Published in final edited form as:
Dev Psychol. 1997 May ; 33(3): 441–452.

The Effect of Talker Variability on Word Recognition in Preschool Children

Brigette Oliver Ryalls and **David B. Pisoni**
Department of Psychology, Indiana University Bloomington.

Abstract

In a series of experiments, the authors investigated the effects of talker variability on children's word recognition. In Experiment 1, when stimuli were presented in the clear, 3- and 5-year-olds were less accurate at identifying words spoken by multiple talkers than those spoken by a single talker when the multiple-talker list was presented first. In Experiment 2, when words were presented in noise, 3-, 4-, and 5-year-olds again performed worse in the multiple-talker condition than in the single-talker condition, this time regardless of order; processing multiple talkers became easier with age. Experiment 3 showed that both children and adults were slower to repeat words from multiple-talker than those from single-talker lists. More important, children (but not adults) matched acoustic properties of the stimuli (specifically, duration). These results provide important new information about the development of talker normalization in speech perception and spoken word recognition.

Understanding spoken language involves mapping from speech—a complex pattern of acoustic energy—to meaning—an abstract representation in the mind. In traditional accounts of language, a number of processing steps intervene between the acoustic input and the comprehension of meaning (Studdert-Kennedy, 1976). These steps include the analysis of the input into sounds or phonemes, the use of strings of phonemes to access words from the lexicon, the semantic processing of the words, and the syntactic processing of strings of words. This article concerns the beginning steps of this process—going from the physical energy that is speech to the abstract representation of a word. The larger theoretical questions concern the nature of this physical sound-to-abstract representation transition—the processes involved and how such processes might develop in the course of language learning (see Best, 1994). The specific issue that we investigated to shed light on these questions is the development of talker normalization.

Talker normalization is how a listener accesses the same word from the lexicon despite wide variation in the acoustic properties from one speaker to the next (Klatt, 1986). There are numerous sources of interspeaker variability, including differences in the size and shape of the vocal tract (Peterson & Barney, 1952), differences in glottal characteristics (Carrell, 1984), idiosyncratic differences in articulatory strategies (Ladefoged, 1980), as well as differences in dialect. To illustrate, Figure 1 shows differences in the acoustic properties of the word *cash* spoken by three different speakers. Although globally similar, the acoustic signals differed from speaker to speaker on several characteristics. Despite these differences in the acoustic signals, a listener hearing these three utterances would hear the same word composed of the same linguistic units. Talker normalization refers specifically to the

Correspondence concerning this article should be addressed to Brigette Oliver Ryalls, who is now at the Department of Psychology, University of Nebraska, Omaha, Nebraska 68182-0274. Electronic mail may be sent via Internet to bryalls@cwis.unomaha.edu.

processes through which the same linguistic units are arrived at despite speaker dependent differences in the acoustic signal.

How is it that listeners hear these various tokens as the same word? How is it that listeners identify the relevant attributes of a highly variable speech signal to specify abstract linguistic units? The dominant view in the field is conveyed by the word *normalization*. The major longstanding assumption is that there is a set of processes that makes the acoustic inputs all the same—a set of processes that corrects for or eliminates speaker differences (Joos, 1948; Krulee, Tondo, & Wightman, 1983; Studdert-Kennedy, 1974). According to this account, what is the same across different tokens of the same word or utterance is the phonemic content, that is, the abstract language-specific categories of sounds present in a given utterance. The assumption, then, is that the idiosyncratic acoustic signals of different speakers somehow contain information that specifies speaker-independent and uniform phonemes (i.e., /k/ /æ/ /ʃ/). Listeners hear the different tokens of the same linguistic category as the same by stripping away the irrelevant speaker dependent “noise” to find what is the same. A simplified representation of this view of normalization is illustrated in Figure 2. This figure shows that, although the acoustic pattern may vary from speaker to speaker and from time to time, the same phonemes are accessed, and it is these specifically linguistic units that contact representations in the mental lexicon. This traditional view of normalization is beginning to be questioned in the adult literature (Pisoni, 1990; Goldinger, 1992). We address this literature and the issues it raises in the General Discussion. The present concern is what the traditional account implies about development.

For the standard view of normalization to be correct, speakers must either be born with normalization processes capable of extracting from the speech signal the phonemes of all languages or they must acquire these processes. It seems unlikely that normalization processes would be completely innate because of the nature of different languages (see Jusczyk, 1994). What constitutes irrelevant variability in one language (e.g., pitch in English) can specify differences in meaning in another language (e.g., tone in Chinese). If normalization processes are at least partially learned, then how is this learning accomplished? One proposal is that children learn to strip away more and more of the irrelevant talker-specific information, leaving only abstract phonemic information—that by stripping away what is noise in their language, they “find” the phonemes. This notion is consistent with current theories of phonological development that posit a developmental trend from nonspecific acoustic representations to speech-specific, segmental representations (Jusczyk, 1993; Walley, 1988). If this proposal is correct, then the degree to which listeners are affected by speaker variability should decrease with age. What is known about the development of talker normalization, however, is confined to the study of infants and adults in tasks so dissimilar that no developmental conclusions are possible.

Studies have shown that infants can discriminate between different speakers (DeCasper & Fifer, 1980; Jusczyk, Pisoni, & Mullennix, 1992) and can generalize speech sounds across different speakers (Kuhl, 1979, 1983). However, research indicate that the ability to normalize across voices comes with some cost to performance. For example, Jusczyk, Pisoni, and Mullennix (1992) tested 2-month-old infants in a habitation paradigm. They found that after being habituated to a single voice repeating a syllable, 2-month-olds dishabituated when either the voice or the syllable was changed. However, when infants were habituated to multiple talkers repeating the same syllable, they dishabituated when the syllable was changed but did not dishabituate when switched to another set of voices repeating the same syllable. In a second experiment, Jusczyk, Pisoni, and Mullennix introduced a delay period between habituation and test. With the addition of a delay, infants no longer dishabituated to a syllable change in the multiple-talker condition, although they did in the single-talker condition. This result suggests that talker variability disrupts infants’ memory for speech sounds. Overall, the evidence indicates that infants are capable of recognizing the similarity of the same speech sounds produced by different speakers but that, in some cases, the demands of processing different talkers disrupts retention of the particular speech sound.

The evidence from adult studies leads to virtually the same conclusion. Although adults are capable of recognizing similar phonetic content across speakers, close examination indicates performance costs for processing speech from multiple talkers. For example, talker variability has been shown to reduce performance in word recognition, naming, and recall paradigms (Creelman, 1957; Goldinger, Pisoni, & Logan, 1991; Martin, Mullennix, Pisoni, & Sommers, 1989; Mullennix, Pisoni, & Martin, 1989). What is not known is what happens to the processing of talker variability in the interval between infancy and adulthood. Although a superficial comparison of infant and adult studies indicates that infants and adults show similar effects of talker variability, the tasks that were used with the two age groups were very different. Studies with infants involved perceptual discrimination, whereas studies with adults involved recognition and recall at the lexical level. The evidence on phonological development suggests that lexical representations specific to a child’s native language may continue to emerge with development post-infancy and in the preschool period (Best, 1994; Walley, 1988; Werker, 1994). Therefore, it is critical that the effects of talker variability are examined across the entire developmental continuum.

By studying normalization in children using methods similar to those used with adults, we begin to close the gap between infant and adult studies and to answer questions about the processes and mechanisms involved in understanding multiple talkers. The specific question we ask in Experiments 1 and 2 is whether talker variability disrupts word recognition for younger preschoolers more than for older preschoolers under ideal conditions and under conditions of stimulus decay. In addition, in a third experiment, we ask if talker variability affects older preschoolers and adults differentially in their ability to repeat spoken words.

Experiment 1

In this study, we investigated the effects of talker variability on word recognition in 3-, 4-, and 5-year-old children. The task was a simple one for children—to point to one picture (out of six possible pictures) that corresponded to a spoken word. We manipulated whether the words in the list were spoken by one talker or by several different talkers. The central question of this experiment was whether speech from multiple talkers would disrupt younger children’s word recognition more than older children’s word recognition performance. This result would be expected if children are in the process of learning to strip away the acoustic information irrelevant to their language in the course of learning this language.

We anticipated that because of their greater experience with language, older children would identify more words correctly than would younger children across both stimulus conditions. Given the evidence that talker variability causes measurable deficits in adult performance under stimulus degradation, it would not be surprising to find poorer word recognition in the multiple-talker condition than in the single-talker condition at all age levels. The critical prediction, then, is that this deficit (performance on multiple-talker list worse than performance on single-talker list of words) will decrease with age, resulting in an interaction between age and talker variability.

Method

Participants—Thirty children, 10 each at ages 3, 4, and 5 years, were recruited from the surrounding community by an ad in the local newspaper. The average ages were 3 years 8 months, 4 years 5 months, and 5 years 6 months. There were an approximately equal number of boys and girls in each age group, and all were monolingual. Each participant was tested separately in a single session lasting approximately 30 min. Children or their parents were paid for their participation. Two participants who partially completed die experiment received payment but were not included in the final analyses.

Stimulus materials—Three word lists (25 words each) from the Word Intelligibility by Picture Identification (WIPI) test were used as stimuli for this experiment (Ross & Lerman, 1970; see the Appendix for a complete listing). The WIPI is a test designed to assess speech discrimination abilities of young children. All words were monosyllabic and have an average adult familiarity of 6.957 out of 7.0 (Nusbaum, Pisoni, & Davis, 1984) and an average adult frequency of 99.45 (Kucera & Francis, 1967). In clinical applications, the person administering the WIPI test reads the test words aloud. The child is shown a display of six pictures (a different display is used for each word) and is instructed to identify the picture of the word they hear. Sound similarity among the six pictures in each display is high. For example, one display contains representations of a crown, a mouth, a mouse, a clown, a cow, and a house. We prerecorded all of the stimulus words on audiotape and played the tapes to children over headphones.

