Journal of Experimental Psychology:
Human Perception and Performance
2000, Vol. 26, No. 5, 1570–1582

0096-1523/00/$5.00 DOI: 10.1037//0096-1523.26.5.1570

The Role of Talker-Specific Information in Word Segmentation by Infants

Derek M. Houston and Peter W. Jusczyk
Johns Hopkins University

Infants' representations of the sound patterns of words were explored by examining the effects of talker variability on the recognition of words in fluent speech. Infants were familiarized with isolated words (e.g., *cup* and *dog*) from 1 talker and then heard 4 passages produced by another talker, 2 of which included the familiarized words. At 7.5 months of age, infants attended longer to passages with the familiar words for materials produced by 2 female talkers or 2 male talkers but not for materials by a male and a female talker. These findings suggest a strong role for talker-voice similarity in infants' ability to generalize word tokens. By 10.5 months, infants could generalize different instances of the same word across talkers of the opposite sex. One implication of the present results is that infants' initial representations of the sound structure of words not only include phonetic information but also indexical properties relating to the vocal characteristics of particular talkers.

A significant part of language acquisition is learning relations between spoken words and concepts. The challenge for the language learner is to extract words from the speech signal and match them to the appropriate referents. A recent investigation of speech directed to an infant between 6 and 9 months of age indicates that only about 7% of the utterances (excluding vocatives, fillers, and social expressions) consist of isolated words (van de Weijer, 1998). Moreover, even when explicitly asked to teach their infants new words, mothers produce words in isolation no more than 20% of the time (Woodward & Aslin, 1990). Thus, many words must be segmented from fluent speech if they are to be learned.

Two main issues bear on the problem of linking sounds and meanings. One is how a child correctly associates the sound patterns of words to the appropriate concepts, given the enormous number of choices that are potentially available (Clark, 1973, 1983; Markman, 1991; Quine, 1960). The other issue involves the complexity and variability of speech itself. Individual words must be perceived from a continuous speech signal that does not reliably mark word boundaries, and each instance of a word must be recognized as equivalent to other acoustically different occurrences (Goldinger, Pisoni, & Luce, 1996; Klatt, 1979). The present set of studies focuses on an important aspect of the second issue, namely, the effect of variability in the input on infants' word representations. In particular, we examine whether infants can recognize words first heard from one talker when they occur in fluent speech produced by a different talker.

The speech signal is rich with information. It carries not only the phonetic and prosodic information of language but also nonlinguistic or indexical information. From speech, a listener can determine the identities of talkers, their physical and emotional state, their sex, their regional dialect, and other qualities that reflect the physiological characteristics of their vocal apparatus and the dynamics of their articulation. Because all this information is contained in the same acoustic signal, any one aspect or type of information is not physically isolated from the rest (Klatt, 1989). Consequently, it is not straightforward to understand how a listener extracts phonetic information from the complex array present in speech.

The traditional approach to the talker variability problem has been to assume that the perceptual processing of speech involves some normalization procedures. One approach has been to assume the existence of invariant acoustic features specifying the identity of particular phonemes (Blumstein & Stevens, 1980; Fant, 1960; Stevens, 1972; Stevens & Blumstein, 1981). By this view, the perceiver normalizes speech by focusing on these invariant features and ignores extraneous information having to do with the articulatory characteristics of particular talkers. The phonetic information in the signal is then compared with idealized abstract representations in the lexicon. This type of normalization process would incur little cost associated with changing talkers or encountering an unfamiliar talker because the detectors are sensitive only to the linguistically relevant aspects of the speech signal. The major challenge that this approach faces is finding evidence for the existence of a complete set of invariant properties that can unambiguously identify all phonetic segments in the speech signal. So far, this set of invariant properties has remained elusive (Klatt, 1989).

Derek M. Houston, Department of Psychology, Johns Hopkins University; Peter W. Jusczyk, Department of Psychology and Department of Cognitive Science, Johns Hopkins University.

This work was supported by research Grant 15795 from the National Institute of Child Health and Human Development, Senior Scientist Award 01490 from the National Institute of Mental Health (NIMH), and NIMH Predoctoral National Research Scientist Award 12232.

We would like to thank Ann Marie Jusczyk, Amy Gambon-Dennis, Debra Dombrowski, and Eileen Crowley for their help in recruiting and testing infants. In addition, Theresa Allocco and Alison Leavitt provided valuable assistance for the acoustic analyses. Finally, we are grateful to Ann Marie Jusczyk and Sven Mattys for comments on earlier versions of this article.

Correspondence concerning this article should be addressed to Derek M. Houston, who is now at the Indiana University Medical School, Department of Otolaryngology—Head and Neck Surgery, DeVault Otologic Research Laboratory, 699 West Drive, Indianapolis, Indiana 46202. Electronic mail may be sent to dmhousto@indiana.edu.

In a related approach, invariant properties are postulated to be assessed via intrinsic normalization mechanisms (Nearey, 1989). That is, there may be static (e.g., Syrdal & Gopal, 1986) and/or dynamic (Strange, 1989) properties in speech that provide the perceptual system with a framework for linguistic interpretation. For example, formant frequency variations between different talkers' productions of vowels may be normalized by automatic auditory mechanisms that use properties in the speech signal (e.g., pitch) to guide the interpretation of formant frequencies (Suomi, 1984). Moreover, it has been proposed that innately specified neuronal cell assemblies may subserve this type of normalization (Sussman, 1984, 1986). Specifically, this approach posits that separate cell assemblies encode for the absolute formant frequencies and relative formant frequencies. These cell assemblies connect to higher order cell assemblies to eliminate information related to vocal tract size and to derive invariant properties. Hence, this biological view assumes an innate normalization mechanism that takes vocal tract information into account when interpreting vowels, predicting that even infants should perceive phonological equivalence across talker variability.

Another view of normalization posits that listeners make perceptual adjustments for characteristics of talkers' vocal tracts and speaking behaviors (e.g., Gerstman, 1968; Shankweiler, Strange, & Verbrugge, 1977). This perceptual adjustment requires recognizing specific indexical qualities of the talker. Once talker-specific information is assessed, the perceptual system becomes adjusted to interpret speech from a particular talker. Such a view necessitates two mechanisms for speech perception: one that encodes the acoustic properties and another that encodes abstract phonetic features.

Evidence for separate mechanisms derives from findings that nonphonological acoustic details play a role in some speech-processing tasks but not others. For example, Schacter and Church (1992) found voice-repetition priming effects in an implicit memory task but not in an explicit memory task. Furthermore, masking speech stimuli in white noise eliminated the voice priming effect. Schacter and Church (1992; Church & Schacter, 1994) interpreted their results as indicating that voice information is processed by a mechanism that can be selectively interfered with by destroying aspects of the acoustic signal while preserving the mechanism involved with processing more abstract phonetic information. They postulate separate subsystems of the perceptual representation systems (PRS) for phonetic and acoustic information. Moreover, Schacter, Church, and Bolton (1995) found that amnesic patients sometimes demonstrate impaired voice repetition priming but intact word repetition priming, suggesting that the phonetic and acoustic PRS may be subserved by different parts of the brain.

Whereas research by Schacter, Church, and others points to different processing for phonetic and indexical information, other evidence suggests that indexical and phonetic information may be more tightly coupled in processing and memory. Specifically, some studies of word recognition suggest that talker-specific information is encoded along with phonetic information and facilitates word recognition in certain settings. For example, Craik and Kirsner (1974) presented participants with words produced by a male and female talker. The participants had to respond after each word whether or not it had been presented to them before. Word recognition was found to be faster and more accurate when words were produced by the same talker than when they were produced by the opposite-sex talker. Palmeri, Goldinger, and Pisoni (1993) extended the findings by showing that words are better recognized when produced by the same talker rather than different talkers, regardless of the sex of the talkers.

These studies suggest that talker-specific information plays some role during language processing. Furthermore, the impact of talker variability is not merely that it necessitates a perceptual adjustment to novel talkers. Participants in the studies described above became equally familiar with all of the talkers. The crucial factor in whether or not a particular word was recognized in these studies was whether the repetitions of a word were by the same or different talkers. The implication of these findings is that listeners appear to store talker-specific information along with the phonetic properties of words.

