, Jusczyk AM. Young infants’ memory for information in speech syllables. in preparation.
- Kaplan, EL. Unpublished PhD dissertation. Cornell University; Ithaca, NY: 1969. The role of intonation in the acquisition of language.
- Kuhl PK. Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America*. 1979; 66:1668–1679. [PubMed: 521551]
- Kuhl PK. Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*. 1983; 6:263–285.
- Kuhl, PK. Perception of speech and sound in early infancy. In: Salapatek, P.; Cohen, L., editors. *Handbook of infant perception*. Vol. 2. New York: Academic Press; 1987. p. 275-381.
- Kuhl PK. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception and Psychophysics*. 1991; 50:93–107. [PubMed: 1945741]
- Kuhl PK, Miller JD. Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception and Psychophysics*. 1982; 31:279–292. [PubMed: 7088673]
- Lieberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychological Review*. 1967; 74:431–461. [PubMed: 4170865]
- Lightfoot, N. Research on Speech Perception Progress Report, 15. Indiana University; 1989. Effects of talker familiarity on serial recall of spoken words lists.

- Marean CG, Werner L, Kuhl PK. Vowel categorization in very young infants. *Developmental Psychology*. in press.
- Martin CS, Mullennix JW, Pisoni DB, Summers WV. Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1989; 15:676–684.
- Mehler J, Bertoncini J, Barriere M, Jassik-Gerschenfeld D. Infant recognition of mother's voice. *Perception*. 1978; 7:491–497. [PubMed: 733440]
- Miller CL, Younger BA, Morse PA. Categorization of male and female voices in infancy. *Infant Behavior and Development*. 1982; 5:143–159.
- Miller, JL. Mandatory processing in speech perception. In: Garfield, JL., editor. *Modularity in knowledge and natural-language understanding*. Cambridge, MA: MIT Press; 1987. p. 309–322.
- Miller JL, Eimas PD. Organization in infant speech perception. *Canadian Journal of Psychology*. 1979; 33:353–367. [PubMed: 546500]
- Mills M, Meluish E. Recognition of the mother's voice in early infancy. *Nature*. 1974; 252:123–124. [PubMed: 4420642]
- Morse, PA. Infant speech perception: Origins, processes, and alpha centuri. In: Minifie, FD.; Lloyd, LL., editors. *Communicative and cognitive abilities: Early behavioral assessment*. Baltimore, MD: University Park Press; 1978.
- Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America*. 1989; 85:365–378. [PubMed: 2921419]
- Nearey, TM. Phonetic feature systems for vowels. Paper published by Indiana University Linguistics Club; Bloomington, IN. 1978.
- Nosofsky RM. Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*. 1986; 115:39–57. [PubMed: 2937873]
- Nosofsky RM. Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1987; 14:700–708.
- Nosofsky RM. Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 1988; 15:282–304.
- Posner MI, Keele SW. On the genesis of abstract ideas. *Journal of Experimental Psychology*. 1968; 77:353–363. [PubMed: 5665566]
- Rand TC. Vocal tract size normalization in the perception of stop consonants. *Haskins Laboratories Status Report on Speech Research*. 1971; SR-25/26:141–146.
- Siegel, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill; 1956.
- Siqueland ER, DeLucia CA. Visual reinforcement of non-nutritive sucking in human infants. *Science*. 1969; 165:1144–1146. [PubMed: 5801599]
- Summerfield, Q. *Report of Speech Research in Progress*. Vol. 2. The Queen's University; Belfast, U.K: 1975. Acoustic and phonetic components of the influence of voice changes and identification times for CVC syllables.
- Summerfield, Q.; Haggard, MP. *Report of Speech Research in Progress*. The Queen's University; Belfast, U.K: 1973. Vocal tract normalisation as demonstrated by reaction times.
- Swoboda P, Morse PA, Leavitt LA. Continuous vowel discrimination in normal and at-risk infants. *Child Development*. 1976; 47:459–465. [PubMed: 1269313]
- Syrdal AK, Gopal HS. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*. 1986; 79:1086–1100. [PubMed: 3700864]
- Trehub SE. The discrimination of foreign speech contrasts by infants and adults. *Child Development*. 1976; 47:466–472.
- Turnure C. Response to the voice of mother and stranger by babies in the first year. *Developmental Psychology*. 1971; 4:182–190.