Seven men and 7 women served as talkers to produce the original stimulus materials. All talkers were from the Midwestern region of the United States and had no accents. The 75 test words were presented randomly on a CRT screen in front of the talker who was seated in a sound-attenuated booth (Industrial Acoustics Co., Bronx, NY; Model 401A). Utterances were recorded on audiotape with an Electro-Voice (Buchanan, MI) Model DO54 microphone and an Ampex (Redwood City, CA) Model AG-500 tape recorder. The talkers were instructed to read the words aloud in a normal voice at a constant speaking rate. The words were then converted into digital form with a 12-bit analog-to-digital converter running at 10-kHz sampling rate. The root mean square amplitude levels of the words were digitally equated, and the test words were edited with a digitally controlled waveform editor (Luce & Carrell, 1981). These operations resulted in a database of 75 words spoken by 14 talkers for a total of 1,050 stimulus tokens.

The 1,050 stimulus tokens were then presented to adult participants to obtain identification scores. Seven adults participated in two, 1-hour sessions. In one session, the adult participants heard the 525 tokens spoken by men, and in the other session they heard the 525 tokens spoken by women. All stimuli were presented by means of headphones, and participants were instructed to record the word that they heard by typing their response into a computer keyboard that was in front of them. Results were tallied and taken as a measure of the intelligibility of each token. The male talker with the highest identification score across all 75 tokens was chosen for use in the single-talker condition. All tokens from this talker were identified correctly by at least 86% (or six out of seven) of the judges. Audiotapes were made with this voice for each of the three word lists that were used for the single-talker condition. The one male voice and two female voices with the lowest identification scores were eliminated from the database, leaving five male and five female voices that were used to construct the multiple-talker tapes. Tokens were chosen at random with the requirement that each stimulus was identified at least above 86% (or 6 out of 7) correct identification. The number of words spoken by each of the 10 talkers was as equal as possible on each list (five of the voices spoke two words each, the other five voices spoke three words each for a total of 25 words per list). Lists were also balanced for gender of the talker.

Design and procedure—Testing occurred in a single session lasting approximately 30 min. All children were tested individually by the same experimenter in a small, well-lit room. All participants were given a pure-tone screening test (at frequencies of 500, 1000, 2000, and 4000) prior to participation in the experiment to ensure that they did not have any major hearing problems. All children passed the screening test. During the experiment, the child sat across from the experimenter either in the parent's lap or in a chair next to the parent. Parents were asked not to assist or coach the child in any way throughout the course of the experiment, and all cooperated with these instructions. (They could not hear the stimulus words and therefore could not tell when children made errors.) Children were told that they were going to play a game with the experimenter and that they could "win" a sticker by playing. Children were instructed to listen to the words that were presented over the earphones and to point to the picture of what they heard. A practice trial was completed to ensure that the child understood the instructions and could carry out the task.

In the practice trial, the child was shown a practice page with six pictures. The experimenter asked, "What would you point to if you heard the word *x*"? with *x* being one of the six pictures (i.e., *cat*). This procedure was repeated with the same practice page one or two more times to ensure that the child understood the nature of the task. None of the children had any difficulty understanding the experimental procedures. The experiment then began with stimulus words presented to the child through Telephonics Corp. (Farmingdale, NY) TDH-39 headphones with a Uher 4000 Report-L tape recorder (Martel, Anaheim, CA). The experimenter would say "show me this," or some analogous prompt, play a test word, then stop the tape recorder until the child responded. Responses were recorded by the experimenter on a response sheet out of view of both the parent and the child. The experiment continued in this fashion until the 25 words on the list were completed. Once or twice per list the experimenter would remind the child of the instructions. After the list was completed, the child chose a "prize" sticker. After a short break, a second (different) list of words was completed in the same fashion.

Each child completed both a single-talker and a multiple-talker list. Half of the participants at each age completed the single-talker list first, and half completed the multiple-talker list first. Word lists were also counterbalanced across participants. An approximately equal number of boys and girls participated at each age.

Results and Discussion

Because the initial analyses including all variables showed no effect of participant gender or word list and no interactions including these variables, the data were collapsed across the levels of both variables. Performance was high at all ages. The overall mean number of words correct was 20.85, or 83%.

The number of correct responses was analyzed with a mixed model analysis of variance (ANOVA) with variables of age, list order, and talker condition. There were three age levels, two list orders (single-first and multiple-first), and two levels of talker condition (single-talker and multiple-talker). The ANOVA revealed a main effect of age, $F(2, 24) = 3.67$, $p = .039$, a main effect of talker condition, $F(1, 24) = 7.40$, $p = .011$, an interaction between age and talker condition, $F(2, 24) = 7.41$, $p = .003$, and an interaction between list order and talker condition, $F(1, 24) = 28.24$, $p < .001$.

Accuracy increased with age as expected. Five-year-olds performed better than 4-year-olds who performed better than 3-year-olds. The overall mean numbers of correct responses were 22.45, 20.5, and 19.6, respectively. Overall, performance was better in the single-talker condition than in the multiple-talker condition, with means of 21.2 and 20.5, respectively.

As shown in Figure 3, however, these overall findings are complicated by interactions. Post hoc comparisons indicated that, in the single-first condition (top), there were no significant differences between the single-talker and multiple-talker lists at any age. In the multiple-first condition (bottom), however, both 3- and 5-year-olds showed significant differences in performance between the single-talker and multiple-talker lists (Tukey's honestly significant difference, $p < .05$). Four-year-olds did not show an effect although their data showed a slight trend in the same direction.

One possible explanation of this interaction is that the two effects of order and talker variability work in opposite directions, having different strengths at different points in development. That is, word recognition may benefit from experience in the task such that performance is better on the second list than on the first list. Further, word recognition may be better on single-than on multiple-talker lists. It is unclear by this account why 4-year-olds do not show this effect in the multiple-talker list first condition but 3- and 5-year-olds do. This finding may reflect that the effects are quite small overall at this near ceiling-level performance. Perhaps with more children or with more trials, the effect would have been significant for 4-year-olds as well.

Alternatively, perhaps the performance of the 4-year-olds indicates that children's general speech perception abilities and their talker normalization abilities do not develop in lockstep. That is, perhaps between 3 and 4 years of age children's ability to deal with multiple talkers develops significantly but their general speech perception abilities change little. Between 4 and 5 years of age, there may be significant development in children's general speech perception abilities but little change in their ability to deal with multiple talkers. This possibility is suggested by the finding that from 3 years to 4 years of age, children's gains are more substantial in performance on the multiple-talker list, whereas between 4 and 5 years of age, their gains are more substantial in performance on the single-talker list. Although quite speculative, this interpretation points to the possible separability of processes in development and deserves further investigation.

Regardless of how the lack of a significant difference for 4-year-olds is explained, there are interesting differences between 3- and 5-year-olds and what they learn (or don't learn) from the first list that they hear. Specifically, 3-year-olds appear to learn something from the single-talker list when it comes first that benefits their performance on the multiple-talker list when it comes second but not vice versa. Their performance on the single-talker lists in the two order conditions is nearly identical (that is, unaffected by order), but performance on the multiple-talker lists is significantly different—performance is higher when the multiple-talker list comes after the single-talker list than when it comes first (see Figure 3). In contrast, 5-year-olds appear to learn something from the multiple-talker list when it comes first that benefits their performance on the single-talker list when it comes second. Five-year olds' performance on the multiple-talker lists in the two order conditions is not significantly different whereas performance on the single-talker lists is significantly different—performance is higher when the single-talker list comes after the multiple-talker list than when it comes first (see Figure 3). Again, these results point to the possible separability of processes in development and merit further investigation.

In summary, with regard to our major question of interest, these results suggest that there can be a cost for talker variability in young children even when identifying words under optimal conditions. However, overall level of performance was quite high, perhaps making it difficult to discern the effects of stimulus manipulations. In adult studies, word recognition tasks must be made difficult by degrading the stimuli before talker variability effects are noticeable in performance. Children in Experiment 1 were presented with words under optimal conditions and still showed an effect of talker variability. These results thus support the proposal that the ability to normalize speech from different talkers develops. To provide further evidence for this hypothesis, Experiment 2 replicated Experiment 1. In Experiment 2, however, words were presented against a background of noise. Experiment 2 thus provides a potentially more sensitive measure of the development of normalization processes.

Experiment 2

We investigated the effect of talker variability on word recognition in 3-, 4-, and 5-year-old children when stimuli were presented against a background of noise. In this experiment, the loudness or amplitude of the words was set to equal the loudness of the white noise (0 signal/noise ratio). In a study of adult performance, Mullennix et al. (1989) found that, at this level of degradation, word identification in a multiple-talker condition is approximately 20% lower than is performance in a single-talker condition. Adults in the Mullennix et al. study, however, did not have a closed set of response alternatives, as did children in the present experiment.

Method

Participants—Thirty-six children, 12 each at ages 3, 4, and 5 years, were recruited to participate in this experiment from the surrounding community by an ad in the local paper. There were approximately equal numbers of boys and girls in each age group, and all children were monolingual. The average age for each group was 3 years 6 months, 4 years 7 months, and 5 years 8 months. Each child was tested separately in a single session lasting approximately 30 min. Children or their parents were paid for their participation.

Stimulus materials—Because we found no differences between the three word lists that were used in the first experiment, we used only two word lists (List 1 and List 2 from the WIPI test that were used in the previous experiment) as stimuli for this experiment to simplify counterbalancing. Words were presented in a background of white noise at a zero signal-to-noise ratio. Both single- and multiple-talker lists were used.

Design and procedure—The procedure that we used for this experiment was identical to that used in Experiment 1. The two word lists (List 1 and List 2; see Appendix 1) and the two talker conditions (single and multiple) were completely counterbalanced by list order (first or second). Three children at each age were assigned to each of the resulting four conditions (Single List 1 Multiple List 2, S1M2; Single List 2 Multiple List 1, S2M1; Multiple List 1 Single List 2, M1S2; and Multiple List 2 Single List 1, M2S1).