These studies and others that have found effects of talker variability (e.g., Mullennix, Johnson, Topcu-Durgun, & Farnsworth, 1995; Sheffert & Fowler, 1995) have led some researchers to consider the lexicon as an episodic or exemplar memory system, whereby each occurrence of a word leaves a detailed trace in memory. So rather than being peeled away by perceptual processes, indexical information is encoded along with phonetic information (Goldinger, 1996, 1998; Jusczyk, 1993, 1997; Nygaard, Sommers, & Pisoni, 1994). The exact nature of the indexical information encoded into these representations is not certain. However, it is clear that such representations are not veridical copies of utterances. For example, there are indications that some acoustic properties, such as the loudness of a particular spoken word, do not appear to be preserved in representations of spoken words (Church & Schacter, 1994; Sommers, Nygaard, & Pisoni, 1994). Indeed, there is some suggestion that the phonetic properties of words may provide information about the identity of particular talkers. Remez, Fellowes, and Rubin (1997) found that their listeners were able to extract information about talker identity from sinewave speech stimuli, even though these stimuli lack the acoustic attributes of natural voice quality. They interpreted their finding as an indication that the phonetic properties of words may themselves encode the necessary information to identify both words and talkers.

The effects of indexical information on speech processing by infants have chiefly been explored by investigations of how talker variability affects the detection of certain phonetic contrasts. In her seminal investigation, Kuhl (1979) showed that 6-month-olds have some capacity to ignore talker variability in discriminating between the vowels [a] and [i]. Specifically, the infants were trained to discriminate a contrast between these vowels produced by a single talker and then were able to transfer to a wide set of different talkers, encompassing both men and women. In a subsequent study, Kuhl (1983) found that 6-month-olds also showed some ability to compensate for talker differences in discriminating a more confusable pair of vowels, [a] and [ɔ]. Similarly, Jusczyk, Pisoni, and Mullennix (1992) found that 2-month-olds could detect syllable changes across different talkers. In one condition, infants were presented with syllables produced either by a single or multiple talkers until they reached a habituation criterion. Immediately following, the experimental groups were presented with novel syllables. The infants in both the single- and multiple-talker conditions were able to detect the change. However, further experiments revealed that when a 2-min delay was introduced between habituation and the presentation of the new syllables, only

the infants in the single-talker condition indicated that they detected the change of syllables.

The results just reviewed clearly show that young infants can detect phonetic change in the midst of talker variability. This pattern is consistent with the biological view, which posits innate neural mechanisms that normalize speech with respect to talker variability. Still, these results do not rule out the possibility that infants store talker-specific information in their representations of words. In fact, the difference in results between the single-talker and multiple-talker conditions in the Jusczyk et al. (1992) 2-min-delay condition suggests that representations of words elicited from multiple talkers are not identical to those from a single talker. It is conceivable that representations based on multiple-talkers may be less well encoded than those based on a single talker because encoding of the former is disrupted by increased demands on a normalization mechanism (Mullennix, Pisoni, & Martin, 1989). However, representations based on multiple talkers may also differ from those based on a single talker because indexical information is an integral part of the representation. Indeed, Jusczyk's word recognition and phonetic structure acquisition (WRAPSA) model (1993, 1997) postulates that infants' lexical representations are talker specific. That is, infants store instances of words that include indexical information, as opposed to general prototypes that include only talker-neutral, abstract phonetic descriptions.

What information listeners encode into lexical representations has implications for language acquisition. If indexical information is encoded with the words children are acquiring, then a word learned in one instance may not be immediately generalizable to other instances of that word. The amount of indexical variability in the infant's environment would thus play a large role in developing a mental lexicon. Exploring the effects of indexical information during the formation of the mental lexicon may also shed some light on the extent to which talker-specific information is encoded during normal speech encoding. It could be argued that adults in word-recognition experiments may focus on more surface details of spoken words if they know or suspect that they will have to recall hearing those specific instances later. It is unlikely that infants form similar expectations.

The experiments reported here focus on the nature of infants' representations of words as they begin to develop a lexicon. Do infants represent words in such a way that indexical information affects their ability to recognize different instances of the same word? Or, do infants represent only an abstract phonetic description of words in the lexicon? We addressed these questions by testing infants' ability to generalize words across different talkers. In four experiments, infants were familiarized with words spoken by one talker and then were presented with passages (two of which contained familiarized words) by another talker to determine if they recognized these words in fluent speech contexts. By exploring the conditions under which infants recognize novel instances of familiarized words, we hope to shed some light on the nature of infants' representations of the sound structure of words.

Experiment 1

To test infants' ability to generalize words across different talkers, we used a method used by Jusczyk and Aslin (1995) to explore infants' ability to recognize words in the context of fluent speech. They found that 7.5-month-olds but not 6-month-olds oriented longer to passages with words they had been familiarized with than to passages without these words, suggesting that 7.5-month-old English-learning infants can extract words from fluent speech and recognize them. Their investigation also demonstrated that infants were able to generalize representations formed during familiarization to acoustically different instances heard during the test period. For instance, the words that the infants were familiarized with were typically spoken in citation form, whereas the acoustic characteristics of the same words presented in the passages during the test phase were influenced by their surrounding words. The infants did generalize from the familiarization words to the test words. Moreover, their representations of the phonetic properties of the words proved to be quite specific. When 7.5-month-olds were familiarized with nonwords that differed by one phoneme from the word in the test passage (e.g., *gike* rather than *bike*), they did not respond to these items. Consequently, Jusczyk and Aslin concluded that infants encoded detailed information about the phonetic characteristics of words heard during the familiarization period (see also Tincoff & Jusczyk, 1996).

Infants appear to represent the phonetic properties of words well enough so that they do not treat words that differ only by a single phoneme as being the same. Does the same hold true for the indexical properties of speech? In the natural course of processing and encoding speech, infants may form representations of words that do not contain talker-specific information. In essence, this processing would amount to an extraction of an abstract phonetic description of the speech signal. However, if infants encode speech without some kind of normalization process, then talker-specific information may be included in their word representations.

As in Jusczyk and Aslin (1995), we familiarized infants with different tokens of two words produced by a female talker. During the test phase, the infants heard four blocks of trials. Each block consisted of a different random ordering of the same four passages (two with the familiarized words and two without). However, in contrast to Jusczyk and Aslin's study, the female talker who recorded the passages differed from the one who recorded the familiarization items. If infants form abstract representations of words (i.e., without talker-specific information), then they should recognize the familiarized words across the different talkers. However, if talker-specific information is encoded into the phonetic representations of particular words, then there is some chance that the familiarized words in the passages may not be recognized when produced by a different talker. On the basis of previous findings (Jusczyk & Aslin, 1995; Jusczyk, Hohne, & Bauman, 1999; Jusczyk, Houston, & Newsome, 1999), infants should, on average, orient longer to the passages containing the familiarized words than the unfamiliar passages if they can recognize the familiarized words. Finally, there is a third possibility. Changing the talker between the familiarization and test phases could divert some attention at first from normal language processing. Listeners may need to perceptually adjust to the voice of a new talker, and so infants may have difficulty recognizing the familiarized words during the first few trials but then detect them on later trials. In this case, longer listening times to passages with the familiarized words would only be apparent during the last blocks.

Method

Participants. Thirty-six American 7.5-month-olds from monolingual English-speaking families were tested. The infants had a mean age of 32 weeks, 6 days (range = 30 weeks, 6 days to 35 weeks, 2 days). Twelve additional infants were tested but not included because of crying (3), failure to complete the full set of test trials due to restlessness (5), failure to look for an average of at least 3 s to each stimulus type (2), extreme left-side bias (1), and equipment failure (1). All were recruited from families living in the greater Buffalo, New York area.

Stimuli. The stimuli consisted of repetitions of isolated words and passages produced by two different talkers. The words and passages were the same as in Jusczyk and Aslin (1995). A passage of six sentences was constructed for each of the four words (*cup, dog, feet, bike*). The target word occurred once in each sentence in variable sentence position (see Appendix). Within each passage, the target word was always preceded by a different word in each sentence. Across the passages, the words preceding the targets were the same (i.e., *his, red, the, big, old, your*). The words following the targets differed for each target type. We used the same recordings as Jusczyk and Aslin (1995) for one of the talkers (Female Talker 1). The new female talker (Female Talker 2) was instructed to read the four passages as if she was speaking to a young child. Next, she was asked to repeat each of the four words (*cup, dog, feet, bike*) successively 15 times with varied intonation, as if speaking to a young infant. These were the same instructions followed by the talker in Jusczyk and Aslin (1995).

