Children's Recognition of Cartoon Voices

Melanie J. Spence
Pamela R. Rollins
Susan Jerger
The University of Texas at Dallas

We examined developmental changes in talker recognition skills by assessing 3-, 4-, and 5-year-old children's recognition of 20 cartoon characters' voices. For each participant, the character set was subdivided into more and less familiar talkers based on the participant's ability to name each character. Four- and 5-year-old children recognized more of the voices (81% and 86%, respectively) than did 3-year-olds (61%), although performance of all age groups was well above chance. All groups of children were more accurate at recognizing more familiar than less familiar characters. These results suggest that indexical information about a talker becomes an integral part of the perceptual record in memory and can be used by children at a very young age. These results are important because children's ability to learn vocal sources may be an important aid to the development of spoken word recognition.

KEY WORDS: talker recognition, implicit memory, speech perception, incidental learning, children

Recent studies demonstrate that talker attributes influence the processing of speech in adult listeners (Nygaard & Pisoni, 1998). Talker attributes refer to the paralinguistic or indexical information of a speech signal, such as a talker's identity, age, gender, social class, and emotional state (Abercrombie, 1967). Currently, there has been a resurgence of interest in the talker recognition skills of adult listeners. The impact of talker-specific acoustic information on the processing of speech is relevant for theoretical issues in the study of speech perception, such as the nature of the representations that are formed (e.g., abstract prototypes vs. instance-specific exemplars), the type of processes implicated for linking a talker's vocal signature to the identity of the talker, and the types of processes and/or systems implicated for the processing of speech versus nonspeech information. Historically, speech perception research has also included a developmental approach that has provided insight into the types of processes involved in speech perception (Eimas & Miller, 1992; Miller & Eimas, 1983) and the role of linguistic experience in speech perception development (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Werker, 1995). Few studies to date, however, have examined children's talker recognition skills or how children's representations of voice attributes are related to semantic knowledge about talker identity. As in the adult literature, data on children's voice recognition may be important for understanding the nature and development of children's linguistic representations. This study examined developmental differences in the talker recognition skills of 3-, 4-, and 5-year-old children.

Adults' Talker-Specific Speech Perception

As noted above, much recent work has examined adults' talker recognition skills and the impact of voice-specific information on the processing of and memory for speech. There is evidence that adults encode and maintain information about characteristics of voices and that this talker-specific knowledge facilitates their processing of linguistic information (Pisoni, 1993). For example, adult listeners identify words and sentences in noise more accurately when they listen to familiar, as opposed to unfamiliar, talkers (Nygaard & Pisoni, 1998). Adults also are more likely to correctly judge a word as "old" in a recognition task if the word is spoken by the same talker at familiarization and test, rather than by different talkers (Palmeri, Goldinger, & Pisoni, 1993). Implicit encoding of indexical characteristics has been evidenced by findings that priming of spoken utterances occurs when perceptual properties, such as affective tone, fundamental frequency, and voice, are unchanged from study to test (Church & Schacter, 1994).

Data relevant for understanding the nature of linguistic and nonlinguistic representations, as well as the relations between them, have been provided by studies examining the effects of talker variability on speech perception. For example, naming latencies are slower and memory is less accurate for words from multiple-talker, relative to single-talker, lists (Mullennix, Pisoni, & Martin, 1989; Martin, Mullennix, Pisoni, & Summers, 1989). Other work using a Garner (1974) selective attention task (speeded classification task) has also provided evidence for processing dependencies between linguistic and talker-specific information (Jerger et al., 1993; Mullennix & Pisoni, 1990). In a Garner task, listeners are required to classify the attributes of utterances along one of two dimensions, linguistic or talker gender, while ignoring irrelevant variation in the other dimension. Listeners have difficulty ignoring irrelevant variation in either dimension, indicating interference between the talker-specific information and linguistic information. However, there is an asymmetric effect such that variability in the talker dimension impairs classification of linguistic information more than vice versa. These results are consistent with the idea that indexical and linguistic information are obligatorily co-processed.

