CHILDREN'S PERCEPTION OF SPEECH IN NOISE

by

Marianne Fallon

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Psychology
University of Toronto

© by Marianne Fallon, 2001
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.
Children’s perception of speech in noise

Doctor of Philosophy, 2001

Marianne Fallon

Department of Psychology

University of Toronto

Abstract

The purpose of the present investigation was to compare children’s and adults’ identification of speech in a background of multitalker babble. The question of particular interest was whether noise disproportionately impairs 5-year-old children’s ability to integrate acoustic and linguistic information, to focus on word-initial information, or to benefit from semantic cues. Chapter 2 presents an experiment in which children 5, 9, and 11 years of age and young adults identified the final word of simple sentences, which were presented in two levels of multitalker babble. Participants responded by selecting one of four pictures. Although children required more favorable signal-to-noise ratios (SNRs) than adults to achieve comparable performance, an equivalent decrease in SNR had comparable consequences for all age groups, indicating that sensory differences accounted for the age differences observed on this task. In the experiment presented in Chapter 3, children aged 5 and 9 and adults identified monosyllabic target words from choices that included phonologically dissimilar foils or foils differing only in their initial or final phoneme. The 5-year-olds experienced disproportionate difficulty on trials with final contrasts. They also performed more poorly than expected on trials with maximal contrasts. On trials with initial contrasts, however, 5-year-olds and adults performed comparably. Chapter 4 documents the attempts of children 5 and 9 years of age and adults to identify the final words of low- and high-context sentences in two levels of babble. Although SNRs were adjusted to equalize differences in stimulus audibility across age, 5-year-olds still performed more poorly than did 9-year-olds and adults. Listeners of all ages, however, showed comparable gains from
context. In sum, when cognitive and linguistic demands of a task are minimized and stimulus audibility is equated across age, children’s and adults’ speech recognition in noise is remarkably similar. When the demands are increased, however, age-related differences emerge. Nevertheless, noise does not impair 5-year-olds’ ability to attend to word-initial information or to profit from semantic contextual cues, both of which make important contributions to speech comprehension. In short, the present findings are inconsistent with the widely held view that noise has more disruptive effects on children than on adults.
Acknowledgements


I wish to thank the members of my committee, Bruce Schneider, Meredyth Daneman, and, in particular, my advisor Sandra Trehub for their support, intellectual insight, and the desire to broaden avenues of thinking and communicating. Special appreciation is extended to Richard Aslin for serving as a thoughtful and insightful external appraiser and examiner.

These projects are a testament to the dedication and exceptional organizational skills of the Infant and Child Studies staff: Marilyn Barras, Elizabeth Olszewska, and Loretta Falco. Without your help, my head would be disconnected from my shoulders.

For their technical assistance, I would like to recognize Peter Mui and Andrew Veglio, with a special nod to Kim Yue for his invaluable expertise in computer programming.

To good friends, I recognize and appreciate all you have given me during this adventure, and I hope that I can repay the favor. To my family, your long-distance encouragement and support has been an invaluable asset. Finally, to my husband, our paths crossed and diverged midstream, but you never left my side.
# Table of Contents

Chapter 1: General Introduction .......................... 1

Chapter 2: Children’s perception of speech in multitalker babble .......................... 10
   Introduction ............................................. 10
   Method ................................................. 14
   Results ................................................. 21
   Discussion ............................................. 27

Chapter 3: Children’s identification of minimally contrastive words in noise .......................... 30
   Introduction ............................................. 30
   Method ................................................. 33
   Results ................................................. 39
   Discussion ............................................. 43

Chapter 4: Children’s use of contextual cues in degraded listening environments .......................... 49
   Introduction ............................................. 49
   Method ................................................. 52
   Results ................................................. 57
   Discussion ............................................. 61

Chapter 5: General Discussion .......................... 64

Appendix .................................................. 70

References .............................................. 73
List of Tables

2.1 Percent correct performance across age, noise level, and order of presentation. 23
2.2 Percentage correct performance per half in low noise. 25
2.3 Average difficulty of targets in descending rank order. 26
2.4 Spearman correlation coefficients relating overall average ranking with rankings of individual age groups. 27
3.1 Word frequencies for initial-, final-, and maximal-contrast trials in log units using three measures of familiarity. 35
3.2 Fundamental frequency and duration for initial-, final-, and maximal-contrast sentences. 36
3.3 Performance (percent correct) of 5-year-olds, 9-year-olds, and adults on maximal-contrast, initial-contrast, and final-contrast trials. 41
3.4 Percentage of errors committed by selecting a similar-sounding foil. 42
3.5 The likelihood (%) of selecting a target or its minimally contrastive foil. 42
3.6 Age-related differences in performance (percent correct) on the maximally contrastive block of trials as a function of when the block occurred. 47
4.1 Fundamental frequency and duration of low- and high-context sentences. 55
4.2 Performance (percent correct) of 5-year-olds, 9-year-olds, and adults on low- and high-context sentences in both levels of noise. 59
4.3 Average latencies in seconds for low- and high-context sentences as a function of age. 61
List of Figures