Results and Discussion

There was no indication of gender, word list, or list order differences (or interactions including these variables) in the initial analysis (The List Order × Talker Condition interaction did not approach significance in this experiment $F(1, 30) = .445$.) Therefore, the data were collapsed across all levels of these factors. Overall, performance fell relative to Experiment 1: The overall mean was 14.68, or 59% of the words chosen correctly. Figure 4 shows the number of words correct plotted as a function of age and talker condition.

The number of correct responses was analyzed with a mixed model ANOVA with the variables age and talker condition. There were three age levels and two levels of the talker condition (single-talker and multiple-talker). The ANOVA revealed a main effect of age, $F(2, 33) = 17.24$, $p < .001$, a main effect of talker condition, $F(1, 33) = 52.69$, $p < .001$, and a marginal Age × Talker condition interaction, $F(2, 33) = 3.06$, $p = .07$. As expected, overall accuracy increased with age. Performance was better in the single-talker condition than it was in the multiple-talker condition at all ages (see Figure 4).

Because younger children performed less well overall than did older children, a direct comparison of number correct on the multiple- and single-talker lists may not be the best measure of the magnitude of the deficit that is caused by talker variability at different age levels. A measure of the percentage drop in performance from single to multiple might be more appropriate. Accordingly, we calculated a percentage score for each child by dividing the number of words correctly identified in the multiple-talker condition by the number of words correctly identified in the single-talker condition and multiplying this number by 100 (see Figure 5). This score indicates the level at which each child performed in the multiple-talker condition relative to the single-talker condition. For example, a percentage score of 70 indicates that performance in the multiple-talker condition was 30% lower than it was in the single-talker condition. These ratio scores were analyzed with a between-subjects ANOVA with the variables of age and list order. This analysis yielded a significant effect of age, $F(2, 30) = 3.38$, $p = .047$. Post hoc Tukey's honestly significant difference (HSD) analyses conducted on the percentage scores revealed that the 5-year-olds differed significantly from the 3-year-olds but that neither was significantly different from the 4-year-olds. By this measure, the multiple-talker list disrupted the performance of younger children more than older children. As predicted, talker variability becomes easier to process with development.

Unlike Experiment 1, there were no significant effects of list order in either the original analysis or the ratio data analysis—performance in the single-talker condition was superior to performance in the multiple-talker condition regardless of the order of presentation. However, although not statistically significant, the difference between talker conditions was larger when the multiple-talker list was presented first than when it was presented second. Specifically, when presented first, children's performance on the multiple-talker list was 22% lower than their performance on the single-talker list. However, when presented second, children's performance on the multiple-talker list was only 17% lower than their performance on the single-talker list.

Experiment 3

The results of Experiment 2 also support the proposal that talker variability decreases the word recognition performance of younger children more than it does older children. In Experiment 3, we sought converging evidence for this idea by examining the effects of talker variability on naming latency in 4- and 5-year-olds and m adults. Latency studies with adults have shown that they are slower to repeat words from a list spoken by multiple talkers than from a list in which all words are spoken by a single talker (Mullennix, Pisoni, & Martin, 1989). Use of processing time as opposed to errors might provide a better measure of any developmental changes that occur in learning to process talker variability. We also asked when in processing talker variability causes difficulty for children by measuring response duration—that is, the time it takes to say a word.

We used two measures of response latency. The first has been typically used in adult studies: the duration from the onset of a stimulus token to the onset of the participant's response (labeled "Total Latency" in Figure 6). Our second, *offset-to-onset*, was a measure of the latency from the offset of the stimulus to the onset of the participant's response (see Figure 6). This was important because one of the acoustic properties that varies across different speakers is speaking rate, and this property was not controlled in the construction of the stimulus tapes. As we report, our single talker did speak faster, on average, than did our multiple talkers. By examining offset-to-onset duration, we can rule out potential confounds that resulted from this difference in speaking rate.

In addition to the two latency measures, we also examined duration of the response ("Response Duration" in Figure 6). The two latency measures—total latency and offset-to-onset latency—measure time prior to the start of a response. This is a reasonable measure to examine if the traditional assumptions (as represented by the simple model in Figure 2) are correct, and all processes relevant to normalization are complete prior to the start of the response. If listeners strip away all linguistically irrelevant information down to the abstract phonemes before formulating a response, then no effects of talker variability would be expected in the duration of the responses. If, however, normalization processes develop over time, then discrete processing stages may emerge only with development. If young children are still in the process of learning to strip away variability, then the normalization processes may not be complete before the response is formulated, and the effects of talker variability might be measurable in the duration of young children's responses. As will be reported, children did indeed speak slower when responding to items from a multiple-talker list than from a single-talker list. This effect, however, was caused at least in part by the difference in speaking rates between our single and multiple talkers.

Method

Participants—Thirty-six participants (12 four-year-olds, 12 five-year-olds, and 12 adult college undergraduates) participated in this study. The mean ages for the children were 4 years 7 months and 5 years 6 months, respectively. There were approximately equal numbers of male and female participants in each age group, and all children were monolingual. Children were recruited from the surrounding community, and their parents were paid for their participation. Adults were recruited from the undergraduate population and were also paid for their participation. All participants were given a pure tone screening test, and none were found to have a hearing deficit. (Three-year-olds were not included because pilot testing indicated that their production was so poor as to preclude judgments about the accuracy of their responses.)

Stimulus materials—The same audiotapes that were used in Experiment 2 were used in this experiment. Word lists were presented in the clear with no white noise.

Design and procedure—The procedure that we used in this experiment was modified from the previous experiments. Participants were asked to repeat words as quickly as possible. No visual display was present. A Realistic PZM microphone was placed on the table in front of the participant to record responses. A Marantz (Aurora, IL) PMD-430 cassette recorder recorded both the stimulus words and the participants' responses in real time. For the children, a stuffed toy monkey was placed in the center of the table directly behind the microphone. Children were told that the monkey wanted to know what they heard through the headphones and that they should tell the monkey as fast as they could.

Before beginning the actual experiment, each child completed several practice trials in which she or he was asked to repeat words spoken aloud by the experimenter. All of the children understood the naming task. As in Experiments 1 and 2, the experimenter conducted each trial by prompting the participant (i.e., "Here's the next word"), playing a test word, and then pausing the stimulus tape until the participant responded.

All participants completed two lists with a short break between lists. During the break, the children were allowed to select a "prize" sticker. As in Experiment 2, the two word lists (List 1 and List 2; see Appendix 1) and two talker conditions were counterbalanced by list order. Three participants at each age were assigned to each of the four resulting conditions (S1M2, S2M1, M1S2, M2S1). Male and female participants were approximately equally distributed.

The audiotapes of each child's performance were measured with a digitally controlled waveform editor (Luce & Carrell, 1981). Four measurements were made for analysis (see

Figure 6): first, the length in milliseconds of each stimulus token ("Stimulus Duration" in Figure 6); second, the length in milliseconds of the latency from the onset of each stimulus to the onset of the participant's response ("Total Latency" in Figure 6); third, the length in milliseconds of the latency from the offset of each stimulus to the onset of the participant's response ("Offset-to-Onset Latency" in Figure 6); and fourth, the length in milliseconds of each response token ("Response Duration" in Figure 6).

Results and discussion

Analyses were conducted on the latencies for all correct responses (see Table 1). Because initial analyses including all variables indicated no effects or interactions involving gender, word list, or list order, the analyses were collapsed across all levels for those variables. Data were analyzed with a mixed model ANOVA with age and talker condition as main variables. There were three age levels (four, five, and adult) and two levels of talker condition (single talker and multiple talker).

Total latency—The ANOVA conducted on the total latency measurement revealed a main effect of age, $F(2, 841) = 54.46$, $p < .01$, and a main effect of talker condition, $F(2, 841) = 56.14$, $p < .01$. As shown in Figure 7, by this measure children and adults responded more quickly to words from single-talker lists than to words from multiple-talker lists. However, the total latency measure included the length of the stimulus word and the time between the stimulus and the response. Thus, because the speaking rate of the different talkers was not controlled for in the construction of the stimulus tapes, increased times could be due to stimulus differences alone. An ANQVA was conducted on the stimulus duration measure. The results revealed a main effect of talker condition, $F(2, 841) = 113.52$, $p < .01$. Words from single-talker lists averaged 426 ms in duration while words from multiple talker lists averaged 460 ms in duration—that is, the talker in the single-talker condition spoke faster than the talkers in the multiple-talker condition who were saying the same words. Thus, the total latency difference reflects at least in part the finding that the stimulus items from the multiple-talker lists were 34 ms longer on average than those same stimulus items from the single-talker lists.

Offset-to-onset latency—An ANOVA on the offset-to-onset measure (see Figure 6) revealed a main effect of age, $F(2, 841) = 51.77$, $p < .01$ and a main effect of talker condition, $F(2, 841) = 6.32$, $p = .01$, (see Figure 8). By this measure, the time to respond to a single-talker list was significantly faster than was the time to respond to a multiple-talker list at all ages. In addition, the main effect of age indicates that 4-year-olds were generally slower to respond than were 5-year-olds and adults, who did not differ (post hoc Tukey's HSD, $p < .05$). These findings indicate that talker variability did not cause special difficulty for younger listeners as compared with older listeners in this task. That is, there was no Age × List interaction for either latency measure.

Response duration—Finally, analyses were conducted on the response durations (see Figure 6). The results revealed a main effect of age, $F(2, 841) = 213.91$, $p < .01$, a main effect of talker condition, $F(2, 841) = 11.45$, $p < .01$, and an Age × Talker Condition interaction, $F(2, 841) = 3.35$, $p < .035$. As shown in Figure 9, adults' responses were shorter in duration than either 4- or 5-year-olds', which did not differ (post hoc Tukey's HSD, $p < .05$). This finding is compatible with previous findings that the duration of speech decreases with age (Smith, 1992). In addition, 4-year-olds' responses to a single-talker list were shorter than were their responses to a multiple-talker list (post hoc Tukey's HSD, $p < .05$). Five-year olds and adults did not show this difference. This effect means that for the youngest participants, but not for the older children and adults, normalization is not complete prior to the beginning of an utterance. The effects of task difficulty are measurable in the production of the word.