These results have led several speech researchers to conclude that representations of both linguistic and nonlinguistic acoustic attributes are stored in memory and that these representations are detailed and exemplar-specific. For example, based on a number of studies such as those described above, Nygaard, Sommers, and Pisoni (1994) have suggested that the "mechanisms responsible for encoding talker information would seem to be linked directly to those that underlie phonetic perception" (p. 43). Tulving and Schacter (1990) have proposed a similar idea: that detailed perceptual information about talkers' voices is implicitly encoded/stored in some type of perceptual representation system (PRS). The notion that perceptual information is implicitly encoded is supported by their work showing that the perceptual attributes of spoken utterances prime later identification, but not explicit recognition, of words (Schacter, Church, & Treadwell, 1994). These views contrast with the traditional abstractionist approach to speech perception, which maintains that linguistic representations are prototypical and abstract in nature (Ladefoged & Broadbent, 1957; Studdert-Kennedy, 1974). Traditionally, variability in acoustic cues due to talker-specific differences is considered to be irrelevant and in need of normalization so that the linguistic input can be matched to prototypical linguistic forms (Shankweiler, Strange, & Verbrugge, 1977; Summerfield & Haggard, 1973).

Another area of research that contributes to our understanding of multidimensional speech processing is the study of talker recognition—namely, what types of attributes are encoded and essential for recognizing a talker and how the attributes are linked to semantic knowledge about the identity of the talker. Adult research using small numbers of familiar talkers (e.g., maximum of 11) producing sentence-length voice samples has found talker recognition rates above 97% (Abberton & Fourcin, 1978; Bricker & Pruzansky, 1966; Hollien, Majewski, & Doherty, 1982; Ladefoged, 1978, and LaRiviere, 1972 [both as cited in Van Lancker, Kreiman, & Emmorey, 1985a]). Recognition accuracy was influenced by stimulus set size (Pollack, Pickett, & Sumby, 1954), the duration of the voice sample (Compton, 1963; Cook & Wilding, 1997; Ladefoged & Ladefoged, 1980 [as cited in Van Lancker et al., 1985b]; Schweinberger, Herholz, & Sommer, 1997), spectral bandwidth (Dukiewicz, 1970; Pollack et al., 1954), and familiarity with the talkers (Papcun, Kreiman, & Davis, 1989).

Talker-recognition research using larger sets of talkers has found lower, but well above chance, accuracy rates (Van Lancker, Kreiman, & Emmorey, 1985a; Van Lancker, Kreiman, & Wickens, 1985b). Adults correctly identified 70% of 45 famous male voices (2-s samples of voices such as Cary Grant and Richard Nixon) in a task in which six names were presented as response choices during presentation of each voice. When the voice samples were presented without response choices, adults correctly identified only 26.6% of the voices. Recognition of famous voices improved as the duration of the voice samples increased from .25 s to 1 s (Schweinberger et al., 1997).

Children's Processing of Talker Attributes and Perceptual Attributes of Words

Few studies have examined whether children also encode talker-specific or indexical information, if children's voice recognition abilities are influenced by

the same factors as those that affect adult voice recognition, and if there are developmental differences in these skills. Studies of grade-school children have shown age-related improvements between the ages of 6 and 10 in recognition of recently learned voices (Mann, Diamond, & Carey, 1979). Other research has demonstrated that 6- to 9-year-old children's gender classification of adult talkers was equivalent to the performance of adult listeners (Bennett & Montero-Diaz, 1982) and that fourth-grade children's recognition of classmates' voices was affected by utterance length, as is the case for adults (Murry & Cort, 1971).

Research examining preschool children's talker encoding and recognition skills has provided some evidence that they encode talker-specific information. The effect of talker variability on young children's word recognition was examined by Ryalls and Pisoni (1997). They found that 3- to 5-year-old children were more accurate on word-recognition tasks when the words were produced by a single talker rather than multiple talkers. This effect was found when words were presented both with and without background noise. This research also demonstrated that talker variability disrupted performance more for the younger than the older groups of children. These results reveal that children, like adults, process talker characteristics of spoken utterances. These findings also are consistent with the suggestion that younger children may focus more on perceptual attributes than older children and adults (Bach & Underwood, 1970). Although infants' processing of talker variability has been examined (Houston, Jusczyk, & Tager, 1998; Jusczyk, Pisoni, & Mullennix, 1992), the work by Ryalls and Pisoni (1997) is the only published report, to the best of our knowledge, examining the impact of talker variability on young children's speech processing.