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Averaged power spectra of the babble used in the present study and the male speaker of the SPIN sentences.</td>
<td>17</td>
</tr>
<tr>
<td>2.2</td>
<td>Averaged power spectra of the babble and the female speaker from the present study.</td>
<td>18</td>
</tr>
<tr>
<td>2.3</td>
<td>Percent correct identification as a function of noise level and age.</td>
<td>22</td>
</tr>
<tr>
<td>2.4</td>
<td>Percent correct identification in high noise as a function of age and order.</td>
<td>24</td>
</tr>
<tr>
<td>3.1</td>
<td>Performance of 5-year-olds, 9-year-olds, and adults on maximal-, initial-, and final-contrast trials.</td>
<td>40</td>
</tr>
<tr>
<td>4.1</td>
<td>Performance of 5-year-olds, 9-year-olds, and adults on low- and high-context sentences in both levels of noise.</td>
<td>58</td>
</tr>
</tbody>
</table>
Children's perception of speech in noise

Chapter 1

General Introduction

The comprehension of speech depends upon the translation of acoustic-phonetic input into linguistically meaningful information. Although rudimentary speech processing (e.g., phoneme discrimination) is evident in early infancy (Eimas, Siqueland, Juszczyk, & Vigorito, 1971), such processing occurs without regard to meaning (Werker, Cohen, Lloyd, Casasola, & Stager, 1998). By contrast, the processing of speech as meaningful material exhibits a protracted developmental course. For example, the identification of words from incomplete information (e.g., *nood__ ; _oodle*) does not become adult-like until the teen years (Elliott, Hammer, & Evans, 1987). This extended developmental timetable may be a by-product of gradual changes in perceptual and/or cognitive processing.

In optimal conditions, children’s speech recognition may appear to be much like that of adults. Degraded conditions, however, reveal substantial discrepancies between young children and adults (e.g., Elliott, 1979; Elliott et al., 1979; Mills, 1975; Nittrouer & Boothroyd, 1990). The suggestion is that noise not only functions as a perceptual masker but also as a significant distractor, one that has disproportionate consequences for young listeners. In addition, specific linguistic factors such as word frequency or semantic cues are thought to have differential consequences for children and adults, especially in suboptimal listening environments (e.g., Elliott, 1979; Elliott, Clifton & Servi, 1983; Nittrouer & Boothroyd, 1990). Careful selection of noise levels would minimize age-related differences in stimulus audibility at early stages of auditory processing, which would permit an examination of perceptual and/or cognitive factors at later stages of processing.

Because the available research on children’s perception of speech in noise is limited, it is useful to consider what is known about the development of auditory word recognition in general. In so doing, one can consider Cole and Jakimik’s (1980) assumptions about the
processes involved. First, word recognition requires the integration of auditory cues with knowledge of language and its uses. Second, word-initial sounds (e.g., *noo*) typically activate several potential word candidates (e.g., *new, news, newt, noodle, noogie, pneumonia*, etc.), with further information required for the specification of a single word (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987). Third, words in fluent speech are recognized sequentially. In other words, the identification of one word in a sentence provides syntactic and semantic cues that facilitate the recognition of the next word.

Returning to the first assumption, it is useful to consider when children begin imbuing elements of the speech signal with linguistic relevance. Although 1- to 4-month-old infants exhibit categorical perception of phonemes (Eimas et al., 1971), there are indications that this skill is not speech-specific. For example, chinchillas and macaques exhibit adult-like categorization of phonemes (Kuhl & Miller, 1975; Kuhl & Padden, 1982), and it is likely that other mammals do likewise (Kuhl, 1987). Speech sounds begin to assume linguistic relevance late in the first year of life, when phoneme discrimination shows influences from the native language environment (e.g., Kuhl, 1992; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Polka, 1993; Werker & Tees, 1984) and infants associate some speech sounds with objects and events in their environment (e.g., Tincoff & Jusczyk, 1999). However, skills that are evident in simple phoneme discrimination tasks are not readily transferable to word learning tasks until at least the second year of life (e.g., Garnica, 1973; Werker et al., 1998).

Aside from learning which sounds "belong" to their native language, young children must learn about phonotactic constraints, or permissible sound sequences. Rudimentary signs of such constraints are apparent by 8 months of age. For example, 2 minutes of familiarization with a continuous string of consonant-vowel syllables is sufficient for 8-month-old infants to distinguish more frequent from less frequent three-syllable sequences within the syllable stream (Saffran, Aslin, & Newport, 1996). It is clear, then, that infants can extract some statistical properties of the input from relatively brief exposure. By 9
months of age, infants' sensitivity to the phonotactic features of their native language is evident in their listening preference for phonotactically "legal" over "illegal" sequences (Friederici & Wessels, 1993). Presumably, the nature of phonotactic knowledge changes once the sounds of a language become meaningful. Children 4 years of age reveal implicit understanding of the phonotactic properties of their native language (Cole, 1981) by superior detection of mispronunciations in "illegal" sequences (e.g., shlocked) than in permissible combinations (e.g., slocked).

Effective processing of one's native language also depends upon accurate representation of its sounds. Children's discrimination of words that differ in a single phoneme (i.e., minimal pairs) can shed light on their representations of speech sounds. Although 2-year-olds can differentiate words contrasting in their initial sound (e.g., bear, pear) at better than chance levels (Barton, 1980), errors persist through the preschool years, particularly for final-phoneme contrasts, (e.g., cat, cap) (Graham & House, 1971; Higgs & Hodson, 1978; Menary, Trehub, & McNutt, 1982; Morgan, 1984). These difficulties imply that young children's phonemic representations differ in some respects from those of adults. Increasing exposure to adult forms seems to promote the construction of mature phonological representations (Vihman, 1996).