- Verbrugge RR, Strange W, Shankweiler DP, Edman TR. What information enables a listener to map a talker's vowel space? *Journal of the Acoustical Society of America*. 1976; 60:198–212. [PubMed: 956527]
- Werker, JF. The ontogeny of speech perception. In: Mattingly, IG.; Studdert-Kennedy, M., editors. *Modularity and the motor theory of speech perception*. Hillsdale, NJ: Erlbaum; 1990. p. 91-109.

Figure 1.

Mean increase in sucking during the test phase for subjects in the talker variability study under the immediate testing condition. Subjects in the single-talker conditions heard the same token throughout the familiarization phase and either the same (no-shift control) or a different token (phonetic change and talker change) during the test phase. Subjects in the multiple-talker conditions heard multiple versions of the syllables during both the familiarization and test phases. (The scores are determined by subtracting the average sucking rates from the last two preshift minutes from the average of the first two postshift minutes.)

Figure 2.

Mean increase in sucking during the test phase for subjects in the talker variability study under the delayed testing condition. Subjects in the single-talker conditions heard the same token throughout the familiarization phase and either the same (no-shift control) or a different token (phonetic change and talker change) during the test phase. Subjects in the multiple-talker conditions heard multiple versions of the syllables during both the familiarization and test phases. (Note that no infants were tested in a multiple-talker talker change condition because infants in the previous experiment failed to discriminate this contrast even without the delay.)

Figure 3. Mean increase in sucking during the test phase for each of the multiple-talker conditions after the 2-min delay period. The phonetic change and talker gender change groups heard different tokens during the two phases of the experiment, whereas the no-shift control group continued to hear the same set of tokens throughout.

Figure 4.

Mean increase in sucking for the test phase for subjects in the token variability study under the immediate testing condition. Subjects in the single-token conditions heard the same token throughout the familiarization phase and either the same (no-shift control) or a different token (token change) during the test phase. Subjects in the multiple-token conditions heard multiple versions of the same syllable during the familiarization phase. During the test phase, the no-shift control group continued to hear the same tokens, whereas the phonetic change group was switched to tokens of the other syllable.

Figure 5.

Mean increase in sucking for the test phase for subjects in the token variability study under the delayed testing condition. Subjects in the single-token conditions heard the same token throughout the familiarization phase and either the same (no-shift control) or a different token (phonetic change) during the test phase. Subjects in the multiple-token conditions heard multiple versions of the same syllable during the familiarization phase. During the test phase, the no-shift control group continued to hear the same tokens, whereas the phonetic change group was switched to tokens of the other syllable.

Table 1

Examples of what is heard in the various conditions of Experiment 1

	Preshift phase	Postshift phase
<i>Single-talker conditions</i>		
Phonetic change	bug (male 1)	dug (male 1)
Talker change	bug (male 1)	bug (male 5)
Control	bug (female 4)	bug (female 4)
<i>Multiple-talker conditions</i>		
Phonetic change	bug (f3), bug (m5),bug(f4),	dug (f4), dug (m3),dug(f2),
	bug (m6), bug (fl), bug (m3),	dug (m6), dug (fl), dug (m4),
	bug(m5),bug(f2),bug(f5),...	dug (m5),dug(f6),dug(f2),
Talker change	bug (f4), bug (m3),bug(f2),	bug (f3), bug (m2),bug(f6),
	bug (m6), bug (fl), bug (m4),	bug (m5), bug (f5), bug (m1),
	bug (m4), bug (fl), bug (f4),...	bug (m1), bug (f6), bug (f5),
Control	bug (m2),bug(f3),bug(f6),	bug (m4),bug(f3),bug(f2),
	bug (m3), bug (m1), bug (f4),	bug (m1), bug (fl), bug (m4),
	bug(f5),bug(f6),bug(m1),...	bug (m3), bug (f5), bug (m2)