Recognition of familiar talkers by 4- and 5-year-old children (age ranging from 4 years 2 months to 5 years 5 months) was investigated by Bartholomeus (1973). Children were presented 18 (3-s) voice samples of their classmates and, in different tasks, asked to name the talker when listening to each voice sample, to match each voice with a picture of the classmate, and to name the talker when listening to voice samples played backward. Children correctly identified 62% and 55%, respectively, of 18 voices by matching the picture of the classmate with the voice sample, and they correctly provided classmates' names for 59% and 54% of the voice samples. As has been found for adults (Van Lancker et al., 1985a), voice identification in the backward voice-naming condition was less accurate (41.6%) than in the forward voice-naming condition (56.5%). Voice-recognition performance was also extremely variable across children in the study, with accurate voice-picture matching ranging from 11% to 100%, and voice naming without pictures ranging from 12% to 95%. Although a few children in the sample received near-perfect scores, the average performance suggests that children perform at lower levels than has been found for adults. However, the tremendous variability in children's performance was reported in terms of ranges, so it is difficult to determine the distributional patterns of scores and the representativeness of the means reported in terms of the children's response rates. Additionally, the use of children's voices as stimuli may have affected the performance values obtained in this study. Children were tape-recorded repeating a short adult-like utterance. ("Hi. I go to your school. Do you know who I am?") In our experience, children in this age range have been observed to alter their voices when speaking into microphones—for example, typically speaking unusually softly, which may drastically reduce talker recognition (Pollack et al., 1954). If the recorded samples of some children's voices were not representative of their typical speaking voices, then this could have also contributed to the variability and/or relatively low performance observed. A third factor that might have affected the findings of this study is variation in the amounts of interaction and voice experience that children may have had with different classmates. Although the study was conducted late in the school year so that children would have had lengthy exposure to their classmates, it is quite possible that there was much variability in the amount of experience or types of experience that children had with different classmates' voices. Some 4-year-old children, for example, are much less verbal and/or socially skilled than others so that these factors could potentially affect their vocal production and ultimately their classmates' ability to encode and subsequently recognize acoustic attributes of their voices in a talker-recognition task.

Data relevant for understanding children's encoding and recognition of vocal attributes is provided by demonstrations that young children more easily encode perceptual stimulus dimensions, relative to more abstract stimulus dimensions, on tasks involving word perception and categorization (Bach & Underwood, 1970; Felzen & Anisfeld, 1970; Scott, Serchuk, & Mundy, 1982). For example, in one paradigm used previously, children were provided sets of target words and later asked to select those target words from a list including both target and distractor words. Preschool-age children, unlike older children and adults, were more likely to confuse acoustically related words (e.g., *fog* with the target word *dog*), rather than conceptually related words (e.g., confusing *cat* with *dog*) (Bach & Underwood, 1970; Felzen & Anisfeld, 1970; Scott et al., 1982).

Additional evidence that young children have difficulty ignoring perceptual dimensions of linguistic stimuli has been reported by Jerger and her colleagues (Jerger et al., 1993; Jerger, Pearson, & Spence, 1999). Investigations of children's ability to ignore irrelevant variation in a nontarget dimension while classifying stimuli along a different target dimension have revealed that young children show the same asymmetric pattern of Garner interference between the indexical and linguistic dimensions of speech that has been found for adults. Three- to 6-year-old children had greater difficulty ignoring irrelevant variation in voice (e.g., talker gender) when classifying target words than when they were required to ignore words and classify talker gender (Jerger et al., 1993). Interestingly, when the target and nontarget dimensions are both perceptual in nature, rather than consisting of one linguistic and one perceptual dimension, young children are just as effective as adults at resisting interference from the irrelevant nontarget dimension (Jerger et al., 1999). Overall, this research demonstrates that young children encode and have difficulty ignoring acoustic attributes of words. These findings also imply that young children may readily encode indexical or perceptual attributes of voices and that their encoding of such talker attributes during their daily experience may occur via an implicit process. Results demonstrating that newborn infants prefer to listen to the mother's voice over an unfamiliar female voice also imply that an implicit learning process is involved in talker recognition (DeCasper & Fifer, 1980; Spence & DeCasper, 1987; Spence & Freeman, 1996).