Children's production of speech provides another window on their processing of speech sounds (Eilers & Oller, 1976). For one thing, their production of particular sounds is likely to involve internal representations that are common to the perception of those sounds. In the preschool period, children typically lack awareness of their mispronunciations (Berko & Brown, 1960), which indicates the absence of self-monitoring. Mispronunciations, such as fis for fish, continue into the early school years (Anisfeld, 1984). The perception of phonemic contrasts is well in advance of production (Zlatin & Koenigsknecht, 1976), with poor production frequently resulting from articulatory difficulty rather than underspecified phonemic representations (Velleman, 1988). The prevailing wisdom is that phonemic
representations are largely mature by the early school years (Morgan, 1984), with notable contributions from the acquisition of reading (Morais, Alegria, & Content, 1987).

Returning to Cole and Jakimik's (1980) second assumption, it is unclear when children begin attending to word-initial information as a cue to lexical identity. What is clear, however, is that young children require more acoustic information to identify isolated words than do older children and adults (Elliott et al., 1987; Walley, 1988). This finding is attributable, in part, to developmental differences in phonetic decoding strategies. Whereas older children and adults are thought to decompose words into phonetic segments, younger children are presumed to use more global or holistic strategies (Jusczyk, 1997; Metsala, 1997; Treiman & Breaux, 1982; Walley, 1993). Global recognition strategies imply approximately equal distribution of attention to acoustic cues throughout a word. By contrast, phonetic strategies involve differential attention to specific portions of a word, such as its onset. Such flexibility of focus is thought to facilitate word identification, especially when acoustic information is incomplete or degraded (Walley, 1988).

Explicit awareness of phonemes is thought to coincide with emerging literacy in the early school years (e.g., Morais et al., 1987). Phonemic awareness is preceded by syllabic awareness, with the syllable serving as the predominant unit of analysis (Liberman, Shankweiler, Fischer, & Carter, 1974; Treiman, 1985a). Indeed, preschoolers can link their actions to syllables but not to phonemes (Liberman et al., 1974). Nevertheless, there are indications that preschoolers can focus on some discrete parts of words, notably the onset and the rime (Treiman, 1985a). For example, preschoolers recognize rhyming words (Knaflé, 1973, 1974; Lenel & Cantor, 1981), which involves separation of the onset (boat) from the rime (boat). Attention to word onset is also evident in 4- and 5-year-olds' superior discrimination of minimal pairs differing in word-initial compared to word-final position (Graham & House, 1971; Higgs & Hodson, 1978; Menary, Trehub, & McNutt, 1982; Morgan, 1984) and their more accurate detection of mispronunciations in word-initial position compared to those in medial and final position (Cole, 1981). Recent evidence
indicates that attention to word onset appears in the second year of life, long before explicit phonemic awareness develops. Infants 18 months of age detect anomalies occurring in the initial position of familiar words, as indicated by longer fixations to a target when labels are correctly pronounced than when they are mispronounced (Swingley & Aslin, 2000). Expansion of the lexicon is thought to promote reorganization of phonological knowledge, with changes occurring on a word-by-word basis rather than an all-or-none fashion. In particular, groups of similar-sounding words (e.g., cat, pat, scat, calf, kit) are likely to undergo segmentation first because of their potential confusability (Metsala, 1997; but see Swingley & Aslin, 2000). For the most part, however, 5-year-olds are thought to be biased towards holistic word recognition (Jusczyk, 1997; Walley, 1993; but see Gerken, Murphy, & Aslin, 1995).

In support of Cole and Jakimik's (1980) assumption of sequential recognition of words, 5-year-olds detect mispronunciations more readily in semantically constrained than in unconstrained sentences (Cole & Perfetti, 1980). Moreover, 6-year-olds find it easier to detect or repeat target words in regular discourse contexts than in semantically and syntactically anomalous prose (Liu, Bates, Powell, & Wulfeck, 1997; Tyler & Marslen-Wilson, 1981). Despite young children's ability to use semantic and syntactic cues to identify target words, there are indications that young children derive less benefit from these cues than do older children and adults (Cole & Perfetti, 1980; Tyler & Marslen-Wilson, 1981). In some circumstances, semantic contextual cues fail to provide any facilitation. For example, when 5-year-olds are required to identify words from word-initial cues, as in a gating task (Grosjean, 1980, 1985), they perform no differently in the presence or absence of semantic cues (Craig, Kim, Rhyner, & Chirillo, 1993). By contrast, adults require considerably less word-initial information to identify words in semantically rich contexts than in semantically impoverished contexts (Craig et al., 1993). It is possible, however, that developmental differences such as these are overestimated in tasks that place large metalinguistic demands on children (Liu et al., 1997). When such demands are minimized,
7-year-olds and adults seem to derive comparable benefit from contextual cues (Liu et al., 1997).

The available data on children's recognition of words in quiet backgrounds provide a rough picture of children's developing skills. By contrast, there is relatively little information about children's identification of speech in noisy conditions. In principle, children could use contextual information to compensate for degraded acoustic cues. To what extent can they do so? We know that 5-year-olds require more favorable signal-to-noise levels than adults to correctly identify isolated monosyllabic words (Elliott et al., 1979; Mackie & Dermody, 1986; Ousey, Sheppard, Twomey, & Palmer, 1989; Summerfield, Palmer, Marshall, Foster, & Twomey, 1994), but it is unclear whether cognitive as well as sensory factors are implicated.