This last finding—that the length of younger children's utterances differ with talker condition—has two potential origins. First, given that the multiple-talker items were longer on average than were the single talker items, this finding could arise if children were simply matching the physical length of the stimuli. This possibility is consistent with the idea that we discussed earlier that younger children may not be able to successfully "strip" the input of its linguistically irrelevant talker-specific information. Second, the slower response duration could arise because of the greater difficulty of the normalization process for younger children. They could be capable of completing this task (repeating the words), but the drain on attention and resources could nonetheless have residual—deleterious—effects on the next step in processing and could result in longer response duration overall (see McClelland, 1979).

To investigate these two possibilities, we removed the 5 longest multiple-talker stimulus items and the 5 shortest single-talker stimulus items from the analyses. This left 20 stimulus words per list for each of the two lists. An ANOVA on this reduced data set indicated no stimulus differences between the single- and multiple-talker conditions in the stimulus duration measure, $F(1, 687) = .35$, $p = .55$. If the youngest children were matching the length of the stimuli, then eliminating these tokens from the analyses should eliminate the difference in response duration between the two talker conditions. If, however, the youngest children's responses to the multiple-talker list were longer in general because of an effect of psychological difficulty, we would still expect a significant effect of talker condition.

ANOVAs on the total latency and on the offset-to-onset measures again revealed main effects of age, $F(2, 687) = 44.06$ and 43.21, respectively, $p < .01$, and talker, $F(1, 687) =$ 15.13 and 13.57, respectively, $p < .01$, equivalent to those that were found in the original analysis. Thus, exclusion of a small number of items effectively eliminated the disparity between talker conditions in the mean duration of the stimuli, without affecting the significant effects of age and talker condition in the two latency measures.

With respect to the main variable of interest, response duration, reducing the data set effectively eliminated the effect of talker condition, leaving only the main effect of age, $F(2, 687) = 186.3$, $p < .01$. In other words, 4- and 5-year-olds' responses were longer in duration than were adults' responses, but there was no difference between the single- and multiple-talker conditions at any age. This suggests that the initial finding of a difference between talker conditions in the 4-year-olds' responses did not indicate a difference in overall psychological difficulty but was due to the younger children "matching" the duration of the stimulus in their response.

Further support for this idea is found in regression analyses of response duration on stimulus duration. The analysis indicated a significant correlation at all ages and at all levels of talker condition (see Table 2; all ps < .05). These significant correlations indicate that the words are similarly ordered by stimulus duration and response duration. Given that some words are naturally longer than others, independent of speaker (i.e., *cat* vs. *caterpillar*), it is not surprising to find some degree of relatedness between the two duration measures. The more relevant variable is the slope of the regression between the two measures. The slope goes beyond the mere ordering of the stimuli to indicate the degree of physical match between stimulus and response. If younger children are matching the duration of the stimulus to a greater degree than are older children and adults, then they should have correspondingly higher slopes.

The individual subject correlations between stimulus duration and response duration were calculated and submitted to an ANOVA. Results indicated that no significant differences for age or talker in the strength of these correlations. This result indicates that, although there is a significant relationship between stimulus duration and response duration, there is no difference in the strength of this relationship across ages. In other words, the degree to which the stimuli and responses are similarly ordered (by duration) does not change with age. Therefore, the initial finding of a difference between single- and multiple-talker response durations for the 4-year-olds must be due to the degree to which the participants matched the stimuli.

An ANOVA performed on the regression slopes for each participant indicated main effects of age, $F(1, 33) = 6.88$, $p < .01$ and talker, $F(1, 33) = 14.15$, $p < .01$, (see Table 2). Slopes decreased with age indicating that children matched the stimulus duration to a greater degree than did adults. In fact, the regression slope for 4-year-olds in the single-talker condition indicated a near one-to-one match (slope = .95). In addition, participants at all ages more closely matched stimuli (had higher slopes) in the single-talker condition than in the multiple-talker condition. What does this difference imply for talker normalization? One possibility is that the normalization process is engaged only by, or to a greater degree by, multiple talkers, which results in a decreased match between the stimulus and response duration (assuming a response is generated from the abstracted phonetic representation). Because there is only a single voice that listeners must deal with in the single-talker condition, the normalization process may be unnecessary (after a few trials). Alternatively, perhaps when listeners are repeatedly presented with the same voice, they are able to access the phonetic content without stripping away as much speaker-specific information as when they are presented with a new voice on every trial (and that this speaker-specific information is incorporated into the response). Either possibility would result in a closer match to the physical stimulus in the single-talker condition. This prediction—that the match between stimulus and response duration should decrease the more normalization processes are engaged—needs to be explored more fully, especially with adults. Normalization is generally thought of as an all-or-none process, and the possibility raised here—that normalization processes may be engaged to differing degrees depending on task variables such as familiarity with a voice—needs to be examined.

In summary, the findings in this experiment indicated that in a naming paradigm, 4-year-olds were slower to respond than were 5-year-olds and adults, who did not differ. Both children and adults were slower to respond to items from a multiple-talker list than to items from a single-talker list. However, latency to respond did not decrease substantially with development (i.e., less than 100 ms between the ages of 4 and 5 years). Rather, the major developmental change was in the duration of the response. That is, the match between the duration of the stimulus and the participant's response decreased with age.

General Discussion

In a series of three experiments, we found that young children's word recognition performance is negatively affected by talker variability both under optimal conditions (words presented in the clear) and under conditions of stimulus degradation (words presented in noise). We also found that, with development, children's ability to process multiple talkers improves relative to processing a single talker—that is, talker normalization develops. In our final experiment, we demonstrated that both children and adults are slower to repeat words from a list spoken by multiple talkers than they are with words from a list spoken by a single talker. Additionally, younger children matched the duration of the stimulus words to a greater degree than did older children and adults, indicating that more speaker-specific information may be retained in young children's representation of speech.

Overall, these results are consistent with the idea that preschool children get better at stripping away or ignoring irrelevant variability as they become more linguistically advanced. From these results, two developmental accounts may be proposed.

The first account consistent with these findings is the traditional view wherein there is a discrete stage of talker normalization (Studdert-Kennedy, 1974). Children, in the course of learning to cope with talker variability, could be developing a stage of processing in which speaker-independent phonemes are abstracted from the auditory signal. This ability or stage of processing, although present to some degree even in infancy, would continue to emerge with the development of adultlike speech perception processes. By adulthood, this discrete stage of normalization would operate prior to lexical access and would result in representations stripped of, or corrected for, speaker-specific information. The question remains as to whether this change with development would be merely a matter of degree or whether the acquisition of a stage of normalization would constitute a qualitative difference in the processing of different talkers by children and adults. In addition, it should be noted that this account has difficulty accounting for the adult data in Experiment 3 and for some of the results discussed in the introduction.

A second possible account consistent with these results differs from the first only in the degree to which a discrete stage of normalization actually develops. Perhaps talker-specific information is never actually eliminated from the representation of speech but always exists to some degree, even in adults. Research has recently shown with adults that recognition of spoken words leads to the creation and storage of detailed perceptual traces that affect later perception and recognition in explicit and implicit memory tasks (Goldinger, 1992; Schacter & Church, 1992; Nygaard, Sommers, & Pisoni, 1994; Palmeri, Goldinger, & Pisoni, 1993; Pisoni, 1993). These findings, among others, have been used to argue against the traditional assumption of a discrete stage of normalization in which speaker-specific information is completely removed from the signal (Goldinger, 1992; Nygaard et al., 1994; Pisoni, 1990). If a discrete stage of talker normalization never develops, the difference between children and adults would remain a quantitative one. This proposal is compatible with the talker variability effects that were found in this investigation, would account for the match in production that was found in Experiment 3, and is consistent with the recent adult evidence for the retention and influence of specific exemplars in later perception (Goldinger, 1992; Nygaard et al., 1994; Palmeri et al., 1993; Pisoni, 1993; Schacter & Church, 1992). In addition, it is reasonable that children would retain more information about the acoustic (or articulatory) structure of the utterance than would adults (as was found in Experiment 3), as this information may be important for learning how to produce speech.

In conclusion, the work presented here is an important first step. Although it makes a significant contribution to our understanding of the effects of talker variability and its effects across development, a great many questions remain unanswered. For example, this work does not directly address the mechanisms or processes responsible for children’s increasing ability to cope with talker variability. If children are learning to strip away variability, are the increases item-specific (word by word) or more general, across the board increases? Does change across development constitute a quantitative or qualitative change in processing? In addition to questions regarding the mechanisms and processes of speech perception, more global questions concerning children’s acquisition of language in real-world contexts remain to be addressed. Work has begun that addresses the effects of talker variability on children with speech–language impairments (see Forrest, Chin, Pisoni, & Barlow, 1994, for a preliminary report), but many other questions have received no attention. For example, our first experiment showed that young children’s performance can be reduced by talker variability even under optimal listening conditions. Basic laboratory research such as this alerts us to the possibility of effects of talker variability in other, less optimal, real-world contexts but does not provide us with any answers. For example, what are the implications regarding children's social context and exposure to different voices—do children who attend day care have an advantage over home-care children (who presumably hear fewer voices)?

These kinds of questions, with regard to preschool children, have largely been ignored but are critical to a developmentally complete understanding of coping with talker variability. In addition, the answers to such questions may have implications for understanding developmental achievements in many areas such as phonological development, speech perception, and language acquisition in general. Given that the traditional assumptions about stimulus variability and talker normalization have recently begun to be seriously questioned, information about the development of such processes is important not only with regard to understanding the development of speech perception and language acquisition but also with regard to general theories of talker normalization, speech perception, and the representation of spoken words in the mental lexicon.

Acknowledgments

This work was supported in part by National Institutes of Health Research Grant DC-0111-14 and National Institutes of Health Training Grant DC-00012-12. We would like to thank Linda B. Smith for her help in the planning and completion of this project.

References

Best, CT. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In: Goodman, JC.; Nusbaum, HC., editors. The Development of Speech Perception. Cambridge, MA: MIT Press; 1994.