Rationale for the Current Study

This study assessed talker recognition by 3-, 4-, and 5-year-olds in order to examine developmental changes in talker recognition skills across the early childhood period. If encoding of vocal attributes facilitates children's processing of and memory for speech as it does for adults, then knowledge concerning children's encoding of indexical and vocal properties of speech is essential for understanding their speech processing. The adult literature has revealed that specific vocal properties are encoded via implicit learning processes and are subject to repetition priming. The literature examining memory development during the preschool period has also shown that children encode certain types of information via implicit learning, and priming effects have been demonstrated as early as age 2. For example, no age differences in priming tasks assessing memory for pictures was found for 3- and 5-year-olds (Greenbaum & Graf, 1989) or among 4-, 5-, and 10-year-old children (Hayes & Hennessy, 1996; although see Parkin, 1993). Few studies have examined auditory priming effects, but there is some evidence of priming effects for word stimuli in 6- and 8-year-olds (Naito, 1990) and 2- to 3-year-olds (Church & Fisher, 1998). Children at 2, 2.5, and 3 years old more accurately identified low-pass filtered words that they had heard previously than ones they had not heard. Additionally, there was no difference in this priming effect between children and adult listeners. Given that children appear to implicitly encode at least some of the indexical properties of spoken words, we hypothesized that young children encode talker-specific attributes during the course of their routine exposure to voices and that children's talker recognition skills should reflect this implicit learning of such talker-specific attributes. However, because the only published report of preschool children's voice recognition (Bartholomeus, 1973) tested children varying in age by only 15 months, we do not have adequate data concerning whether encoding and recognition of vocal information changes during the preschool years. As previously noted, because adults are more accurate in recognizing more familiar than less familiar talkers, we also predicted that children would be more likely to correctly identify the voices of more familiar than less familiar talkers.

One set of talkers to which many young children are likely to have had frequent exposure is that of cartoon characters. Selection of this set of stimuli also allowed us to perform an experiment with children that was analogous to studies examining adults' recognition of famous voices (Van Lancker et al., 1985a; Van Lancker et al., 1985b). Additionally, the use of this stimulus set avoided problems one might encounter when recording children's voices and provided a stimulus set that was relatively more heterogeneous in perceptual attributes than a set of children's voices.

The task that was required of children was to identify talkers in a voice-picture matching paradigm. This task can be conceptualized within the functional face-recognition model proposed by Bruce and Young (1986). The presentation of a picture of a person activates the mental representation corresponding to the person's face, which also activates other attributes of that person, including the name of the person, the person's voice signature, information about his/her activities, social characteristics, and the like. For our task, children were shown six pictures of cartoon characters while hearing a sample of one voice, and they were asked to point to or identify the character who was talking. If the child recognizes the utterance as belonging to a particular character, the child should correctly identify the character by pointing to the picture of that particular character. If children have not implicitly stored vocal information about these talkers, then they should be equally likely to identify any of the six characters as the talker.

Method

Participants

Seventy-two children enrolled in a local preschool located in an upper-middle-class neighborhood participated in the study. The sample included 24 children in each of three groups, 3, 4, and 5 years of age. The mean ages of the groups were 3 years, 7.5 months (range = 36–47 months); 4 years, 7.8 months (range = 48–59 months); and 5 years, 4.6 months (range = 60–71 months). None of the children had any diagnosed disabilities. All passed vision and hearing screening measures for normalcy (Texas Department of Health, 1998). All also had normal listening skills as determined by their teachers and parents.

Stimuli and Design

Pictures and voice samples for 20 target characters and pictures for 40 foil characters from popular children's television and video programs served as stimuli. Both target and foil characters were sampled from television listings representing weekday afternoon and Saturday morning broadcasts. Table 1 details the 20 specific cartoon characters.

The voice stimuli consisted of two utterances produced by each of the 20 characters. Stimuli were selected that were approximately 4 s in length and that were judged subjectively to be of equivalent loudness and of high fidelity. The beginning of each utterance represented a sentence onset. The utterances were recorded from the videocassette audio line output directly onto cassette tape using a Marantz PMD 201 cassette recorder. The recording system was adjusted to approximately 0 VU for each character during stimulus recording and playback. Two sets of utterances, each containing one utterance for each character, were used to ensure that children's recognition of specific talkers was not utterance-specific. Each set of utterances was edited onto cassette tape such that the order of the characters' voices was presented in random order with 3-s intervals between utterances. The utterances were presented at approximately 70 dB SPL using a Marantz PMD 201 cassette recorder and a Realistic MPS-5 loudspeaker. One half of the children in each age group listened to one of the two sets. No effects of utterance set were found in statistical analyses.

Table 1. Characters presented in the voice recognition task.

Alvin Chipmunk	Kermit the Frog
Barney Rubble	Lucy Van Pelt
Big Bird	Miss Piggy
Bugs Bunny	Mr. Rogers
Cookie Monster	Olive Oil
Eeyore	Popeye
Ernie	Porky Pig
Fred Flintstone	Raphael Turtle
Garfield	Ted
Heathcliff	Winnie the Pooh

Six 8 x 10.5 cm laminated color pictures were presented with each voice sample. The pictures were attached with Velcro to a 28 x 35.5 cm white card, positioned in two columns and three rows with equidistant spacing between pictures. Each card contained the picture of the talker as well as pictures of five foil characters. All children saw the same sets of pictures. The 20 target pictures were presented twice during the session, once as the target character and once as a foil picture (within a set of 5 foils and 1 target picture). At least two intervening trials were always presented between a picture's serving as the target and a foil. Each of the 40 foil pictures was also presented twice during the experimental session. Positions of the target pictures were counterbalanced so that the target character appeared in each of the six card positions at least three times.