Auditory development continues into middle and late childhood. For example, absolute thresholds for narrow-band noise do not reach adult levels until approximately 10 years of age (Trehub, Schneider, Morrongiello, & Thorpe, 1988). Masked thresholds for narrowband noise presented in a broadband masker also decrease with age (Schneider, Trehub, Morrongiello, & Thorpe, 1989). Although the relation between absolute thresholds and speech identification is unclear (Elliott et al., 1979; Mackie & Dermody, 1986; Nilsson, Soli, & Sullivan, 1994; Palmer, Sheppard, & Marshall, 1991; Siegenthaler, Pearson, & Lezak, 1954; Summerfield et al., 1994), young children's lesser auditory sensitivity undoubtedly contributes to age-related differences in word identification in noise.

Fluctuating attention is thought to contribute to children's elevated auditory thresholds (e.g., Wightman & Allen, 1992, but see Schneider & Trehub, 1992). For listeners who are periodically inattentive and resort to random guessing, the rate of improvement as a function of signal intensity will be slower than that of optimally attentive listeners. The usual assumption is that young children are less attentive than older children and adults and that their motivation is affected by task demands. As a result, adaptive procedures may yield less accurate estimates of ability than does the method of constant stimuli (Schneider & Trehub,
1992). Although corrective feedback has favorable effects on performance (Smith & Hodgson, 1970), it is often absent in children's tasks (e.g., Elliott, 1979; Elliott et al., 1979; Nittrouer & Boothroyd, 1990).

There are suggestions that the distracting effects of noise are greater for younger than for older listeners. For example, classroom noise is associated with poor performance on everyday school tasks, an effect that is more pronounced at younger ages (Hétu, Truchon-Gagnon, & Bilodeau, 1990; Houtgast, 1981; Whaley & Hanson, 1984). However, noise of constant amplitude may not differentially disrupt children's attention. Schneider et al. (1989) found that increments of broadband noise led to comparable performance decrements across age, which ruled out attention as a contributing factor. The situation may be different when noise has a fluctuating amplitude (e.g., multitalker babble).

Other factors that are likely to influence children's identification of words are linguistic knowledge and history of exposure to the target words (Elliott et al., 1979; Elliott et al., 1983). For example, adults show more rapid recognition of words acquired early in life compared to those acquired later (Brown & Watson, 1987). Similarly, 5-year-olds are better at detecting mispronunciations in familiar words than in less familiar words (Walley & Metsala, 1990).

Stimulus unpredictability is known to depress performance (Spiegel & Green, 1982; Spiegel, Picardi, & Green, 1981), but the consequences may be greater for children than for adults (Allen & Wightman, 1995). Random variations in the frequency of a masker have greater consequences for preschoolers than for adults, as evidenced by higher thresholds and slower rates of improvement with more favorable signal-to-noise levels (Allen & Wightman, 1995). Similarly, the identification of words produced by multiple talkers is more difficult for preschoolers than for older children (Oliver Ryalls & Pisoni, 1997). These findings are consistent with reports of age-related improvement in the ability to attend to relevant aspects of an auditory stimulus (e.g., Zukier & Hagan, 1978).
Cognitive efficiency improves throughout childhood (e.g., Kail, 1991), but maximal efficiency may not be evident under challenging circumstances, such as those involving noise. For example, noise imposes greater cognitive demands on elderly than on young adult listeners, as reflected in their performance on short-term memory tasks presented in noise (Murphy, Craik, Li, & Schneider, 2000; Pichora-Fuller, Schneider, & Daneman, 1995). However, elderly adults derive comparable, if not greater, benefit from contextual cues when they identify words in noisy backgrounds (Dubno, Ahlstrom & Horwitz, 2000; Pichora-Fuller et al., 1995).

Young children's holistic bias in word recognition (Jusczyk, 1997; Walley, 1993) may also impair their performance in adverse listening conditions. When the target words (e.g., bat) have similar-sounding alternatives (e.g., cat), phonetic segmentation skills become critical (Bradlow & Pisoni, 1999). Although young children accord priority to word-initial information in some circumstances (Cole & Perfetti, 1980; Knafle, 1973, 1974; Lenel & Cantor, 1981), noise may interfere with such left-to-right decoding processes.

As noted, contextual cues are beneficial to adults in degraded listening environments (e.g., Dubno et al., 2000; Kalikow, Stevens & Elliott, 1977; Pichora-Fuller et al., 1995). Young children also profit from semantic and syntactic cues in some situations (e.g., Cole, 1981; Cole & Perfetti, 1980; Liu et al., 1997; Tyler & Marslen-Wilson, 1981), but there are indications that noise interferes with the optimal use of such cues (Elliott, 1979; Nittrouer & Boothroyd, 1990).