Carrell, TD. Contributions of fundamental frequency, formant spacing, and glottal waveform to talker identification. Unpublished doctoral dissertation. Bloomington: Indiana University; 1984.

Creelman CD. Case of the unknown talker. Journal of the Acoustical Society of America. 1957; 29:655.

DeCasper AJ, Fifer WP. Of human bonding: Newborns prefer their mothers' voices. Science. 1980 Jun 6.208:1174–1176. [PubMed: 7375928]

Forrest, K.; Chin, SB.; Pisoni, DB.; Barlow, N. Research on Spoken Language Processing: Technical Report No. 19. Bloomington, IN: Speech Research Laboratory; 1994. Talker normalization in normally articulating and phonologically delayed children: Methodological considerations; p. 229-251.

Goldinger, S. Research on Speech Perception Technical Report No. 7. Bloomington, IN: Speech Research Laboratory; 1992. Words and voices: Implicit and explicit memory for spoken words; p. 1-128.

Goldinger S, Pisoni DB, Logan J. On the nature of talker variability effects on recall of spoken word lists. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1991; 17:152–162.

Joos MA. Acoustic phonetics. Language. 1948; 24(Suppl. 2):1–136.

Jusczyk, P. Developing phonological categories from the speech signal. In: Ferguson, CA.; Menn, L.; Stoel-Gammon, C., editors. Phonological development: Models, research, and implications. Timonium, MD: York Press; 1993.

Jusczyk, P. Infant speech perception and the development of the mental lexicon. In: Nusbaum, HC.; Goodman, JC., editors. The transition from speech sounds to spoken words: The development of speech perception. Cambridge, MA: MIT Press; 1994.

Jusczyk P, Pisoni DB, Mullenix J. Effects of talker variability on speech perception by 2-month-old infants. Cognition. 1992; 43:253–291. [PubMed: 1643815]

Klatt, DH. The problem of variability in speech recognition and in models of speech perception. In: Perkell, JS.; Klatt, DH., editors. Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum; 1986.

Krulee GK, Tondo DK, Wightman FL. Speech perception as a multilevel processing system. Journal of Psycholinguistic Research. 1983; 12:531–554.

Kucera, E.; Francis, W. Computational analysis of present-day American English. Providence, RI: Brown University Press; 1967.

Kuhl PK. Speech perception in early infancy: Perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America. 1979; 66:1668–1679. [PubMed: 521551]

Kuhl PK. Perception of auditory equivalence classes for speech in early infancy. Infant Behavior and Development. 1983; 6:263–285.

Ladefoged P. What are linguistic sounds made of? Language. 1980; 56:485–502.

Luce, PA.; Carrell, TD. Research on Speech Perception Progress Report No. 7. Bloomington, IN: Speech Research Laboratory; 1981. Creating and editing waveforms using WAVES.

Martin C, Mullennix J, Pisoni DB, Sommers M. Effects of talker variability on recall of spoken word lists. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1989; 15:676–684.

McClelland JL. On the time relations of mental processes: An examination of systems of processes in cascade. Psychological Review. 1979; 86:287–330.

Mullennix J, Pisoni DB, Martin C. Some effects of talker variability on spoken word recognition. Journal of the Acoustical Society of America. 1989; 85:365–378. [PubMed: 2921419]

Nusbaum, H.; Pisoni, DB.; Davis, S. Research on Speech Perception Progress Report No. 10. Bloomington: Speech Research Laboratory, Indiana University; 1984. Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words.

Nygaard LC, Sommers MS, Pisoni DB. Speech-perception as a talker-contingent process. Psychological Science. 1994; 5:42–46. [PubMed: 21526138]

Palmeri TJ, Goldinger SD, Pisoni DB. Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology: Learning, Memory and Cognition. 1993; 19:1–20.

Peterson GH, Barney HL. Control methods used in a study of the vowels. Journal of the Acoustical Society of America. 1952; 24:175–184.

Pisoni, DB. Effects of talker variability on speech perception: Implications for current research and theory. In: Fujisaki, H., editor. Proceedings of the 1990 International Conference on Spoken Language Processing; Kobe, Japan: Acoustical Society of Japan; 1990. p. 1399-1407.

Pisoni DB. Long-term memory in speech perception: Some new findings on talker variability, speaking rate and perceptual learning. Speech Communication. 1993; 13:109–125. [PubMed: 21461185]

Ross M, Lerman J. A picture identification test for hearing impaired children. Journal of Speech and Hearing Research. 1970; 13:44–53. [PubMed: 4192711]

Schacter DL, Church BA. Auditory priming: Implicit and explicit memory for words and voices. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1992; 18:915–930.

Smith B. Relationships between duration and temporal variability in children’s speech. Journal of the Acoustical Society of America. 1992; 91:2165–2174. [PubMed: 1597607]

Studdert-Kennedy, M. The perception of speech. In: Sebeok, TA., editor. Current trends in linguistics. The Hague, The Netherlands: Mouton; 1974. p. 2349-2385.

Studdert-Kennedy, M. Speech perception. In: Lass, NJ., editor. Contemporary issues in experimental phonetics. New York: Academic Press; 1976. p. 243-293.

Walley AC. Spoken word recognition by young children and adults. Cognitive Development. 1988; 3:137–165.

Werker, JF. Cross-language speech perception: Development change does not involve loss. In: Goodman, JC.; Nusbaum, HC., editors. The development of speech perception. Cambridge, MA: MIT Press; 1994. p. 93-120.

Appendix

Word Lists From the Word Identification by Picture Intelligibility (WIPI) Test

List 1	List 2	List 3
School	Broom	Moon
Ball	Bowl	Bell
Smoke	Coat	Coke
Floor	Door	Corn
Fox	Socks	Box
Hat	Flag	Bag
Man	Fan	Can
Bread	Red	Thread
Neck	Desk	Nest
Stair	Bear	Chair
Eye	Pie	Fly
Knee	Tea	Key
Street	Meat	Feet
Wing	String	Spring
Mouse	Clown	Crown
Shirt	Church	Dirt
Gun	Thumb	Sun
Bus	Rug	Cup
Train	Cake	Snake
Arm	Barn	Car
Chick	Stick	Dish
Crib	Ship	Bib
Wheel	Seal	Queen
Straw	Dog	Saw
Pail	Nail	Jail

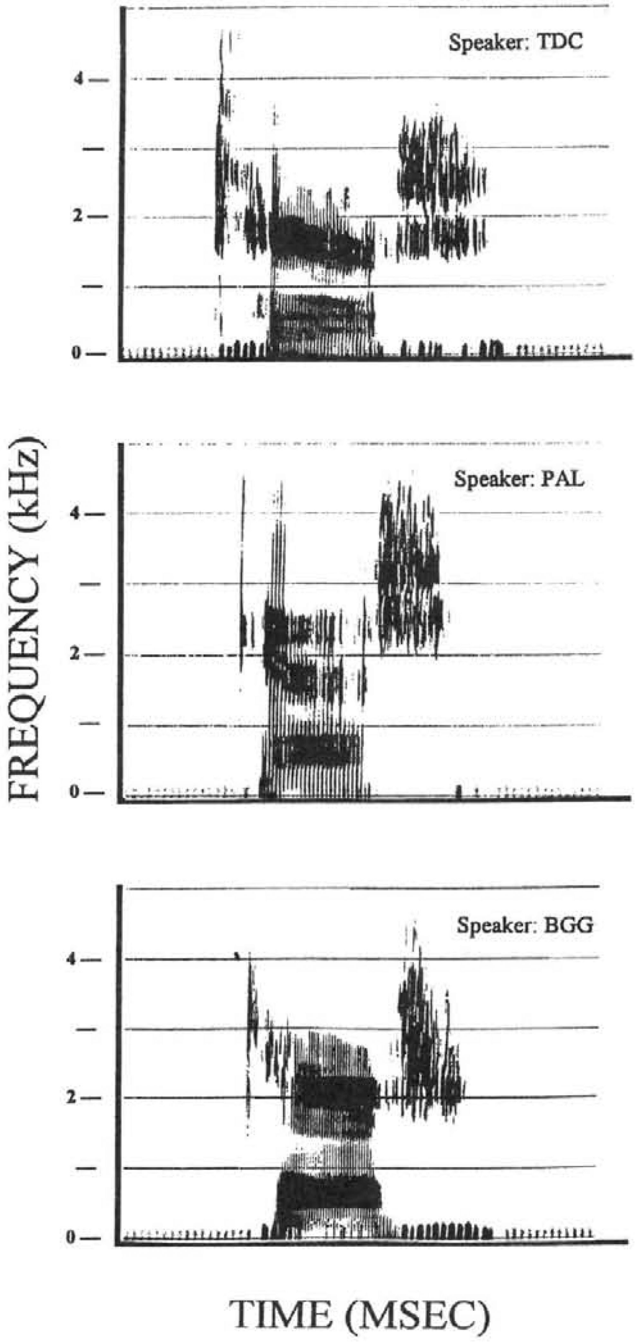

Figure 1.
Spectrograms of three adult speakers (TDC, PAL, BGG) uttering the word “cash.”