For Female Talker 1, the average duration of the passages was 19.72 s (ranging from 18.51 s for the *bike* passage to 20.60 s for the *feet* passage). The average duration of the lists was 26.53 s (ranging from 25.84 s for the *feet* list to 27.13 s for the *dog* list). For Female Talker 2, the average duration of the passages was 19.36 s (ranging from 18.96 s for the *cup* passage to 20.12 s for the *feet* passage). The average duration of the lists was 18.55 s (ranging from 18.34 s for the *feet* list to 18.88 s for the *bike* list).

Apparatus. The experiment was conducted in a three-sided test booth constructed of pegboard, with panels of 4 × 6 ft on three sides and open at the back. This structure made it possible for an observer to look through one of the existing holes to monitor the infant's headturns. Except for a small section for viewing the infant, the remainder of the pegboard was backed with white cardboard to guard against the possibility that the infant might respond to movements behind the panel. The test booth had a red light and a loudspeaker mounted at eye level on each of the side panels, and a green light mounted on the center panel. A white curtain suspended around the top of the booth shielded the infant's view of the rest of the room. A Macintosh Centris 650 computer and response box were located behind the center panel, out of view of the infant. The response box, which was connected to the computer, was equipped with a series of buttons that started and stopped the flashing center and side lights, recorded the direction and duration of headturns, and terminated a trial when the infant looked away for more than 2 s. Information about the direction and duration of headturns and the total trial duration were stored in a data file on the computer.

Procedure. We used the headturn preference procedure, as modified by Jusczyk and Aslin (1995), to test the infants. Half of the infants were familiarized with words produced by Female Talker 1 and were tested on passages produced by Female Talker 2; the other half heard Female Talker 2 in the familiarization phase and Female Talker 1 during the test phase. Each infant sat on the lap of a caregiver who was seated on a chair in the center of the test booth. At the beginning of each trial, the center light flashed until the infant oriented to the center. Then, the center light was turned off, and one of the side lights began flashing. When the infant oriented at least 30° in the direction of the light, the speech stimulus was presented to the same side as the flashing light. The stimulus continued until the infant looked away for 2 s or until the end of the trial. The amount of time the infant oriented to the stimulus side while the stimulus was playing was recorded for each trial.

The experiment had two phases. During the familiarization phase, on a given trial, each infant was presented with different tokens of one of the two familiarization words. Half of the infants heard *cup* and *dog*, and the other half heard *bike* and *feet*. The familiarization words were presented on alternating trials until at least 30 s of looking time was accumulated for each familiarization word. During the test phase, all four passages were presented once in each of four blocks. The order of the passages within each block was randomized. For each participant, two of the passages contained the target words presented during familiarization, and the other two passages contained target words not heard during familiarization. An average orientation time difference between the passages with the familiarized and unfamiliar target words is taken as an indication that the infants differentiated the two types of passages, presumably because they recognized the familiarized words in the passages.

An observer hidden behind the center panel looked through a peephole and recorded the direction and duration of the infant's headturns using a response box. The observer was not informed of which items served as familiarization words for a given infant. The loudness level for the samples was set at 72 ± 2 dB (C) SPL using a Quest (Model 215) sound level meter by an assistant who was not involved in the observations. During the experiment, both the observer and the caregiver listened to music over tight-fitting closed headphones (SONY MDR-V600), so they were unaware of which particular stimulus was presented at any given time. Reliability checks between the live observer and observers of video tapes of test sessions are high, with correlations ranging from .92 to .96 (Kemler Nelson et al., 1995).

Results and Discussion

The infants' average orientation time to each word type (familiar/unfamiliar) was computed for each block of trials. A two-way analysis of variance (ANOVA) (word familiarity and test block) revealed main effects of word familiarity, $F(1, 35) = 4.21$, $p < .05$, and test block, $F(3, 105) = 18.89$, $p < .001$. There was no Word Familiarity × Test Block interaction, $F(3, 105) < 1.00$. Figure 1 displays the average orientation times across all four blocks of trials to the passages containing the familiarized

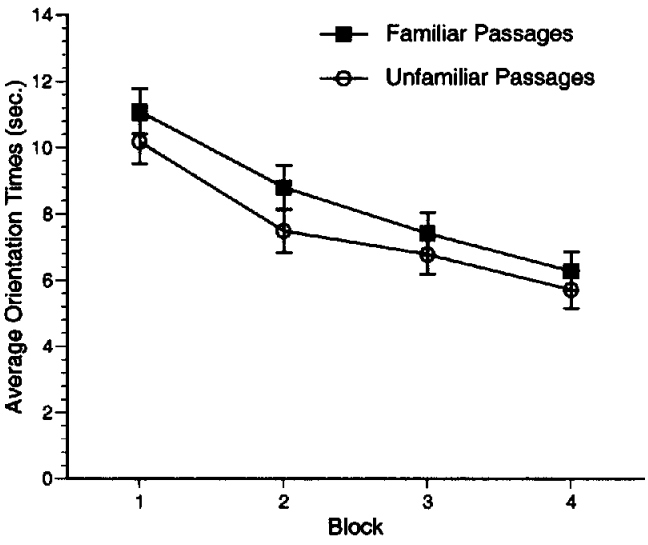

Figure 1. Displays 7.5-month-olds' average orientation times and standard error (±1 *SE*) to test passages in Experiment 1 across four blocks of trials. Passages were produced by a female talker different from the female talker used during the familiarization phase.

(M = 8.39 s; SD = 2.02 s) and the unfamiliar (M = 7.53 s; SD = 2.74 s) target words. The means differed by 0.86 s (95% CI: 0.05 < 0.86 < 1.69). Overall, 24 of the 36 infants oriented longer to the familiar passages.

The results obtained in this experiment essentially replicate those reported by Jusczyk and Aslin (1995) for 7.5-month-olds. In both cases infants oriented longer to the passages containing the familiarized words than the passages that did not, indicating that they noticed the occurrence of the familiarized words in the passages. The main effect of test block reflects a steady decrease in orientation times as the infants became less attentive over the course of the test phase. Therefore, the present findings suggest that infants' representations of words can be generalized to new tokens spoken by another talker of the same sex. At least to some degree, infants are able to overcome acoustic variability in speech produced by different talkers and to generalize from words presented in isolation to the same words in the passages. Moreover, the absence of any significant interaction between word familiarity and test blocks suggests that any perceptual adjustment was evident early in the test phase and did not develop across the test blocks.

One possible interpretation of the present results is that infants represent words in the lexicon without their indexical properties (i.e., that only an abstract phonetic description of words is encoded). However, the findings do not rule out the possibility that indexical information is preserved in the lexicon (or elsewhere) or that perceptual systems adjust to talker differences. The two talkers in the present experiment were both female, from the same region of the country, and were very close in age. The infants may not have had difficulty recognizing the similarity of the same words spoken by different talkers because the indexical properties of the two talkers were similar. This raises the question of whether infants might have more difficulty generalizing among productions of the same word produced by talkers whose voices are more distinctive from one another, namely, a male versus a female talker. To explore this possibility, we conducted the following experiment.

Experiment 2

At first glance, findings from several previous investigations provide some basis for predicting that infants may generalize the productions of words from a talker of one sex to those of a talker of the opposite sex. For instance, in her pioneering investigation, Kuhl (1979) found that 6-month-olds were able to maintain a discrimination between two vowels, [a] and [i], even when the range of talkers was increased to encompass both men and women. A subsequent study by Jusczyk et al. (1992) demonstrated that even younger infants (i.e., 2-month-olds) correctly detected a consonantal contrast between [b] and [d] when these sounds were produced by a set of 12 different talkers that included both male and female adults. However, Jusczyk et al. (1992) also found that talker variability did appear to impair 2-month-olds' retention of information about the phonetic contrasts, particularly when a short delay interval was introduced between familiarization with tokens of one of the syllables and testing on the tokens of the other syllable. This task required the infants to retain some memory trace of the familiarization over the delay interval. Jusczyk et al. (1992) found that infants who had been exposed to tokens produced by many different talkers no longer detected the phonetic contrast, whereas those who were exposed to a pair of tokens produced by a single talker did detect the contrast. This last finding raises the possibility that acoustic variability in the production of the same words by distinctive talkers does affect how infants process and retain speech information.