Procedure

Each child was seated at a table in a quiet room with the loudspeaker placed 40 cm in front of the child. A small easel containing the picture cards was positioned to the left of the loudspeaker. The experimenter explained to the children that they were going to play a game in which samples of "talking" would be presented and that they should point to the picture of who was talking. The experimenter positioned the picture card before the child and then played the 4-s utterance. The tape was paused if necessary to allow time for the child's response and for the experimenter to record each response on a data sheet. This sequence was repeated for presentation of all 20 utterances. Children typically produced responses within 2 s and without prompting.

This talker recognition task was followed by a control task in which a series of color pictures (21.5 x 15.5 cm) of the 20 target characters was presented in one of two random orders. Participants were asked to name each character, and responses were recorded by the experimenter. This control task was included as a measure of each participant's familiarity with the characters. If a child could spontaneously name the character, that character was defined as a "more familiar" character for that child. If a child could not provide the name for a character, but demonstrated some knowledge of the character's context or activities (e.g., "He's on Sesame Street"), then that character was designated as a "less familiar" character for that child. There were no completely unfamiliar characters for any child. Familiarity

served as a factor in the data analyses, allowing us to address whether recognition differed for more familiar and less familiar talkers.

Data Analysis

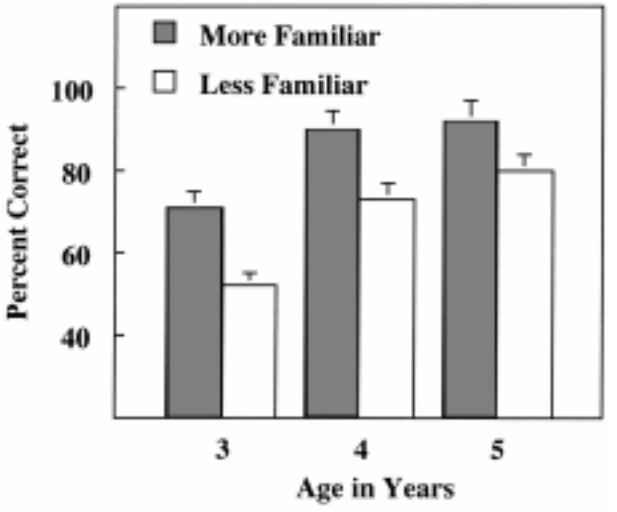
Each child's percentage of correct responses in the voice-recognition task was coded separately for more familiar and less familiar characters. These percentage measures were entered into a mixed-design Analysis of Variance (ANOVA) by regression (Pedhazur, 1982), in which the percentage of correct responses was regressed on age and familiarity. Thus, the percentage correct measure served as the dependent variable in which age was the between-subject factor and familiarity was the within-subject factor.[1] Post hoc analysis was conducted using the Scheffé test of multiple comparisons.

Results and Discussion

Figure 1 shows mean percent correct voice recognition scores for each group for more familiar and less familiar characters. Results revealed a main effect of age [$F(2, 141) = 48.19$, $p < .01$]. Overall, 3-year-olds recognized fewer voices ($M = 61.44\%$, $SE = 3.29$) than 4-year-olds ($M = 81.36\%$, $SE = 2.95$) or 5-year-olds ($M = 86.03\%$, $SE = 3.11$) (S critical = 6.6). The 4- and 5-year-olds did not differ from each other. A significant main effect for familiarity also resulted [$F(1, 143) = 93.51$, $p < .01$], such that children were more likely to recognize voices of characters they could name in the picture recognition task ($M = 84.27\%$, $SE = 2.14$) than characters they could not name ($M = 68.29\%$, $SE = 4.19$). There was no significant interaction of age and familiarity [$F(2, 140) = 1.91$]. The coefficient of determination revealed that age and familiarity together explained 78% of the variance in percent correct responses.[2]

In short, this study assessed talker-recognition abilities in preschool children between the ages of 3 and 5 years. Our data provide new information about the implicit encoding and recognition of voice signatures by children and also demonstrate that children, like adults, recognize more familiar talkers better than less familiar talkers.