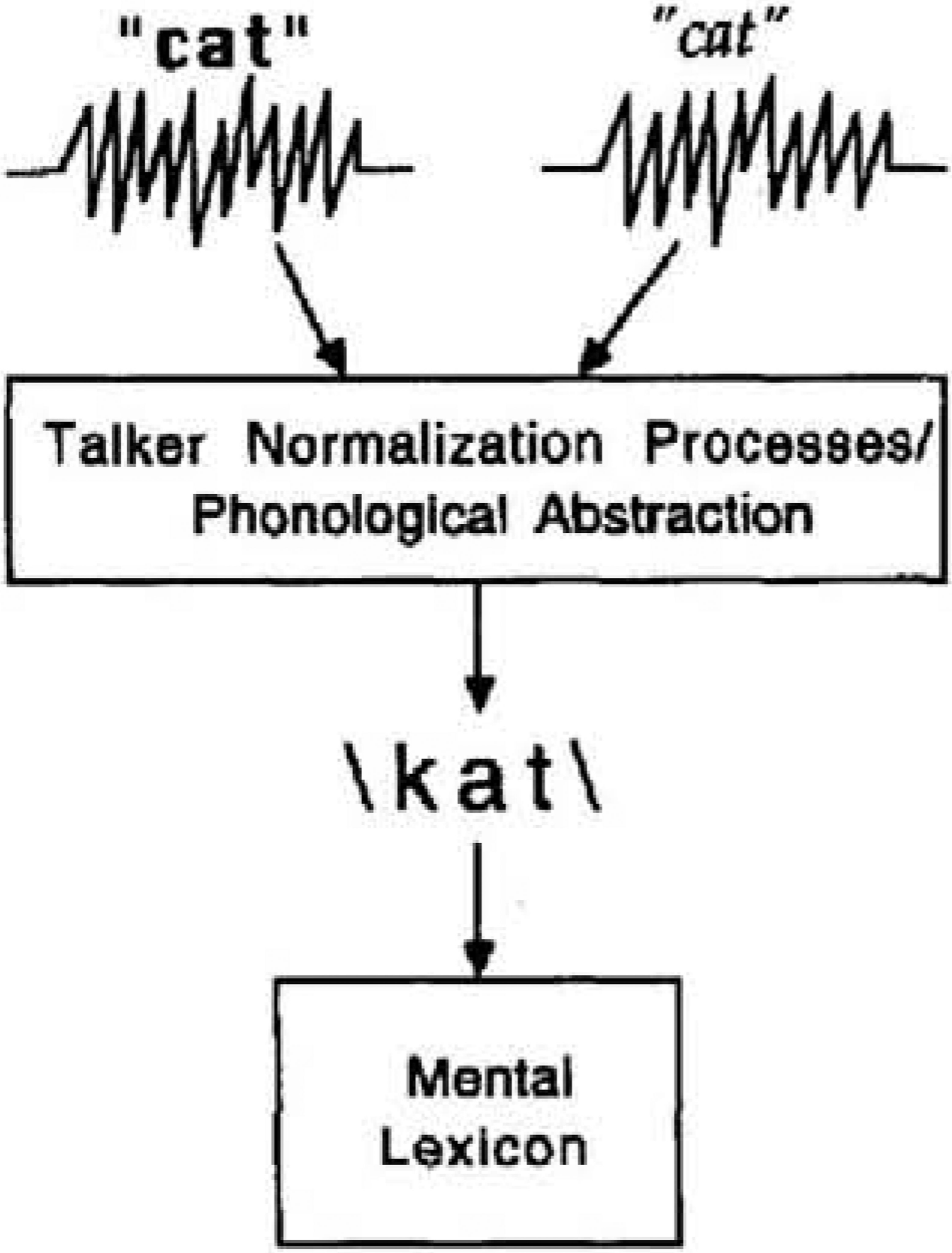

Figure 2.
A simplified model of the process of talker normalization.

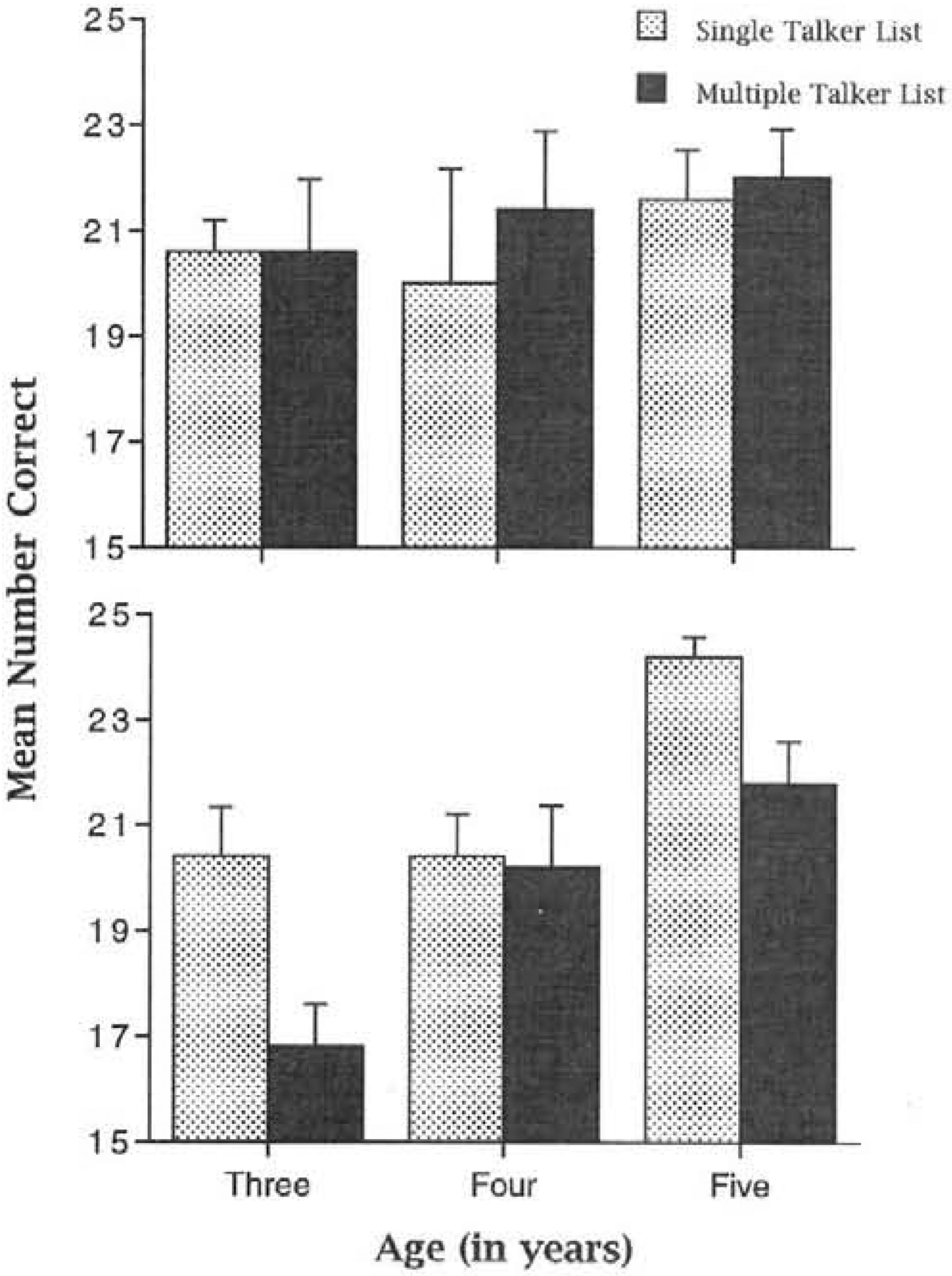

Figure 3.

(Top) Single-talker list first. (Bottom) Multiple-talker list first. Experiment 1: Mean number of words correct by age, talker condition, and list order. Error bars represent the standard errors of the means.

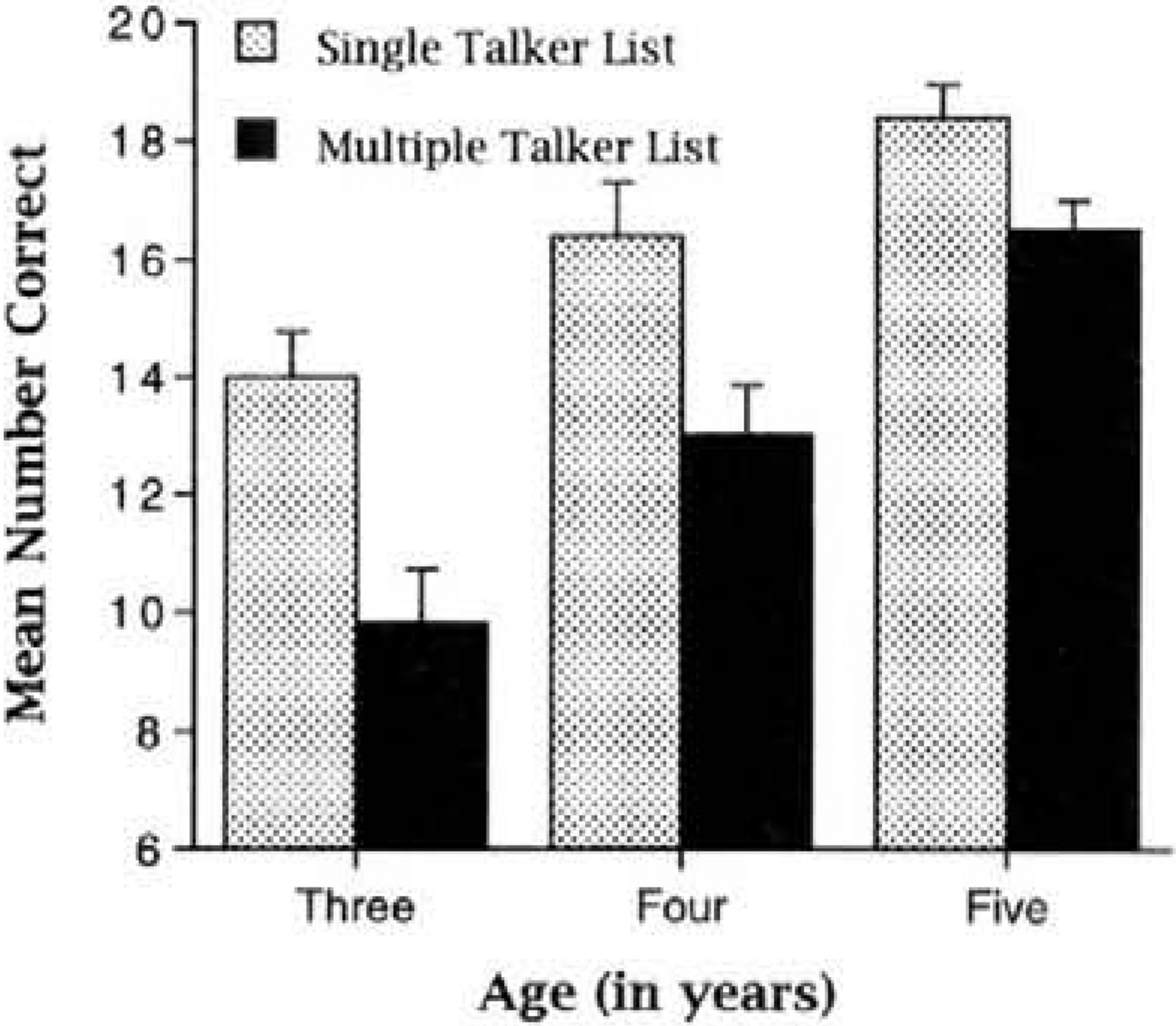

Figure 4.
Experiment 2: Mean number of words correct by age and talker condition. Error bars represent the standard errors of the means.