To investigate this issue in the context of developing word-segmentation abilities, we familiarized 7.5-month-olds with words from a male or female talker and then tested them on passages recorded by the talker of the opposite sex. If infants do not encode indexical information into their representations of words, then despite the greater acoustic differences between the voices of the talkers in the present experiments, infants should perform much the way they did in Experiment 1. That is, they should orient significantly longer to the passages containing the familiarized words. Alternatively, if infants do retain indexical information in their representations of the familiarization words, the greater acoustic dissimilarity between these words and the comparable ones in the passages may interfere with infants' detection of the latter.

Method

Participants. Thirty-six American 7.5-month-olds from monolingual English-speaking families were tested. The infants had a mean age of 32 weeks, 6 days (range = 30 weeks, 3 days to 36 weeks, 0 days). Fourteen additional infants were tested but not included due to crying (3), failure to complete the full set of test trials (6), failure to look for an average of at least 3 s to each stimulus type (3), and looking only to one of the sides of the test apparatus (2). All were recruited from families living in the greater Buffalo, New York area.

Stimuli. The stimuli consisted of repetitions of isolated words and passages by a male and a female talker. The female-talker stimuli were identical to those of Female Talker 2 in Experiment 1. The male talker (Male Talker 1) recorded the same words and passages and was instructed to produce these as if speaking to a young child. For Male Talker 1, the average duration of the passages was 19.09 s (ranging from 18.27 s for the *bike* passage to 19.74 s for the *dog* passage). The average duration of the lists was 20.27 s (ranging from 19.86 s for the *bike* list to 20.65 s for the *feet* list).

Apparatus and procedure. These were identical to Experiment 1.

Results and Discussion

The infants' average orientation time to each word was computed. As in Experiment 1, an ANOVA revealed a main effect of test block, $F(3, 105) = 10.46$, $p < .001$. However, there was no main effect of word familiarity, $F(1, 35) < 1.00$. The Test Block × Word Familiarity interaction was also not significant, $F(3, 105) < 1.00$. Figure 2 displays the average orientation times for the passages with the familiarized (M = 7.73 s; SD = 2.46 s) and the unfamiliar (M = 7.72 s; SD = 2.53 s) target words. The means differed by 0.001 s (95% CI: −0.59 < 0.001 < 0.59). Overall, 22 of the 36 infants oriented longer to the passages containing the familiarized words.

The results indicate that 7.5-month-olds did not orient longer to the passages with the familiarized than those with the unfamiliar target words when the passages were spoken by a talker of the opposite sex to the one heard during familiarization. These findings suggest that the infants in the present experiment had more difficulty generalizing words across talkers whose voices were

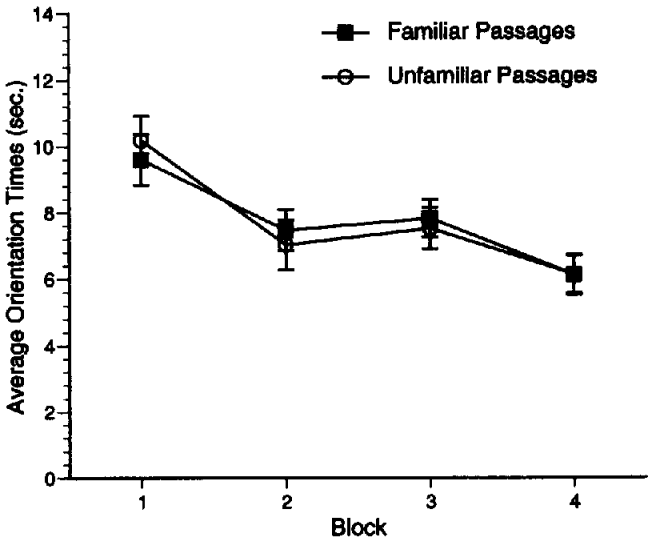

Figure 2. Displays 7.5-month-olds' average orientation times and standard error (±1 *SE*) to test passages in Experiment 2 across four blocks of trials. Passages were presented by either a male or female talker—the opposite-sex talker than whichever was used during the familiarization phase.

more distinctive from each other than the similar-sounding female talkers in Experiment 1. Apparently, the greater distinctiveness of the talkers used in the present experiment interfered with 7.5-month-olds' ability to recognize the familiarized words when they occurred in the test passages. Hence, there was no indication that the infants were able to ignore talker differences in the present experiment. Instead, the differences in the indexical properties did affect the likelihood that infants recognized the familiarized words in the test passages. Moreover, note that there was no indication that infants' ability to recognize the familiarized words improved across the four test blocks. Although this last observation does not preclude the possibility that infants can eventually make perceptual adjustments for different talkers' voices, it suggests that the process of making such adjustments is quite protracted for infants at this age. By comparison, the present findings appear to be more consistent with the view that infants include indexical information in their representations of words.

Given that the present line of argument depends crucially on the assumption that the talkers used in Experiment 2 were more distinctive than those used in Experiment 1, it would be useful to have some empirical support for this assumption. To examine the differences between the talkers' voices, we conducted acoustic analyses of the test words to ascertain which properties of the voices contributed to our subjective judgment that the two female talkers were more similar to each other than the male and female talkers. The analyses focused mainly on the pitch and duration of the test words in both isolated and sentential contexts. These two acoustic dimensions were selected because previous research has found that pitch difference consistently plays a role in word recognition and vowel identification (Carrell, Smith, & Pisoni, 1981; Wannemacher, 1995) and that duration also can play a role in word recognition (Sommers et al., 1994; Tomiak, Green, & Kuhl, 1991). Furthermore, Walden, Montgomery, Gibeily, Prosek, and Schwartz (1978) found that pitch and word duration were the most important (out of 14) acoustic properties for judgments of talker similarity. Using CSL software developed by Kay Elemetrics, we measured the duration, mean pitch, and peak pitch for every instance of the four test words for each talker.

The mean values for each dimension for each talker are shown in Table 1. The comparisons of most interest were those that compared the differences between the two female talkers with the differences between the male and female talkers. The mean values were subjected to an ANOVA. The two talkers in each experiment were treated as a repeated-measures variable, and the experimental condition (i.e., two female talkers vs. one male/one female talkers) was treated as a between-subjects variable. For mean pitch, there was a main effect of experimental condition, $F(1, 163) = 37.48$, $p < .001$, reflecting higher mean pitch of the talkers in Experiment 1 than Experiment 2. More importantly, there was a significant interaction between experimental condition and talker, $F(1, 163) = 4.40$, $p < .04$. The analysis of pitch peak showed the same pattern as with mean pitch. There was a main effect of experimental condition, $F(1, 163) = 33.86$, $p < .001$ and a significant interaction between experimental condition and talker, $F(1, 163) = 6.86$, $p < .01$. The analyses indicate that with respect to mean pitch and pitch peak, the talkers in Experiment 1 were more similar to each other than the talkers in Experiment 2. By comparison, another analysis indicated the talkers did not differ across experiments in overall duration of the target words. For duration, neither a main effect of experimental condition, $F(1, 163) < 1.00$, nor the experimental condition/talker interaction, $F(1, 163) < 1.00$, was significant. Therefore, there was no evidence that the durations of the target words produced by the talkers in Experiment 1 were more similar to each other than were those produced by the talkers in Experiment 2. Overall, the results of the acoustic analyses provide support for the view that the talkers in Experiment 2 were more dissimilar from those in Experiment 1. However, this dissimilarity is manifested in pitch differences of the talkers' voices rather than in any durational difference in their productions of the target words.

Perceptual judgments from adult listeners were elicited as a further check on whether the two female talkers were more similar to each other than were the male and female talkers. These similarity ratings were obtained as part of a study that included a larger array of different male and female talkers (Houston, 1999). Listeners made judgments about the similarities of isolated words

Table 1
Acoustic Analyses of the Target Words (Sentential and Citation Contexts)

Talker	Mean duration (ms)	Mean pitch (Hz) & *SD*[a]	Pitch peak (Hz)
Female 1: Exp. 1	388	345 (67)	422
Female 2: Exps. 1, 2, & 4	547	306 (53)	382
Male 1: Exps. 2, 3, & 4	372	227 (39)	280
Male 2: Exp. 3	511	182 (37)	232

Note. Exp. = experiment.
[a] *SD* is given in parentheses.

produced by the talker pairs in Experiments 1 and 2. Perceptual distances were derived from a multidimensional scaling program (ALSCAL, using SPSS). The mean squared distance between the two female talkers of Experiment 1 was 2.33, whereas the mean squared distance between the male and female talkers of Experiment 2 was 6.64. Thus, the perceptual judgment data confirm the trends noted in the acoustic analyses: The female talkers were more similar in their vocal characteristics than were the male and female talkers.