Figure 1. Mean percent correct voice recognition (and standard errors) for each age group for more familiar and less familiar characters.



With regard to talker recognition ability per se, our results indicate that preschool children encode talker-specific information via implicit learning. Although the literature on preschool children's voice recognition is scarce, our results are consistent with the few other findings on this topic. First, children recognized voices of cartoon characters they had experienced through their daily activities, which is similar to the finding that 4-year-olds recognized classmates' voices (Bartholomeus, 1973). The recognition rates of children in our study were noticeably higher than those reported by Bartholomeus (1973), presumably due to the factors we noted earlier. Children's recognition of cartoon characters' voices learned via their normal activities is also similar to auditory priming effects that have been reported. Specifically, young children more accurately identified low-pass filtered words that they had heard previously than novel words (Church & Fisher, 1998). Finally, our demonstration of children's voice recognition is also consistent with work reported by Ryalls and Pisoni (1997) that 3- to 5-year-old children process talker characteristics of spoken utterances.

Voice recognition ability of the children in our study also showed significant age-related improvement between the ages of 3 and 4 years. No other studies have compared voice recognition of 3- to 5-year-olds, but there

[1] Preliminary analyses examined the contribution of linguistic content and speaking style to the identification of talkers in the absence of auditory information. Written transcriptions of the speech samples were presented to 50 adult participants who were asked to read each sample and to select the character who produced that particular speech sample out of six possible choices. We sorted talker utterances into four linguistic categories (20%, 40%, 60%, & 80%) as a function of the percentage of raters who correctly identified the talker producing each utterance from the written transcriptions. The estimated change in performance as a function of linguistic cue did not vary with age for either the more familiar [$F(1, 91) = 0.110$, $p = 0.74$] or the less familiar [$F(1, 61) = 0.423$, $p = .518$] condition, so this variable was omitted from subsequent analyses.

[2] Analysis of Variance (ANOVA) by regression (Pedhazur, 1982) in which the arcsine transformations of percentage correct scores were regressed on age and familiarity revealed the same results as the ANOVA by regression on the raw percentage correct scores. There was a main effect of age [$F(2, 141) = 86.01$, $p < .01$] and a main effect of familiarity [$F(1, 143) = 32.55$, $p < .01$], but no interaction of these variables resulted [$F(2, 141) = .26$].

is other evidence for developmental change in children's voice processing. Changes have been found in processing of talker-specific information by children between the ages of 3 and 5 years (Ryalls & Pisoni, 1997), and children's recognition of recently learned voices improved between the ages of 6 and 10 years (Mann et al., 1979).

Although these few studies imply that encoding of vocal information may change as a function of age, most research examining memory for information acquired via implicit processes has not found developmental differences. More specifically, visual priming effects for pictures do not appear to show age-related change between about 3 and 10 years (Drummey & Newcombe, 1995; Greenbaum & Graf, 1989; Hayes & Hennessy, 1996; although see Parkin, 1993). Auditory priming effects for low-pass filtered words also do not appear to differ for 2- to 3-year-olds and adults (Church & Fisher, 1998). Differences in the findings regarding developmental change observed in our study versus priming studies may be related to information-processing demands imposed by the different tasks. Priming studies typically assess differences in responses to previously experienced and novel stimuli using tasks in which a single stimulus is presented on each trial. Most of the priming studies with children have involved identification tasks; children simply identify the picture or the word (presented in a degraded version or in noise). In the talker-identification task, however, children were presented trials in which a single voice sample was presented in conjunction with pictures of six characters. To perform this task, it was necessary that children identify the person from the vocal signature and then select the correct picture on the response card. This latter task was chosen for our work because it had previously been used with children and adults and because it was a task that did not require a verbal response from young children. Nonetheless, this task may have imposed greater processing demands on 3-year-old children than the tasks that have been used in priming studies. We should emphasize, however, that despite the finding of a significant effect of age for talker recognition, performance of all three age groups of children was impressively good.

With regard to familiarity, children at all three ages were more accurate in recognizing the voices of characters they could name than the voices of characters they could not name. This result is consistent with findings in adults (Papcun, Kreiman, & Davis, 1989). A possible explanation for the familiarity effect on our task may be provided by Bruce and Young's (1986) model of face recognition. If a child can access the name of the character when presented the face of that individual, then the child has encoded several types of identity-specific information. If the child has had sufficient exposure to a particular character for encoding both name and face information, then she is also more likely to have been exposed to the character's voice and to have encoded vocal attributes of that character than if she has had less exposure to a character.