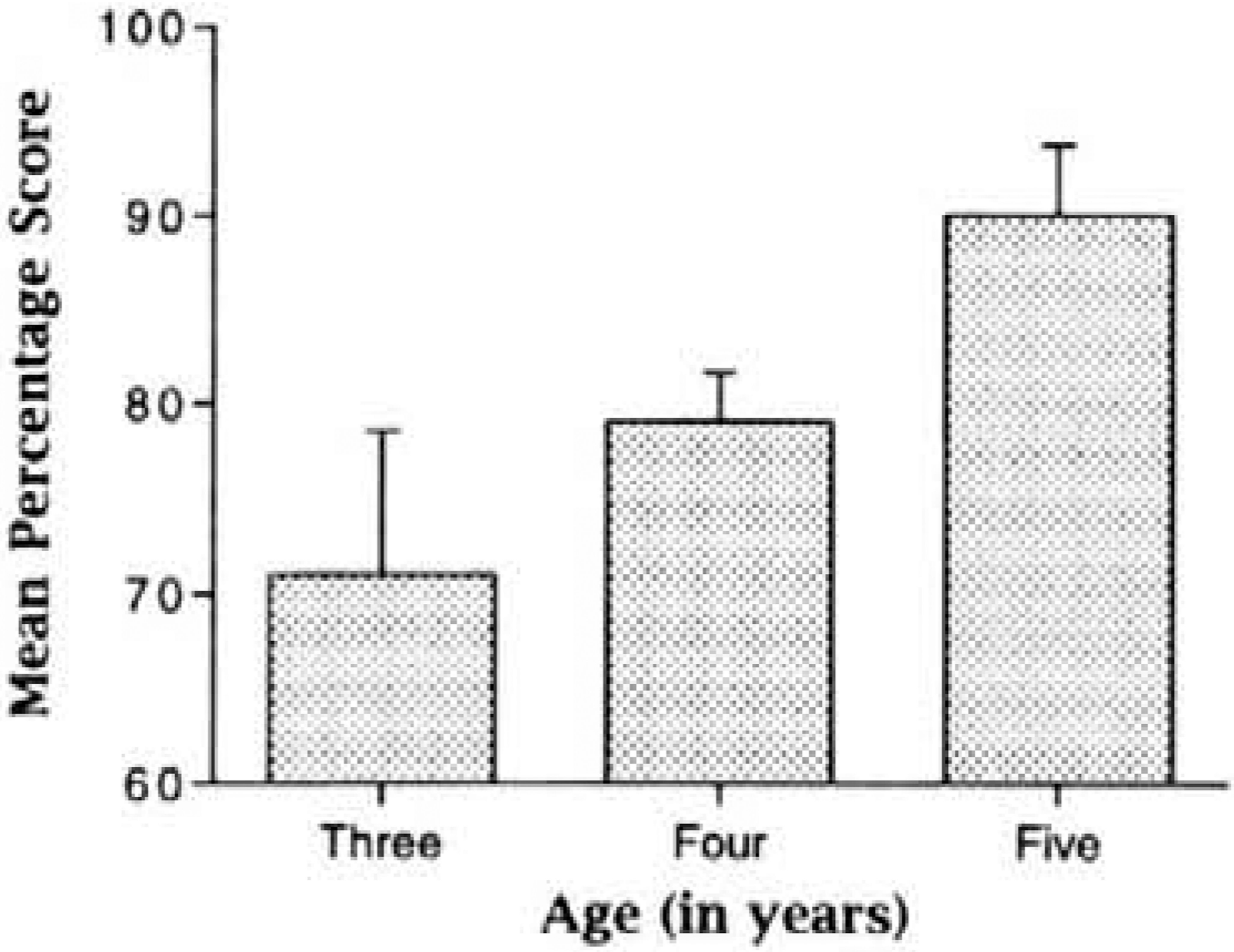

Figure 5.
Experiment 2: Mean percentage scores presented by age. Error bars represent the standard errors of the means.

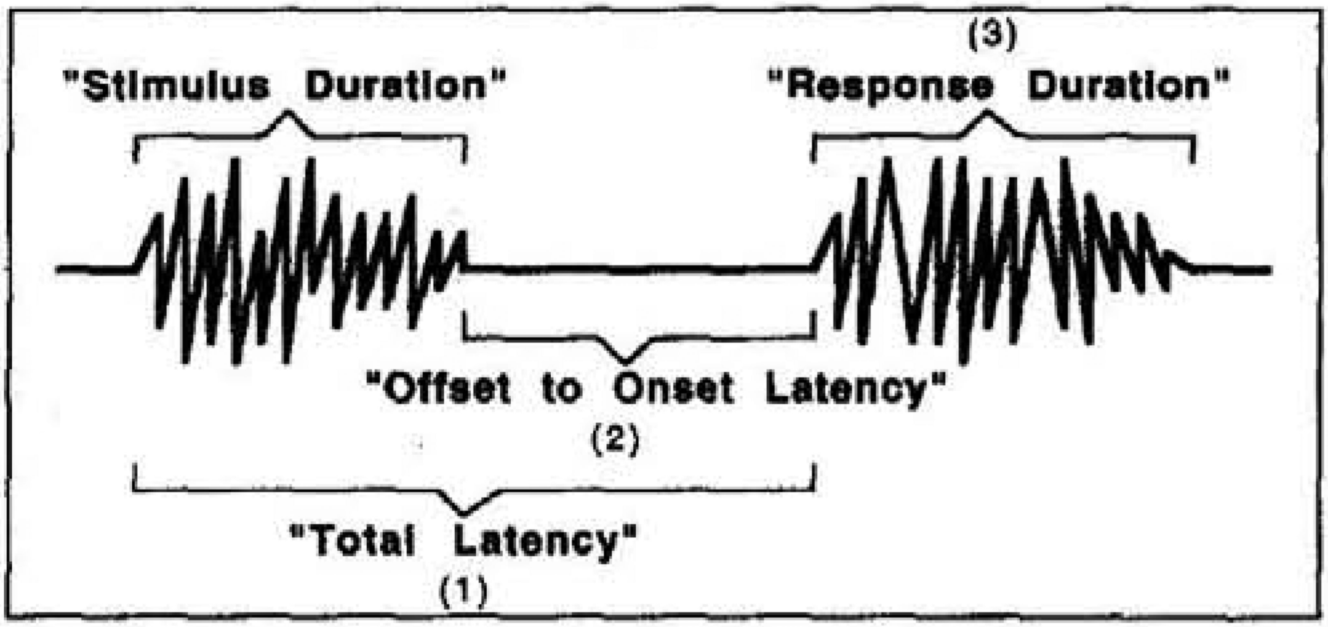

Figure 6.
Dependent measures analyzed in Experiment 3.

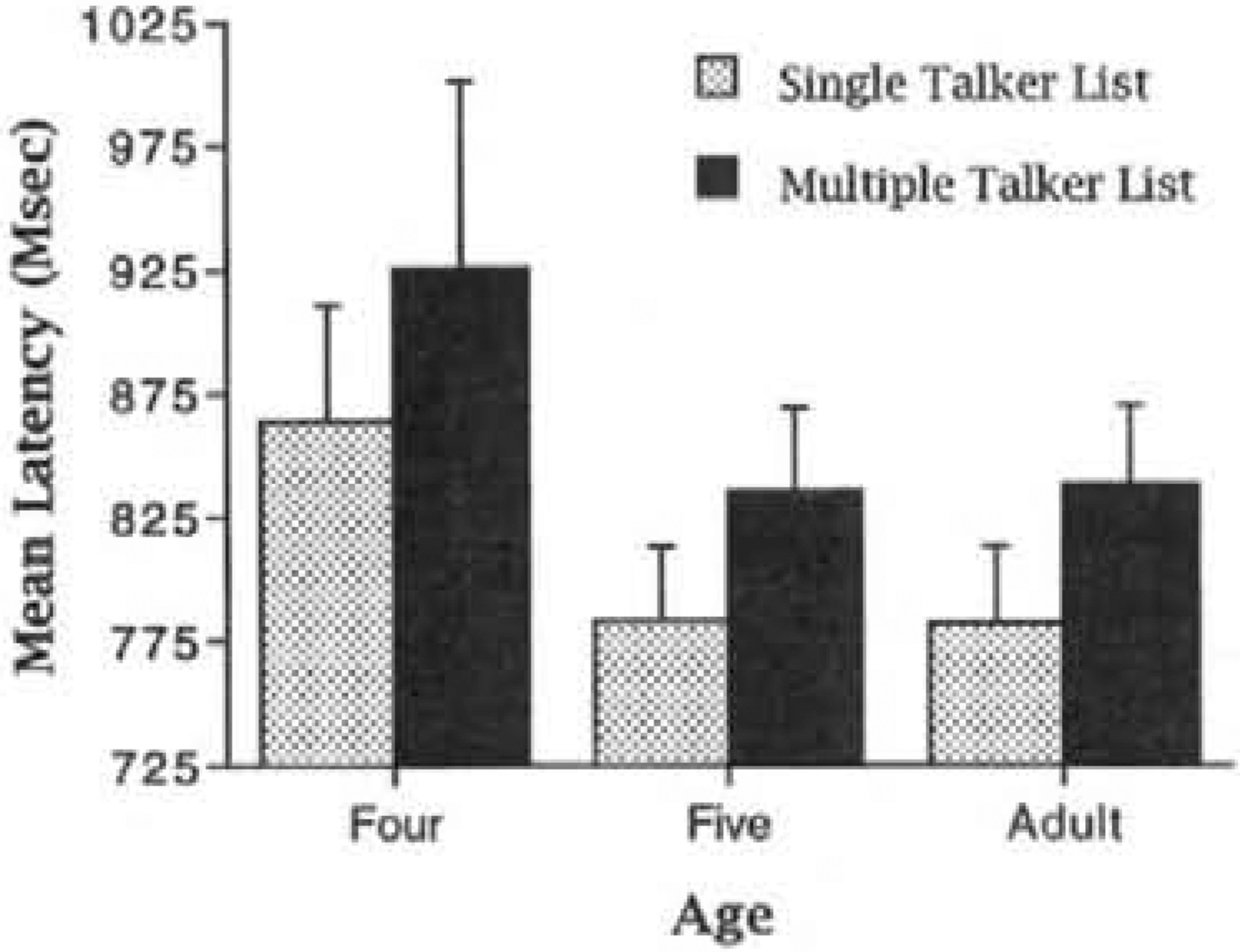

Figure 7.
Experiment 3: Total latency by age and talker condition. Error bars represent the standard errors of the means.