We have argued that the greater dissimilarity of the talkers in the present experiment was responsible for the failure of the 7.5-month-olds to recognize the familiarized words when they occurred in the passages spoken by a talker of the opposite sex. However, it is worth considering another possible explanation for the pattern of results obtained. Infants may simply have had more difficulty processing the speech from the male talker than from the female talker. Consequently, the infants may have been unable to extract from the speech of the male talker the critical information for recognizing the target words. In fact, previous investigations of word segmentation abilities of infants have used either natural female voices (e.g., Jusczyk & Aslin, 1995; Jusczyk, Hohne, et al., 1999; Jusczyk, Houston, et al., 1999) or synthetic versions of female voices (Saffran, Aslin, & Newport, 1996). Furthermore, when we examined the overall orientation times to the male and female talkers in Experiment 2, there was a slight, though nonsignificant, tendency to orient longer to the passages presented with the female talker (M = 8.04 s, SD = 2.41 s) than the male talker (M = 7.41 s, SD = 2.30 s). Thus, to evaluate the ability of 7.5-month-olds to segment words from fluent speech produced by male talkers, we conducted the following experiment.

Experiment 3

In the present experiment, we decided to familiarize infants with words produced by one male talker and then test them on passages recorded by a different male talker. This experimental manipulation had two purposes. First, it allowed us to test whether the null result in Experiment 2 was due to a general inability of 7.5-month-olds to segment words from speech produced by male talkers. Second, the experimental design provided an opportunity to replicate the findings of Experiment 1 but with two male talkers rather than with two female talkers. A finding that the infants listen longer to the passages with the familiarized words would not only demonstrate that they can segment words from speech produced by male talkers but would also provide a further indication of their ability to generalize their representations of words across talkers of the same sex. Alternatively, if infants do not recognize the familiarized word in the test passages, then the difference in results for Experiments 1 and 2 may not be attributable to their inclusion of indexical information in representations of words but rather to their inability to segment words from speech produced by men.

Method

Participants. Twenty-four American 7.5-month-olds from monolingual English-speaking families were tested. The infants had a mean age of 33 weeks, 6 days (range = 30 weeks, 0 days to 40 weeks, 5 days). Seven additional infants were tested but not included due to crying (2), failure to complete the full set of test trials (1), failure to look for an average of at least 3 s to each stimulus type (1), parental interference (1), and equipment failure (2). All were recruited from families living in the greater Baltimore, Maryland area.

Stimuli. The stimuli consisted of repetitions of isolated words and passages by two male talkers. The male talker from Experiment 2 was used (Male Talker 1), and an additional male talker recorded the same words and passages (Male Talker 2). For Male Talker 2, the average duration of the passages was 19.37 s (ranging from 18.99 s for the *cup* passage to 19.65 s for the *dog* passage). The average duration of the lists was 18.85 s (ranging from 17.91 s for the *cup* list to 19.73 s for the *bike* list).

Apparatus and procedure. These were identical to Experiments 1 and 2.

Results and Discussion

The infants' average orientation time to each word was computed. A two-way ANOVA revealed main effects of word familiarity, $F(1, 23) = 7.30$, $p < .02$, and test block, $F(3, 69) = 20.49$, $p < .001$. The Word Familiarity × Test Block interaction was also significant, $F(3, 69) = 6.91$, $p < .001$. Figure 3 displays the average orientation time to the passages with the familiarized (M = 8.50 s; SD = 3.10 s) and the unfamiliar (M = 7.10 s; SD = 3.61 s) target words. The means differed by 1.40 s (95% CI: 0.38 < 1.40 < 2.41). Overall, 18 of the 24 infants oriented longer to the passages containing the familiarized words.

To determine whether infants who heard the two male talkers in the present experiment differed from those in Experiment 1 who heard the two female talkers, we submitted the data from the two experiments to a mixed-design ANOVA with experiment as a between-subjects variable and word familiarity as a within-subjects variable. Only the main effect of word familiarity proved to be significant, $F(1, 58) = 12.64$, $p < .001$. Neither the main effect of experiment ($F < 1.00$) nor the interaction between experiment and familiarity ($F < 1.00$) approached statistical significance. Hence, there was no indication that infants had more difficulty segmenting the words from the two male talkers in the

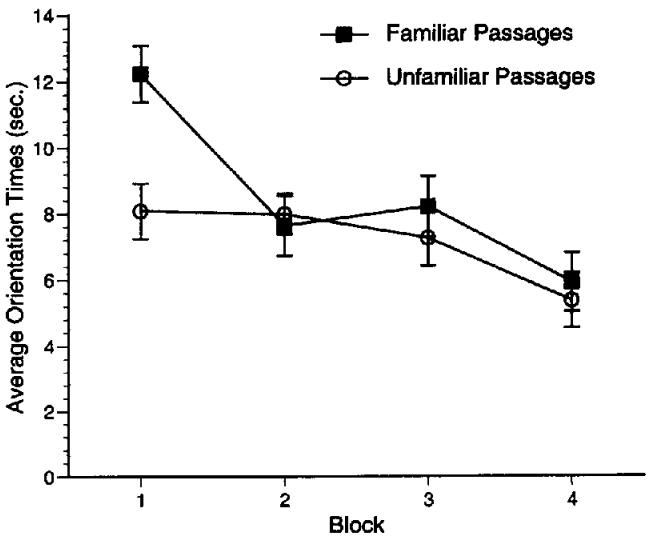

Figure 3. Displays 7.5-month-olds' average orientation times and standard error (±1 *SE*) to test passages in Experiment 3 across four blocks of trials. Passages were presented by a male talker different from the male talker used during the familiarization phase.

present experiment than they had segmenting the words from the two female talkers of Experiment 1.

The findings suggest that by 7.5 months of age, infants are able to generalize words across two male talkers. The results replicate the findings from the first experiment and undermine the notion that the infants have particular difficulty in segmenting words from speech produced by male talkers. Hence, 7.5-month-olds do display some ability to generalize their representations of words across talkers of the same sex.

Although there was evidence of a significant Word Familiarity × Test Block interaction, it did not follow the pattern expected for a perceptual adjustment to the new voice of a new talker. In particular, there was no indication of a consistent improvement in recognizing the familiarized words across blocks. Instead, the infants oriented much longer to the passages with the familiarized words in the first block (12.23 s and 8.08 s, respectively), slightly longer to the ones with the unfamiliar words in the second block (7.63 s vs. 7.98 s), and moderately longer to the ones with the familiarized words in the last two blocks (see Figure 3). Had the infants adjusted their perceptual systems to the new talker, one would have expected a much different pattern, such as a small difference in the first block and then a larger one in the following blocks. The dramatic change in orientation time differences between the first and second blocks may be an artifact. Perhaps the reduction in orientation times to the familiar passages in the second block stems from very long orientation times in the first block.

Acoustic analyses conducted on the talkers' voices of the first two experiments suggested that pitch properties, which correlated with the talker's sex, may be one of the important acoustic cues that infants use in word recognition. Infants may be able to recognize the words when the pitch of the two talkers is somewhat similar but not when the pitch is different. If so, we would expect the pitch characteristics of the male talkers used in Experiment 3 to be more similar to each other than those of the male and female talkers used in Experiment 2. To determine this, we conducted acoustic analyses for the second male talker's voice in Experiment 3, and compared the differences between the two males to those of the male and female talkers' voices used in Experiment 2. Table 1 provides a comparison of the pitch and durational characteristics for the talkers used in Experiments 2 and 3.