In sum, the available information about children's identification of words in noise is fragmentary. Accordingly, the goal of this research was to enrich the limited database in this realm by documenting age-related differences in auditory word recognition in multitalker babble. The first study (Chapter 2) explored differences in the identification of monosyllabic words by 5-, 9-, and 11-year-old children and adults. Controlling for age-related differences in stimulus audibility and presenting target words in two levels of babble made it possible to examine the relative contributions of perceptual and cognitive factors. The second study
(Chapter 3) compared 5-year-olds, 9-year-olds, and adults on their identification of words when the response alternatives involved similar-sounding words (i.e., minimal pairs). The third, and final, study (Chapter 4) evaluated 5-year-olds', 9-year-olds', and adults' ability to make use of semantic cues when identifying words in sentential contexts.
Chapter 2

Children’s perception of speech in multitalker babble

Introduction

Children are believed to have considerable difficulty perceiving speech in noise (e.g., Elliott, 1995; Mills, 1975). For example, normative estimates of speech perception (e.g., The Goldman-Fristoe-Woodcock Test of Auditory Discrimination, Goldman, Fristoe, & Woodcock, 1976) reflect young children’s poor speech identification in “cafeteria noise” relative to that of older children and adults. In classroom settings, younger children are more “distracted” by noise than are older children (Hétu et al., 1990). Moreover, young children are significantly less accurate at identifying the last word of a sentence presented in multitalker babble than are older children and adults (Elliott, 1979; Elliott et al., 1979).

Similarly, children 4 to 6 years of age are significantly poorer than adults at identifying words and sentences in spectrally matched noise (Nittroeur & Boothroyd, 1990).

A number of factors may contribute to children’s apparent difficulty with speech in noise. First, young children have higher auditory thresholds than do older children and adults (Berg & Smith, 1983; Elliott & Katz, 1980; Roche, Siervogel, Himes, & Johnson, 1978; Yoneshige & Elliott, 1981; Sinnott, Pisoni, & Aslin, 1983; Schneider, Trehub, Morrongiello, & Thorpe, 1986; Trehub et al., 1988). Although the relation between pure-tone sensitivity and speech identification thresholds in quiet and in noise is unclear (Elliott et al., 1979; Summerfield et al., 1994), age-related differences in pure-tone sensitivity could underlie children’s difficulty identifying speech in noise. From 5 years of age, age-related changes in absolute thresholds for octave-band noise mirror the changes in masked thresholds for octave-band noise presented in broadband noise (Schneider et al., 1989). Moreover, thresholds for the identification of words presented in quiet are higher for younger children than for older children and adults (Elliott et al., 1979). Nevertheless, most developmental investigations of speech identification in noise have used identical signal-to-noise ratios.
(SNRs) across age without adjusting for age-related differences in detection or identification thresholds (e.g., Elliott, 1979; Nittrouer & Boothroyd, 1990). Chermak and Dengerink (1981) found, however, that when age-related differences in word identification thresholds were taken into account, adult-child performance differences in noise were minimal.

Second, young children's limited language experience may have adverse effects on their performance. For example, native speakers of English are more proficient at identifying speech in noise than are nonnative speakers with several years of exposure to English (Gat & Keith, 1978; Mayo, Florentine, & Buus, 1997). As words become increasingly familiar, less acoustic information is required for their identification (Elliott et al. 1979; Elliott et al., 1983; Rosenweig & Postman, 1957). In some cases, however, words (e.g., "oath") and sentences ("Tough guys sound mean") that are unfamiliar to many children have been used as test stimuli (e.g., Elliott, 1979; Nittrouer & Boothroyd, 1990). Moreover, limited phonological awareness on the part of young children (Hnath-Chisolm, Laipply, & Boothroyd, 1998; Treiman, 1985b), especially pre-readers (Wimmer, Landerl, Linortner, & Hummer, 1991), may also impair performance on speech identification tasks. For example, a young child hearing "_ike" might not generate "bike" as a candidate word. Even if a child can use phonological strategies to aid identification, noise may disrupt this process.

Third, the typical tasks used in speech identification studies may pose disproportionate difficulty for young children. As Wightman and Allen (1992) note, some auditory performance differences between children and adults "may reflect nothing more than the influence of nonsensory factors such as memory and attention" (p. 133). These nonsensory factors could contribute to young children's difficulty with verbal material that has been designed for use with adults. For example, some investigators have adapted the well-known SPIN (Speech Perception In Noise) test for use with children (e.g., Elliott, 1979; Elliott & Katz, 1983, in Elliott, 1995). The original (Kalikow et al., 1977) and revised (SPIN-R) versions (Bilger, Neutzel, Rabinowitz, & Rzeczowski, 1984) of the SPIN test, which were normed for native English-speaking adults, require listeners to repeat the last
word of low-predictability ("I had not thought about the growl.") and high-predictability ("The watchdog gave a warning growl.") sentences presented in multitalker babble. The presumption is that low-predictability sentences depend primarily on sensory function, and that high-predictability sentences engage cognitive as well as sensory functions (Kalikow et al., 1977). Children's (9- and 11-year-olds') poor performance on both types of sentences led Elliott (1979) to conclude that the SPIN task was inappropriate for use with children under the age of 15. The required verbal responses are potentially problematic for listeners with limitations in articulation and memory. Indeed, picture-pointing responses yield substantially higher performance levels for young children than do verbal responses (Elliott et al., 1979). Although feedback about performance is also known to enhance accuracy (Green & Swets, 1966; Smith & Hodgson, 1970), such feedback is often excluded from developmental investigations (e.g., Elliott et al., 1979, Nittrouer & Boothroyd, 1990). Its absence may have particularly negative consequences on children's motivation.

Nittrouer and Boothroyd (1990) tested 4- to 6-year-old children on a task that may be even more demanding than the SPIN task. Children were required to repeat various types of verbal material presented in noise, including nonsense syllables, monosyllabic words, and four-word sentences, some of which were semantically and/or syntactically anomalous (e.g., "Sing his get throw," "Lend them less joy"). Semantic and syntactic anomalies are likely to be especially confusing for young children. Unfortunately, Nittrouer and Boothroyd (1990) did not confirm that young children were capable of repeating such anomalous sentences under optimal listening conditions.