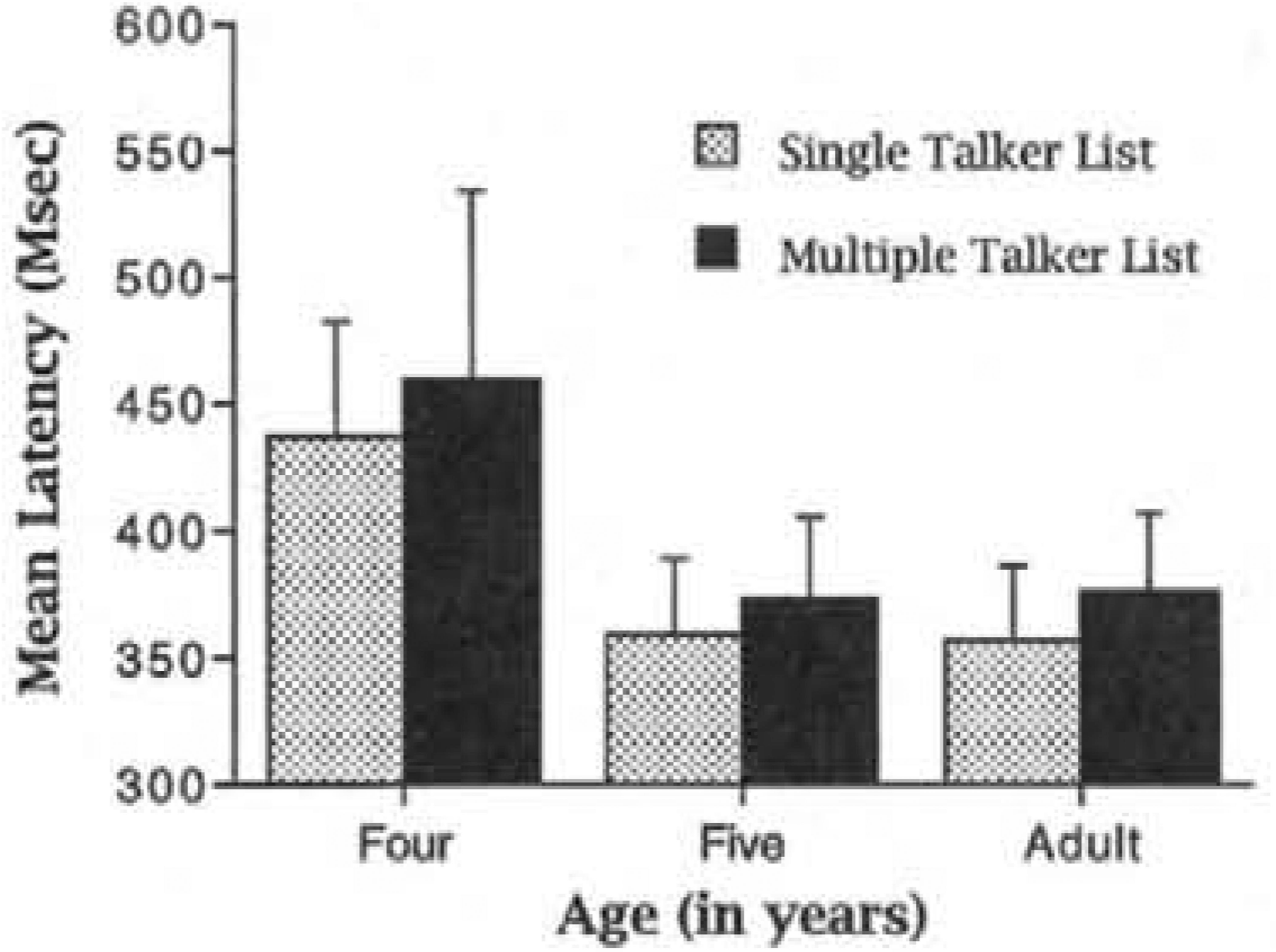

Figure 8.
Experiment 3: Offset-to-onset latency by age and talker condition. Error bars represent the standard errors of the means.

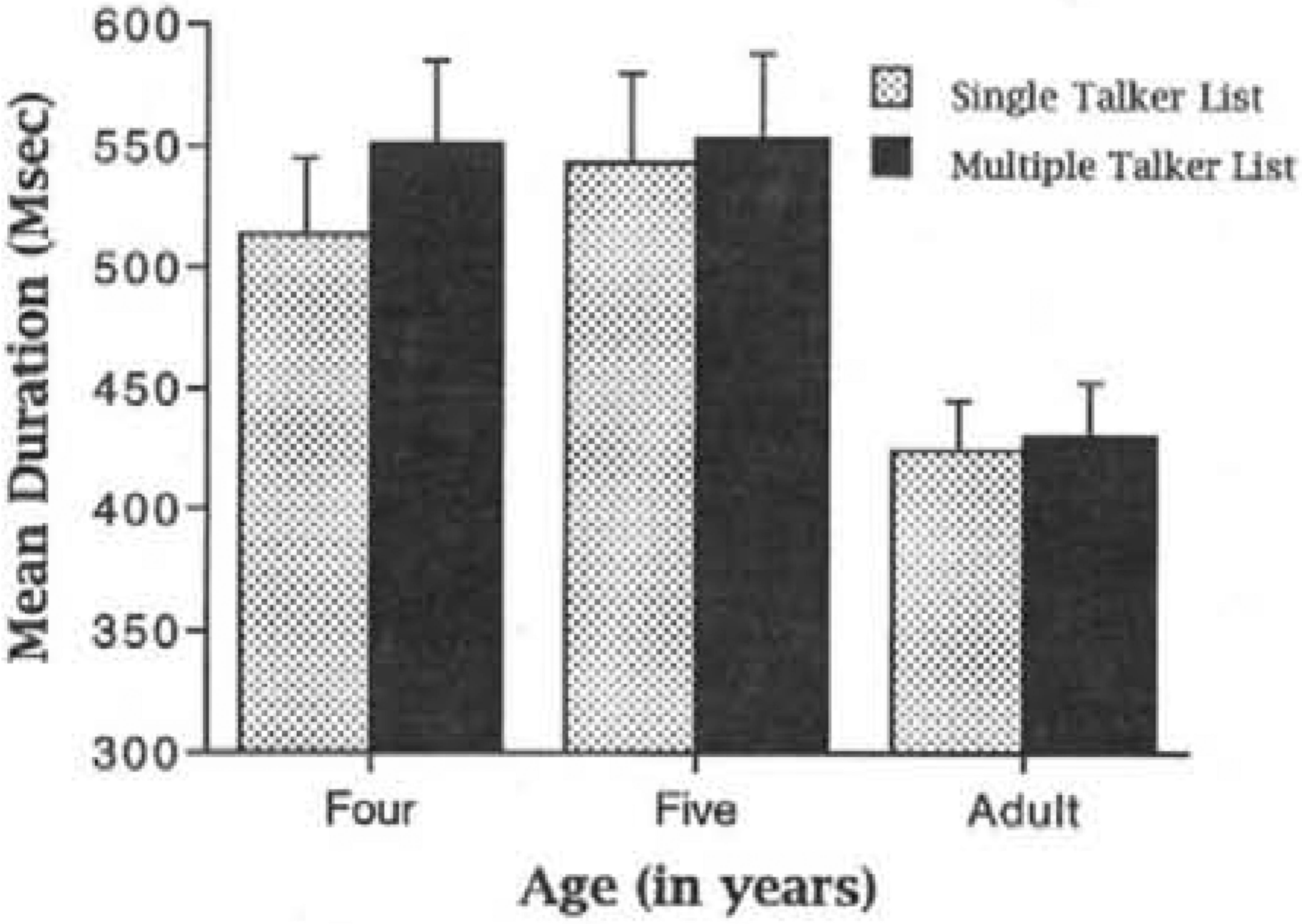

Figure 9.
Experiment 3: Response duration by age and talker condition. Error bars represent the standard errors of the means.

Table 1

Experiment 3: Mean Correct Responses and Standard Deviations by Age and Talker Condition

	Age		
Talker	**4 years**	**5 years**	**Adult**
Single	23.3 (1.74)	24.1 (2.23)	24.7 (1.03)
Multiple	22.6 (3.45)	24.2 (2.86)	24.5 (1.94)

Note. Standard deviations are presented in parentheses.

Table 2

Experiment 3: Regression Analysis of Response Duration on Stimulus Duration

	Age					
	4-year		5-year		Adult	
Regression analysis	**Single talker**	**Multiple talker**	**Single talker**	**Multiple talker**	**Single talker**	**Multiple talker**
R^2	.56	.57	.44	.38	.56	.32
Slope	.95	.78	.89	.58	.67	.50
Y intercept	110	185	167	285	139	199