With respect to their mean pitch and pitch peak, the talkers in Experiment 3 were lower than those in Experiment 2, $F(1, 166) = 53.79$, $p < .001$, and $F(1, 166) = 47.02$, $p < .001$, respectively, and the experimental condition/talker interaction was marginally significant, $F(1, 166) = 3.58$, $p = .06$, for mean pitch and significant, $F(1, 166) = 5.50$, $p = .02$, for pitch peak. Once again, the differences between the talkers in the same-sex condition are smaller than the differences between the talkers in the opposite-sex condition with respect to mean pitch and pitch peak. The analyses of the durational characteristics of the target words provided no indication of greater differences between talkers of the opposite sex than between the two male talkers. Thus, the main effect of experimental condition on the average duration of the words spoken by the two talkers, $F(1, 166) = 2.10$, $p > .14$, was not significant, nor was the interaction between experimental condition and talker, $F(1, 166) = 1.89$, $p > .17$. Overall, then, the acoustic analyses comparing the opposite-sex talkers with two male talkers mirror those for the comparison of the female talkers with the opposite-sex talkers. In both instances, there was significant evidence for the greater dissimilarity of the opposite-sex talkers in the mean pitch and pitch peak measures but not for the durational measure. In addition, judgments gathered on the perceived similarity of the two male talkers (Houston, 1999) indicated that the mean squared distance between the two male talkers in the present experiment (2.55) was less than that for the male and female talkers in Experiment 2 (6.64).

Taken together, the results of the first three experiments reveal that 7.5-month-olds can recognize instances of the same words produced by different talkers, under certain circumstances. In particular, when differences between talkers' voices are not great, such as when two talkers are of the same sex and speak a similar dialect, 7.5-month-olds display some ability to generalize their representations of words spoken by one talker to those of another talker. However, when the differences between talkers' voices are greater, such as when they are of the opposite sex, 7.5-month-olds do not display recognition of the same word produced by different talkers. To verify this apparent effect of talker similarity in 7.5-month-olds' ability to recognize words across different talkers, we submitted the average familiar and unfamiliar passage orientation times from Experiments 1–3 to an ANOVA of a 3 (experiment) × 2 (word familiarity) mixed design. The main effect of word familiarity was significant, $F(1, 93) = 11.02$, $p < .01$, but the main effect of experiment was not, $F(2, 93) < 1.00$. Moreover, there was a significant interaction between these two factors, $F(2, 93) = 3.16$, $p < .05$. Paired comparisons revealed a marginally significant experiment/word familiarity interaction comparing Experiments 1 (female talkers) and 2 (female/male talkers), $F(1, 70) = 2.81$, $p = .097$, a significant interaction comparing Experiments 2 (female/male talkers) and 3 (male talkers), $F(1, 58) = 5.88$, $p < .02$, and no interaction comparing Experiments 1 (female talkers) and 3 (male talkers), $F(1, 58) < 1.00$.

In sum, the findings suggest that the likelihood that 7.5-month-olds will recognize a word in fluent speech is not based solely on the abstract phonetic properties of the word. Rather, the similarities among the indexical properties of words also appear to matter. Consequently, at this early phase in the development of word segmentation abilities, instances of the same word spoken by different talkers may not always be categorized by 7.5-month-olds as the same word, especially if the talkers' voices differ substantially in acoustic characteristics, such as pitch. Of course, eventually language learners must be able to generalize words across a wide variety of talkers. To gain some indication of when this ability may develop, we decided to conduct a version of Experiment 2 with an older group of infants.

Experiment 4

The second half of the first year of life marks a period of important changes in infants' speech-perception capacities. For example, between 6 and 10 months, infants' ability to discriminate non-native speech contrasts begins to decline for many distinctions (Best, 1995; Lalonde & Werker, 1995; Tsushima et al., 1994; Werker & Lalonde, 1988; Werker & Tees, 1984) but not all (Best, McRoberts, & Sithole, 1988; Polka & Bohne, 1996). During the same period, infants are becoming more attuned to certain characteristics of native language sound patterns such as their typical prosodic (Jusczyk, Cutler, & Redanz, 1993; Morgan & Saffran,

1995; Turk, Jusczyk, & Gerken, 1995) and phonotactic properties (Friederici & Wessels, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk, Luce, & Charles-Luce, 1994). It is during this period of changing sensitivity to native and nonnative language input that the beginnings of word-segmentation abilities have been first reported (e.g., Jusczyk & Aslin, 1995; Saffran et al., 1996). However, it also appears that infants' abilities to segment words from fluent speech contexts undergo considerable development between 7.5 and 10.5 months of age. For example, Jusczyk, Houston, et al. (1999) found that at 7.5 months, English-learners do not segment words with weak–strong stress patterns from fluent speech, although they do segment ones with strong–weak stress patterns. However, by 10.5 months, English-learners are able to segment words with weak–strong patterns. Similarly, Jusczyk, Hohne, et al. (1999) found no indication that 9-month-old English-learners are sensitive to allophonic markers of word boundaries but that 10.5-month-olds are sensitive to such cues. In light of the changes taking place in the development of word segmentation abilities at around 10.5 months, it seemed reasonable to ask whether infants at this age might also generalize words produced by a talker to those produced by a talker of the opposite sex. To explore this possibility, we tested 10.5-month-olds with the same stimuli that the 7.5-month-olds heard in Experiment 2.

Method

Participants. Thirty-two American 10.5-month-olds from monolingual English-speaking homes were tested. The infants had a mean age of 45 weeks, 6 days (range = 43 weeks, 3 days to 49 weeks, 0 days). Fifteen additional infants were tested but not included due to crying (4), failure to complete the full set of test trials (5), equipment failure (1), and experimenter error (5). All were recruited from families living in the greater Baltimore, Maryland area.

Stimuli, apparatus, and procedure. These were identical to Experiment 2.

Results and Discussion

The infants' average orientation time to each word was computed. An ANOVA of these data revealed significant main effects of word familiarity, $F(1, 31) = 6.45$, $p < .02$, and test block, $F(3, 93) = 17.44$, $p < .001$. However, the Word Familiarity × Test Block interaction was not significant, $F(3, 93) < 1.00$. Figure 4 displays the average orientation times to the passages with the familiarized ($M = 7.22$ s; $SD = 2.45$ s) and the unfamiliar ($M = 6.25$ s; $SD = 2.38$ s) target words. The means differed by 0.97 s (95% CI: $0.20 < 0.97 < 1.74$). Overall, 21 of the 32 infants oriented longer to the passages containing the familiarized words.

To evaluate the differences in abilities between 7.5- and 10.5-month-olds to generalize words across talkers of the opposite sex, we submitted the data from Experiments 2 and 4 to an ANOVA of a 2 (age) × 2 (word familiarity) mixed design. There was a marginally significant main effect of age, $F(1, 66) = 3.23$, $p = 0.08$, reflecting shorter overall orientation times of the older infants. Also, there was a significant main effect of word familiarity, $F(1, 66) = 4.29$, $p < .05$. Most importantly, there was a significant interaction between age and word familiarity, $F(1, 66) = 4.27$, $p < .05$.

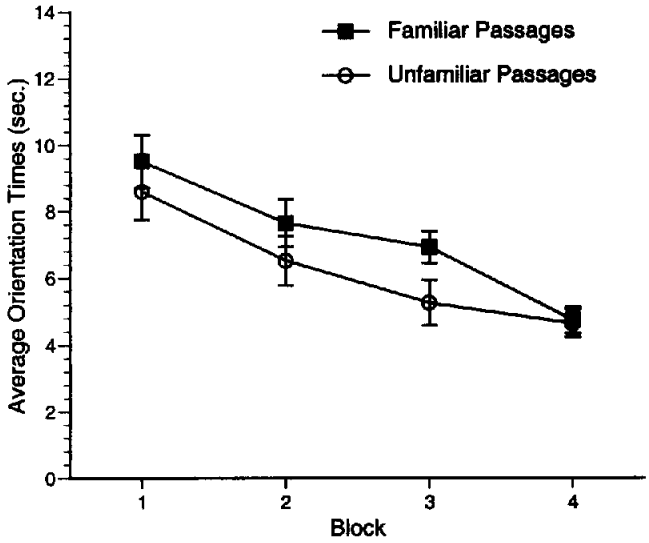

Figure 4. Displays 10.5-month-olds' average orientation times and standard error (±1 *SE*) to test passages in Experiment 4 across four blocks of trials. Passages were presented by either a male or female talker—the opposite-sex talker than whichever was used during the familiarization phase.