The primary goal of the present investigation was to determine whether noise impairs children's identification of speech to a greater extent than it does for adults. In other words, does noise affect children's perception of speech beyond what would be expected from adult-child differences in auditory sensitivity? Unlike previous adult-child comparisons that evaluated all listeners at identical SNRs (e.g., Elliott, 1979; Nittrouer & Boothroyd, 1990), the present study identified SNRs that yielded 85% correct performance (i.e., a low-noise
condition) for 5-year-olds, 9-year-olds, 11-year-olds, and adults. A high-noise condition was created by adding 7 dB of noise. Based on adult-child differences in the detection of signals in quiet (e.g., Schneider et al., 1986; Trehub et al., 1988) and in noise (Schneider et al., 1989), 5-year-olds were expected to require at least 5-dB less background noise than adults to achieve comparable (85% correct) performance in low noise.

Comparable decrements in performance across age in the high-noise condition would imply that the primary determinant of performance was stimulus audibility (i.e., speech awareness threshold) rather than some combination of factors (e.g., distraction, interference with lexical access, or higher-level auditory processing difficulties). Alternatively, disproportionate performance decrements on the part of 5-year-olds may reflect, among other things, difficulty focusing attention on the target signal or retrieving words from the lexicon (i.e., lexical access). A final possibility is that 5-year-olds’ performance in high noise exceeds that of adults, an outcome consistent with less effective higher-level auditory processing on the part of 5-year-olds (see Pichora-Fuller et al., 1995). Although this notion seems counterintuitive, it becomes clear when we consider the entire range of performance as a function of SNR. If young children process higher-level auditory information less effectively than adults, they would require more favorable SNRs to reach 85% correct performance than would be expected from sensory functioning alone (5 dB), and, most importantly, their rate of improvement with more favorable SNRs would not be as dramatic as that of adults. Thus, the addition of a fixed increment of noise would result in greater performance decrements for adults than for 5-year-olds.

The available evidence does not favor a single hypothesis. According to Elliott (1979), perceptual and cognitive factors contribute jointly to children’s poor performance on the SPIN task. Nittrouer and Boothroyd (1990) argue, however, that children’s poor performance on their task resulted from perceptual factors.

A second goal of the present investigation was to develop a procedure that would minimize cognitive demands and maximize comparability across a broad age range. A four-
alternative, picture-pointing task was used because of its documented success with young children (Geffner, Lucker, & Koch, 1996; Goldman et al., 1976) and its obvious advantages over word-generation tasks (Elliott et al., 1979). Target words were presented in a low-context carrier phrase ("Touch the X") because of the present focus on perceptual, or bottom-up, factors and the reported advantage of sentential contexts over words presented in isolation (Craig, 1988). Target words were restricted to those that were familiar to 5-year-olds, the youngest age group in the present study. Moreover, target words and foils were highly contrastive phonologically so that children would not be penalized for their lesser phonological awareness (Treiman, 1985b). Sentences were presented in multitalker babble (Bilger et al., 1984) rather than spectrally-matched noise (Boothroyd & Nittrouer, 1988; Nittrouer & Boothroyd, 1990) because babble is a more effective masker of speech (Carhart, Tillman, & Greetis, 1969; Elliott et al., 1979; Lewis, Benignus, Muller, Mallott, & Barton, 1988). A female speaker was used because of the predominance of female caregivers and educators in the lives of young children. Finally, a motivating, game-like atmosphere was created by presenting the pictorial response options on a touch-sensitive screen and providing automated visual feedback for correct and incorrect responses.

Method

Participants.

The participants were 24 children 5;0-5;6 years of age ($M = 5;3$ years), 24 children 9;0-9;6 years of age ($M = 9;3$ years), 24 children 11;0-11;6 years of age ($M = 11;3$ years) and 24 adults 19-28 years of age ($M = 22.7$ years), none of whom had health problems or a history of hearing loss. Equal numbers of males and females were included in each age group. No participant had experienced frequent ear infections or pressure-equalizing tubes in the past; none had a cold on the day of testing. Children's age and their family's middle-to upper-middle-class status made it unlikely that serious middle-ear problems had gone undetected; thus no tympanometric screening was undertaken. The children were all native English speakers and the adults were either native speakers or had learned English by 6 years
of age (n = 3). Additional children were excluded because of experimenter error (one 9-year-old) or inattentiveness (four 5-year-olds and one 9-year-old), which included the following behaviors: responding before hearing the entire sentence on more than two trials, excessive fidgeting, talking during sentence presentation, talking excessively between test trials, or not completing one or both experimental conditions.

**Stimuli and Apparatus.**

Testing occurred in a double-wall sound-attenuating booth, 3 x 2.8 x 2 m in size. Participants were seated facing a non-glare touch screen monitor (Goldstar 1465DLs) 33 cm x 33 cm. Loudspeakers (KEF Model 101) were 45° to the left and right of the participant (distance of 70 cm) at approximate ear level. All sentences, which were spoken by the same young woman, were digitized at a rate of 20 kHz by means of a 16-bit Tucker Davis (DD1) analog-to-digital converter. The babble portion of SPIN forms used by Pichora-Fuller et al. (1995) was similarly digitized and stored. Sentence files and babble files were converted to analog form using Tucker-Davis digital-to-analog converters under the control of a computer with a Pentium processor. Sentence and babble amplitudes were controlled separately by means of programmable attenuators. After mixing, the combined signals were amplified (SAE 2600) and presented over loudspeakers located inside the testing booth. Sound-field levels were determined in the absence of the listener with a Brue and Kjaer ½-in microphone.