Overall, our findings suggest that talkers' vocal attributes are stored by children in long-term memory and that these vocal attributes are linked to other knowledge about the identity of the talker, such as the talker's face and name. These results also provide further support for the idea that perceptual information can be processed efficaciously by young children. Although more work is needed in this research area, the ability to encode and recognize voices, which is quite good by age 3, may play an important role in the development of children's speech processing skills. Studies of adult voice perception have consistently shown that listeners' familiarity with particular voices influences their perception of and memory for speech (Pisoni, 1993). Specifically, adults identify words produced by familiar talkers better than words produced by unfamiliar talkers (Nygaard et al., 1994). Children's ability to encode and later recognize words, which is rapidly developing in preschool children, may also be influenced by their familiarity with vocal signatures. Our research provides evidence that children are quite skilled at recognizing vocal sources, which may be an important aid to the development of spoken word recognition.

Acknowledgments

This research was supported in part by Grant DC 00421 from the National Institute on Deafness and Other Communication Disorders to The University of Texas at Dallas and by Texas Advanced Research Project Grant #009741035 from the Texas Higher Education Coordinating Board. Funding for this research also was provided by the Callier Excellence in Education Fund and a Callier Scholar award to M. Spence. We thank the children who participated and the preschool personnel who allowed us to conduct this research within their schools. We also are grateful for the assistance of Karen Thierry, Eric Wood, and Laura Woessner with data collection and stimulus preparation.

References

Abberton, E., & Fourcin, A. J. (1978). Intonation and speaker identification. *Language and Speech, 21,* 305–318.

Abercrombie, D. (1967). *Elements of general phonetics.* Chicago: Aldine.

Bach, M., & Underwood, B. (1970). Developmental changes in memory attributes. *Journal of Educational Psychology, 61,* 292–296.

Bartholomeus, B. (1973). Voice identification by nursery school children. *Canadian Journal of Psychology, 27,* 464–472.

Bennett, S., & Montero-Diaz, L. (1982). Children's perception of speaker sex. *Journal of Phonetics, 10,* 113–121.

Bricker, P. D., & Pruzansky, S. (1966). Effects of stimulus content and duration on talker identification. *Journal of the Acoustical Society of America, 40,* 1441–1449.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77,* 305–327.

Church, B. A., & Fisher, C. (1998). Long-term auditory word priming in preschoolers: Implicit memory support for language acquisition. *Journal of Memory and Language, 39,* 523–542.

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, & Cognition, 20,* 521–533.

Compton, A. J. (1963). Effects of filtering and vocal duration upon the identification of speakers, aurally. *Journal of the Acoustical Society of America, 35,* 1748–1752.

Cook, S., & Wilding, J. (1997). Earwitness testimony: Never mind the variety, hear the length. *Applied Cognitive Psychology, 11,* 95–111.

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

Drummey, B., & Newcombe, N. (1995). Remembering versus knowing the past: Children's explicit and implicit memories for pictures. *Journal of Experimental Child Psychology, 59,* 549–565.

Dukiewiez, L. (1970). Frequency-based dependence of speaker identification. In W. Jassen (Ed.), *Speech analysis and synthesis* (Vol. II, pp. 41–50). Warsaw: Institute of Fundamental Technical Research, Polish Academie of Sciences.

Eimas, P. D., & Miller, J. L. (1992). Organization in the perception of speech by young infants. *Psychological Science, 3,* 340–345.

Felzen, E., & Anisfeld, M. (1970). Semantic and phonological relations in the false recognition of words by third- and sixth-grade children. *Developmental Psychology, 55,* 163–168.

Garner, W. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Greenbaum, J., & Graf, P. (1989). Preschool period development of implicit and explicit remembering. *Bulletin of the Psychonomic Society, 27,* 417–420.

Hayes, B. K., & Hennessy, R. (1996). The nature and development of nonverbal implicit memory. *Journal of Experimental Child Psychology, 63,* 22–43.

Hollien, H., Majewski, W., & Doherty, E. T. (1982). Perceptual identification of voices under normal, stress, and disguise speaking conditions. *Journal of Phonetics, 10,* 139–148.

Houston, D., Jusczyk, P. W., & Tager, J. (1998). Talker-specificity and persistence of infants' word representations. *Proceedings of the 22nd Annual Boston University Conference on Language Development* (Vol. 1, pp. 385–396). Somerville, MA: Cascadilla Press.

Jerger, S., Pearson, D. A., & Spence, M. J. (1999). Developmental course of auditory processing interactions: Garner interference and Simon interference. *Journal of Experimental Child Psychology, 74,* 44–67.