The pattern of results suggest that, in contrast to 7.5-month-olds, 10.5-month-olds familiarized with words produced by one talker, subsequently recognized these words in passages produced by a talker of the opposite sex. Clearly, the ability to generalize words across different talkers improves between 7.5- and 10.5-months of age. This improvement in the ability to generalize across the productions of the same words by different talkers is another manifestation of the developments occurring in infants' word-segmentation abilities during this short period. English learners move from a heavy reliance on the location of stressed syllables to identify word onsets in fluent speech to the use of multiple sources of information in word segmentation. The latter change allows infants to segment words with a wider range of stress patterns from fluent speech. Similarly, the improvements documented here, in 10.5-month-olds' abilities to generalize across productions of the same words by different talkers, allow them to recognize particular words in a wider range of contexts. In this sense, infants are becoming more adult-like in their word-segmentation abilities.

General Discussion

The findings reported here suggest that talker-specific speech characteristics can affect infants' recognition of words in fluent speech. Although English-learning 7.5-month-olds displayed some capacity to generalize words across two female talkers and two male talkers, they did not generalize across productions of the same words produced by a female and a male talker. Acoustic analyses revealed that the voices in the same-sex talker conditions were more similar to each other than those in the opposite-sex talker condition with respect to their pitch characteristics. By 10.5 months, English-learners who were familiarized with words from one talker did recognize these words in fluent speech contexts, when spoken by a talker of the opposite sex.

The findings are at odds with a biological view that assumes prewired neural mechanisms normalize speech with respect to talker-specific information (e.g., Sussman, 1984, 1986). Such an approach predicts that talker-specific information is stripped from infants' representations. Instead, the present findings indicate that talker specific characteristics of speech do affect infants' recognition of words and that the ability to cope with talker variability develops as language is acquired.

The overall pattern of findings across the present series of experiments raises several questions that merit further discussion. First, why do talker-voice differences pose any problem in 7.5-month-olds' ability to recognize words in fluent speech contexts? Second, what is responsible for the improvement in the ability of 10.5-month-olds to tolerate talker voice differences in recognizing words in fluent speech? Third, how do the present findings bear on the issue of whether infants' representations of the sound structure of particular words are abstract phonetic descriptions or are ones that include indexical information relating to talker voice characteristics? In the following discussion, we attempt to address each of these questions.

As noted earlier, several previous studies have demonstrated that infants younger than 7.5 months are able to ignore talker differences in discriminating certain phonetic contrasts (e.g., Jusczyk et al., 1992; Kuhl, 1979, 1983). Given this apparent capacity for dealing with the acoustic variability in different talkers' pronunciations of the same syllables, why were the 7.5-month-olds in Experiment 2 who were familiarized with words produced by one talker unsuccessful in recognizing the same words in the passages produced by a talker of the opposite sex? One potential factor has to do with a change in the nature of the task demands between the earlier investigation and the present one. In particular, in the earlier studies in which infants were able to ignore talker voice differences, the discrimination task involved an immediate shift from one phonetic stimulus to another. The infants in this setting simply had to detect that a phonetic change had occurred. By comparison, in the present study, infants had to retain information about two different words produced in citation form, then subsequently recognize the occurrence of fluent speech versions of these items in the midst of a number of surrounding words. In the face of these additional processing demands, the infants' ability to ignore differences between talkers with dissimilar voices may have broken down. Some empirical support for this view comes from two previous investigations. First, when Jusczyk et al. (1992) required 2-month-olds to retain information over a short delay period, they found that talker variability disrupted their ability to detect a phonetic change between their test stimuli. Second, Stager and Werker (1997) found that 14-month-olds who were able to discriminate a phonetic contrast in a simple discrimination experiment, failed to discriminate the same contrast in a word-learning task. They attributed the failure in the latter task to the greater demands it imposed on infants' language-processing capacities.

Let us now address the second question. What promotes the change in the ability of 10.5-month-olds to recognize productions of the same word by talkers of the opposite sex? One possibility is that a perceptual normalization process that strips away talker-specific information underlies infants' ability to ignore talker differences and that the greater experience and practice of older infants allows them to execute this process more rapidly and efficiently. Such an account would certainly be consistent with influential views of perceptual learning (e.g., Gibson, 1969). However, an alternative explanation is that indexical information (including talker-voice characteristics) is retained in infants' representations of the sound structure of words and that the developmental improvement results because of a shift in the weighting of the information stored in infants' lexical representations. For instance, the WRAPSA Model proposes that infants learn to weigh more heavily those properties of speech that are most relevant for making meaningful distinctions among words in the native language (Jusczyk, 1993, 1997). This would mean that although infants retain information relating to talker-specific characteristics, they tend to give greater weight to those characteristics that are most relevant to making phonemic distinctions (see also Nosofsky, 1988). Thus, although infants at this age appear to be in the process of developing phonemic categories (e.g., Werker, 1991; Werker & Tees, 1984), there is also some indication that they still retain information about the specific voice characteristics of different talkers. In addition to the present findings, Jusczyk, Hohne, Jusczyk, and Redanz (1993) reported that 8.5-month-olds demonstrated recognition of an unfamiliar female talker's voice when subsequently tested on this talker and another female talker two weeks later. Although the present data do not rule out the first possible explanation (i.e., that 10.5-month-olds omit talker-specific information), more recent research in our laboratory suggests that talker-specific information still plays a role in word recognition by infants at this age. Specifically, Houston and Jusczyk (2000) found that although infants recognized familiarized words after a 1-day delay when the words and passages were produced by the same talker, both 7.5- and 10.5-month-olds failed to do so when familiarization words and test passages were produced by different talkers.

Turning to the third question, what do the present findings have to say about the nature of infants' representations of the sound structure of words? The finding that certain talker voice differences affect 7.5-month-olds' ability to recognize the familiarized words in fluent speech contexts is certainly consistent with the view that indexical information is encoded in their representations of words. If such information is not part of their representations of the sound structure of words, then we need another explanation of why 7.5-month-olds have difficulty generalizing across talkers of the opposite sex.

Even if we accept the possibility that indexical information is included in 7.5-month-olds' representations of words, other questions arise. For instance, in the present study, the type of talker differences that affected 7.5-month-olds' recognition of words was one that involved talkers of the opposite sex. This raises the possibility that the kind of indexical information encoded into infants' representations of words relates not to the identity of a particular talker, but to the particular sex of the talker. In other words, infants at this age may simply have separate representations of words produced by men and women. Consistent with this view are findings from a study with adults by Geiselman and Bellezza (1977) who reported that it is the sex of the talker that is incidentally retained in processing speech information. However, a more recent investigation with adults, Palmeri et al. (1993) used a greater range of talkers (up to 20 different talkers in some of their test conditions). Palmeri et al. found that same-talker repetitions of words were better recognized than different-talker repetitions of the same words and that there was no evidence of any significant

interaction with talker sex on these effects. In other words, the magnitude of the difference between same- and different-talker repetitions was not greater for two talkers of the opposite sex than for two of the same sex. These findings seem to show that it is talker-specific information and not just information about the sex of talkers that adults retain in their representations. The plausibility of the view that infants retain sex-specific but not talker-specific information is also strained by certain findings with infants. Recall also that Jusczyk, Hohne, et al.'s (1993) found that 8.5-month-olds retained information about a particular female talker's voice versus that of another female talker's voice. Other evidence shows that newborn infants recognize their own mother's voice from that of another female (DeCasper & Fifer, 1980). Such findings appear to demand a more detailed encoding of talker-specific characteristics than is provided by the view that only sex-specific information is included in 7.5-month-olds' representations of words.

Although we have argued that the present findings are consistent with views that representations of words include information about talker-specific characteristics (e.g., Goldinger, 1996; Jusczyk, 1997; Remez et al., 1997), we readily acknowledge that more evidence is required to establish the validity of our claim. We are currently exploring whether with a 1-day delay period following familiarization, infants are better at recognizing words produced by the same talker or a different talker (Houston & Jusczyk, 2000). Similarly, a more systematic scaling of talker voice differences will provide a more definitive account of how acoustic dissimilarities between talkers affect 7.5-month-olds' ability to recognize words in fluent speech contexts (Houston, 1999). In the meantime, the present findings provide new information about infants' developing capacities for word segmentation. They extend previous research by demonstrating that even at the very beginning phase of segmenting words from fluent speech, 7.5-month-olds have some capacity to generalize across different talkers of the same sex. Three months later, this ability has improved to the point at which they are able to generalize more readily across talkers of the opposite sex.

References

Best, C. T. (1995). Learning to perceive the sound patterns of English. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (Vol. 9, pp. 217–304). Norwood, NJ: Ablex.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of the perceptual re-organization for speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345–360.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America, 67,* 648–662.