The multitalker babble, which did not contain energy above 8 kHz, consisted of eight voices, both male and female, reading from newspapers (see Bilger et al., 1984 for further details). The degree to which a speech signal is masked by babble will depend on the spectral characteristics of the speech relative to that of the background babble. Figure 2.1 (dashed line) shows the distribution of average spectral power in the background babble. This distribution was obtained by averaging the power spectra of 40 independent 1-s samples of the babble background. The average power spectrum was then normalized by dividing it by the total power in the average spectrum. This normalized average power spectrum was
converted to decibels and plotted in Figure 2.1. The average power in the babble background declines with increasing frequency at a rate of approximately 3 dB/octave for frequencies up to approximately 700 Hz, and 12 dB/octave for frequencies higher than 700 Hz. Also shown in Figure 1 is the average power spectrum for the male voice used in the modified SPIN test (Bilger et al., 1984). This spectrum was obtained by averaging the power spectra (1-s samples taken from the beginning of each sentence) of the first 40 SPIN sentences from Form 1. The average power spectrum was then normalized by dividing it by the total power in the average spectrum. Clearly, the relative power spectrum of the male voice from the SPIN test closely matches that of the babble background. By contrast, Figure 2.2 plots the relative power spectrum of the female voice used in the present experiment. This spectrum was obtained by averaging the power spectra of the 40 test sentences (1-s samples taken from the beginning of each sentence). Not only was the fundamental frequency of the female speaker (255 Hz) much higher than that of the male speaker, but the distribution of power in the female voice was heavily weighted toward the higher frequencies.
Figure 2.1: Averaged power spectra of the babble used in the present study and the male speaker of the SPIN sentences (Bilger et al., 1984).
Figure 2.2: Averaged power spectra of the babble and the female speaker from the present study.
The female speaker’s high degree of power at the high frequencies vastly improves the SNR at these frequencies. If we adjust the total power in the babble to equal the total power in each of the speech signals (an overall SNR of 0 dB), the average SNR for the male speaker is -0.06 dB in the 0-1.5 kHz region and +4.49 dB in the 1.5-5 kHz region. By contrast, the average SNR for the female speaker is -2.40 dB in the 0-1.5 kHz region and +16.92 dB in the 1.5-5 kHz region. Thus, the male speaker enjoys a 2.34-dB advantage at the low frequencies whereas the female speaker enjoys a 12.4-dB advantage at the high frequencies. Because it is usually assumed that the low- and high-frequency portions of the spectrum contribute equally to speech recognition, we might expect the female speaker to enjoy a 12-dB advantage relative to the male speaker in the babble background.

Speech stimuli consisted of the prompt, “Touch the ____,” and a target word (e.g., “ball”). Target words consisted of 40 monosyllabic nouns for the test phase and an additional 20 words (monosyllabic and polysyllabic) for the training phase. The sentences were presented at approximately 44 dB (A scale). This level was chosen to ensure that young children would not be exposed to sound levels exceeding 80 dB (A). Root-mean-square (RMS) values were calculated and adjusted such that each sentence was presented at an equal RMS value following the procedure described in Schneider, Daneman, Murphy, and Kwong-See (2000). SNR was varied by adjusting the level of babble (F₀: 185 Hz). Pilot-testing established the SNR at which each age group achieved approximately 85% correct performance: -28 dB for 5-year-olds, -30 dB for 9-year-olds, -31 dB for 11-year-olds, and -33 dB for adults. These levels were designated low-noise conditions. High-noise conditions were created for each age group by decreasing the SNR in the low-noise condition (i.e., the level yielding 85% correct performance) by 7 dB, resulting in SNRs of -35 dB for 5-year-olds, -37 dB for 9-year-olds, -38 dB for 11-year-olds, and -40 dB for adults. SNRs for the training phase were set lower than those in the low-noise conditions: 0 dB, -5 dB, -6 dB, and -8 dB for 5-year-olds, 9-year-olds, 11-year-olds, and adults, respectively.
Visual stimuli consisted of 60 black-and-white Snodgrass line drawings of familiar, concrete objects (Snodgrass & Vanderwart, 1980). All pictures were chosen on the basis of 4- and 5-year-olds’ ability to correctly name the image. An independent sample of 32 children 4;0 to 5;6 years of age was asked to verbally identify each of the 60 pictures. Inclusion of a picture in the stimulus set required at least 88% of children correctly identifying it. The average correct identification of pictures was 96.8%.

Procedure.

All participants were tested individually. A trial, which was initiated by means of a button box located inside the testing booth, consisted of the simultaneous presentation of vocal stimuli (sentence and noise) and visual stimuli (pictures). A sentence in low or high noise was accompanied by an array of four different images, one appearing in each corner of the touch screen. Sentences were selected randomly without replacement. The multitalker babble began with the onset of the sentence and terminated when the sentence ended. The visual array included the target image and three foils selected randomly from the remaining images. A picture could appear as a target only once and as a foil three times during the test phase. The only other restriction on foils was that an item could not serve as a foil immediately after it was presented as a target. The locations of targets and foils were selected randomly on each trial. Feedback for correct performance consisted of the target picture flashing in the middle of the screen. Incorrect selections resulted in the screen going blank.