Jerger, S., Pirozzolo, F., Jerger, J., Elizondo, R., Desai, S., Wright, E., & Reynosa, R. (1993). Developmental trends in the interaction between auditory and linguistic processing. *Perception & Psychophysics, 54,* 310–320.

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,* 630–645.

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

Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America, 29,* 98–104.

Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology, 27,* 153–165.

Martin, C. S., Mullennix, J. W., Pisoni, D. B., & Summers, W. V. (1989). Effects of talker variability on recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 15,* 676–684.

Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition, 13,* 135–165.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47,* 379–390.

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.

Murry, T., & Cort, S. (1971). Aural identification of children's voices. *Journal of Auditory Research, 11,* 260–262.

Naito, M. (1990). Repetition priming in children and adults: Age-related dissociation between implicit and explicit memory. *Journal of Experimental Child Psychology, 50,* 462–484.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics, 60,* 355–376.

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.

Papcun, G., Kreiman, J., & Davis, A. (1989). Long-term memory for unfamiliar voices. *Journal of the Acoustical Society of America, 85,* 913–925.

Parkin, A. J. (1993). Implicit memory across the lifespan. In P. Graf & M. W. J. Masson (Eds.), *Implicit memory: New directions in cognition, development, and neuropsychology* (pp. 191–206). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pedhazur, E. (1982). *Multiple regression in behavioral research: Explanation and predictions* (2nd ed.). New York: Holt, Rinehart, & Winston.

Pisoni, D. B. (1993). Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. *Speech Communication, 13,* 109–125.

Pollack, I., Pickett, J. M., & Sumby, W. H. (1954). On the identification of speakers by voice. *Journal of the Acoustical Society of America, 26,* 403–406.

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology, 33,* 441–452.

Schacter, D. L., Church, B., & Treadwell, J. (1994). Implicit memory in amnesic patients: Evidence for spared auditory priming. *Psychological Science, 5,* 20–25.

Schweinberger, S. R., Herholz, A., & Sommer, W. (1997). Recognizing famous voices: Influence of stimulus duration and different types of retrieval cues. *Journal of Speech, Language, and Hearing Research, 40,* 453–463.

Scott, M. S., Serchuk, R., & Mundy, P. (1982). Taxonomic and complementary picture pairs: Ability in two-to five-year olds. *International Journal of Behavioral Development, 5,* 243–256.

Shankweiler, D. P., Strange, W., & Verbrugge, R. 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.

Spence, M. J., & DeCasper, A. J. (1987). Prenatal experience with low-frequency maternal voice sounds influences neonatal perception of maternal voice samples. *Infant Behavior and Development, 10,* 133–142.

Spence, M. J., & Freeman, M. S. (1996). Newborn infants prefer the maternal low-pass filtered voice, but not the maternal whispered voice. *Infant Behavior and Development, 19,* 199–212.

Studdert-Kennedy, M. (1974). The perception of speech. In T. A. Sebeok (Ed.), *Current trends in linguistics* (pp. 2349–2385). The Hague: Mouton.

Summerfield, Q., & Haggard, M. P. (1973). Vocal tract normalization as demonstrated by reaction times. *Report of speech research in progress* (Vol. 2, pp. 12–23). Belfast, Northern Ireland: The Queen's University of Belfast.

Texas Department of Health. (1998). *Special Senses and Communication Disorders Act, Texas health and safety codes.* Austin, TX: Bureau of Children's Health, Vision and Hearing Screening.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 247,* 301–306.

Van Lancker, D. R., Kreiman, J., & Emmorey, K. (1985a). Familiar voice recognition: Patterns and parameters. Part I: Recognition of backward voices. *Journal of Phonetics, 13,* 19–38.

Van Lancker, D. R., Kreiman, J., & Wickens, T. D. (1985b). Familiar voice recognition: Patterns and parameters. Part II: Recognition of rate-altered voices. *Journal of Phonetics, 13,* 39–52.

Werker, J. F. (1995). Exploring developmental changes in cross-language speech perception. In L. R. Gleitman & M. Liverman (Eds.), *An invitation to cognitive science* (2nd ed., Vol 1., Language). Cambridge, MA: MIT Press.

Received November 29, 2000

Accepted October 15, 2001

DOI: 10.1044/1092-4388(2002/016)

Contact author: Melanie J. Spence, PhD, Box 830688, GR 4.1, The University of Texas at Dallas, Richardson, TX 75083-0688. E-mail: mspence@utdallas.edu