Carrell, T. D., Smith, L. B., & Pisoni, D. B. (1981). Some perceptual dependencies in speeded classification of vowel color and pitch. *Perception & Psychophysics, 29,* 1–10.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 521–533.

Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65–110). New York: Academic Press.

Clark, E. V. (1983). Meanings and concepts. In J. H. Flavell & E. M. Markman (Eds.), *Cognitive development* (Vol. III, pp. 787–840). New York: Wiley.

Craik, F. I. M., & Kirsner, K. (1974). The effect of speaker's voice on word recognition. *Quarterly Journal of Experimental Psychology, 26,* 274–284.

DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science, 208,* 1174–1176.

Fant, C. G. M. (1960). *Acoustic theory of speech production.* The Hague: Mouton.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception & Psychophysics, 54,* 287–295.

Geiselman, R., & Bellezza, F. (1977). Incidental retention of speaker's voice. *Memory & Cognition, 5*(6), 658–665.

Gerstman, L. (1968). Classification of self-normalized vowels. *IEEE Transactions on Audio and Electroacoustics, ACC-16,* 78–80.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105,* 251–279.

Goldinger, S. D., Pisoni, D. B., & Luce, P. A. (1996). Speech perception and spoken word recognition: Research and Theory. In N. J. Lass (Ed.), *Principles of experimental phonetics* (pp. 277–327). St. Louis, MO: Mosby.

Houston, D. M. (1999). *The role of talker variability in infant word representations.* Unpublished doctoral dissertation, Johns Hopkins University.

Houston, D. M., & Jusczyk, P. W. (2000). *Infants' memory for words and voices over a 1-day delay.* Manuscript in preparation.

Jusczyk, P. W. (1993). From language-general to language-specific capacities: The WRAPSA model of how speech perception develops. *Journal of Phonetics, 21,* 3–28.

Jusczyk, P. W. (1997). *The discovery of spoken language.* Cambridge, MA: MIT Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology, 29,* 1–23.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development, 64,* 675–687.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32,* 402–420.

Jusczyk, P. W., Hohne, E. A., & Bauman, A. (1999). Infants' sensitivity to allophonic cues for word segmentation. *Perception & Psychophysics, 61,* 1465–1476.

Jusczyk, P. W., Hohne, E. A., Jusczyk, A. M., & Redanz, N. J. (1993). Do infants remember voices? *Journal of the Acoustical Society of America, 93,* 2373.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39,* 159–207.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33,* 630–645.

Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month old infants. *Cognition, 43,* 253–291.

Kemler Nelson, D. G., Jusczyk, P. W., Mandel, D. R., Myers, J., Turk, A., & Gerken, L. A. (1995). The Headturn Preference Procedure for testing auditory perception. *Infant Behavior & Development, 18,* 111–116.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279–312.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.

Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America, 66,* 1668–1679.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development, 6,* 263–285.

Lalonde, C. E., & Werker, J. F. (1995). Cognitive influences on cross-language speech perception in infancy. *Infant Behavior and Development, 18,* 459–475.

Markman, E. M. (1991). The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on language and thought* (pp. 72–106). Cambridge, UK: Cambridge University Press.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. *Child Development, 66*(4), 911–936.

Mullennix, J. W., Johnson, K. A., Topcu-Durgun, M., & Farnsworth, L. M. (1995). The perceptual representation of voice gender. *The Journal of the Acoustical Society of America, 98,* 3080–3095.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85,* 365–378.

Nearey, T. M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America, 85,* 2088–2113.

Nosofsky, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 54–65.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker contingent process. *Psychological Science, 5,* 42–46.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 309–328.

Polka, L., & Bohn, O. S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America, 100,* 577–592.

Quine, W. V. O. (1960). *Word and object.* Cambridge, MA: MIT Press.

Remez, R. E., Fellowes, J. M., & Rubin, P. E. (1997). Talker identification based on phonetic information. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 651–666.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996, December). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 915–930.

Schacter, D. L., Church, B., & Bolton, E. (1995). Implicit memory in amnesic patients: Impairment of voice-specific priming. *Psychological Science, 6,* 20–25.

Shankweiler, D., Strange, W., & Verbrugge, R. (1977). Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 315–345). Hillsdale, NJ: Erlbaum.

Sheffert, S. M., & Fowler, C. A. (1995). The effects of voice and visible speaker change on memory for spoken words. *Journal of Memory and Language, 34,* 665–685.

Sommers, M. S., Nygaard, L. C., & Pisoni, D. B. (1994). Stimulus variability and spoken word recognition. I. Effects of variability in speaking rate and overall amplitude. *Journal of the Acoustical Society of America, 96,* 1314–1324.

SPSS (Version 8.0) [Computer software]. (1998). Chicago: SPSS.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature, 388,* 381–382.

Stevens, K. N. (1972). The quantal nature of speech. In J. E. E. David & P. B. Denes (Eds.), *Human communication: A unified view* (pp. 51–66). New York: McGraw-Hill.

Stevens, K. N., & Blumstein, S. E. (1981). The search for invariant acoustic correlates for phonetic features. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 1–38). Hillsdale, NJ: Erlbaum.

Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America, 85,* 2135–2153.

Suomi, K. (1984). On talker and phoneme information conveyed by vowels: A whole spectrum approach to the normalization problem. *Speech Communication, 3,* 199–209.

Sussman, H. M. (1984). A neuronal model for syllable representation. *Brain and Language, 22,* 167–177.

Sussman, H. M. (1986). A neuronal model of vowel normalization and representation. *Brain and Language, 28,* 12–23.

Syrdal, A. K., & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America, 79,* 1086–1100.

Tincoff, R., & Jusczyk, P. W. (1996, July). *Are word-final sounds perceptually salient for infants?* Poster presented at the Fifth Conference on Laboratory Phonology, Evanston, IL.

Tomiak, G. R., Green, K. P., & Kuhl, P. K. (1991). Phonetic coding and its relationship to talker and rate normalization. *Journal of the Acoustical Society of America, 90,* 2363.

Tsushima, T., Takizawa, O., Sasaki, M., Siraki, S., Nishi, K., Kohno, M., Menyuk, P., & Best, C. (1994, October). *Discrimination of English /r-l/ and /w-y/ by Japanese infants at 6–12 months: Language specific developmental changes in speech perception abilities.* Paper presented at the International Conference on Spoken Language Processing, Yokohama, Japan.

Turk, A. E., Jusczyk, P. W., & Gerken, L. A. (1995). Do English-learning infants use syllable weight to determine stress? *Language and Speech, 38,* 143–158.

van de Weijer, J. (1998). *Language input for word discovery* (Vol. 9). Nijmegen, the Netherlands: Max Planck Institute.

Walden, B. E., Montgomery, A. E., Gibeily, G. J., Prosek, R. A., & Schwartz, D. M. (1978). Correlates of psychological dimensions in talker similarity. *Journal of Speech and Hearing Research, 21,* 265–275.

Wannemacher, C. (1995). *The relationship between processing of talker voice and phonetic information in speech perception.* Unpublished doctoral dissertation, State University of New York at Buffalo, Buffalo, NY.

Werker, J. F. (1991). The ontogeny of speech perception. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 91–109). Hillsdale, NJ: Erlbaum.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24,* 672–683.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7,* 49–63.

Woodward, J. Z., & Aslin, R. N. (1990, April). *Segmentation cues in maternal speech to infants.* Paper presented at the 7th biennial meeting of the International Conference on Infant Studies, Montreal, Quebec, Canada.

(Appendix follows)

Appendix

Passage Recorded as Stimuli in Test Phases

His bike had big black wheels. The girl rode her big bike. Her bike could go very fast. The bell on the bike was really loud. The boy had a new red bike. Your bike always stays in the garage.

The dog ran around the yard. The mailman called to the big dog. He patted his dog on the head. The happy red dog was very friendly. Her dog barked only at squirrels. The neighborhood kids played with your dog.

The cup was bright and shiny. A clown drank from the red cup. The other one picked up the big cup. His cup was filled with milk. Meg put her cup back on the table. Some milk from your cup spilled on the rug.

The feet were all different sizes. This girl has very big feet. Even the toes on her feet are large. The shoes gave the man red feet. His feet get sore from standing all day. The doctor wants your feet to be clean.

Received May 11, 1999
Revision received November 3, 1999
Accepted December 13, 1999 ■