The instructions were tailored to the age of participants. The experimenter explained to 5-year-old children that if they only hear part of a word, they should choose the picture that sounds similar to what they hear (e.g., “If you hear ‘irt’ and there are pictures of a can, plate, shirt, and boat on the screen, you should pick the shirt since ‘shirt’ sounds the most like ‘irt’.”). Older children and adults were told that the pictures were identifiable by basic-level terms. For example, a picture of a shirt would be identified by the word “shirt,” not “button-down” or “clothing.” No other explicit strategies were provided.
The test session consisted of a training phase and a test phase. All participants had to meet a training criterion of correctly identifying 4 targets in a row within 16 trials; on average, listeners achieved the training criterion in 6.04 trials. After reaching the criterion, participants advanced to the test phase, consisting of 40 trials in the low-noise condition and another 40 trials in the high-noise condition. The two conditions, which were separated by a short break, were counterbalanced such that half of the participants received the low-noise condition first and the other half received the high-noise condition first. Adults and older children initiated trials at their preferred pace. The experimenter initiated trials for 5-year-olds when she judged them to be ready and attentive. The experimenter remained in the testing booth during the entire session for children, offering verbal reinforcement and encouragement when appropriate. An additional motivational technique was used with the 5-year-olds. After every four trials, children received a colored sticker to place in an "incomplete" black-and-white picture. At the end of the 40 trials, the child had completed the picture. A new picture was made in the second condition.

Results

Figure 2.3 plots correct performance at each noise level for each age group. Performance (percent correct) in the low-noise condition did not differ significantly across age groups, as confirmed by a one-way ANOVA, \( F(3,92) = .49, p = .689 \). Recall, however, that SNRs were selected to equalize performance in low noise across age levels. To examine whether the high-noise condition differentially affected younger children, a \( 2 \times 4 \times 2 \times 2 \) repeated-measures ANOVA was calculated with noise level as the within-subject factor and age, sex, and presentation order (low or high noise first) as between-subjects factors. As expected, listeners performed significantly better in low noise than in high noise, \( F(1,80) = 678.00, p < .00001 \) (see Table 2.1). There was no effect of age and no effect of gender, but there was a significant effect of presentation order, \( F(1,80) = 11.42, p = .001 \). Participants made more correct selections when the low-noise condition was presented before the high-noise condition than the reverse order (\( M_{\text{low}, \text{high}} = 74.79\%, M_{\text{high}, \text{low}} = 70.96\% \)). The only
two-way interaction that achieved statistical significance was a noise x order interaction, $F(1,80) = 24.90, p < .001$, reflecting listeners' improved performance in the high-noise condition when it was preceded by the low-noise condition. No higher-order interactions were significant.

Figure 2.3: Percent correct identification as a function of noise level and age. Error bars represent the standard error of the mean.
To examine potential age-related differences with regard to the noise x order interaction, a 2 x 4 ANOVA was calculated using performance in high noise as the dependent variable, with order and age as independent variables. Not surprisingly, there was a significant effect of order, $F(1, 88) = 27.15, p < .001$, but no effect of age or age x order interaction (see Figure 2.4). However, one-way ANOVAs conducted on each age group separately revealed that the order effect in high noise was somewhat more pronounced for 9-year-olds, 11-year-olds, and adults [$F(1, 22) = 5.62, p < .03, F(1, 22) = 7.16, p < .02$, and $F(1, 22) = 11.48, p < .005$, respectively] than for 5-year-olds [$F(1, 22) = 3.47, p = .076$].
Figure 2.4: Percent correct identification in high noise as a function of age and order. Error bars represent the standard error of the mean.

Patterns of performance were also examined for potential age-related improvement within the test sessions. Gender was excluded from this analysis because it had no effect in the main analysis. Given that the order effect was noise-level specific, analyses were conducted separately for each noise condition. A $2 \times 4$ repeated measures ANOVA was performed using percent correct per half in low noise (i.e., first 20 trials vs. final 20 trials) as the within-subject factor and age as the between-subjects factor. Order was excluded from this analysis because it had no effect in low noise. Listeners performed more accurately in the second half of the test session than in the first half, $F(1, 92) = 12.40$, $p = .001$, ($M = 86.09\%$, $M = 81.98\%$, respectively). No age effect or age x half interaction was observed (see Table 2.2). In the high-noise condition, a $2 \times 2 \times 4$ repeated measures ANOVA, with half as the within-subject variable and order and age as between-subjects variables, revealed
an order effect, $F(1, 88) = 27.591$, $p < .001$, reflecting the effect found in the main analysis.

No additional main effects or higher-order interactions were observed.

Table 2.2: Percentage correct performance per half in low noise.

<table>
<thead>
<tr>
<th></th>
<th>1&lt;sup&gt;st&lt;/sup&gt; Half</th>
<th>2&lt;sup&gt;nd&lt;/sup&gt; Half</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td>81.04 (8.97)</td>
<td>85.21 (7.44)</td>
</tr>
<tr>
<td>9-year-olds</td>
<td>82.71 (9.09)</td>
<td>84.58 (7.65)</td>
</tr>
<tr>
<td>11-year-olds</td>
<td>83.12 (8.05)</td>
<td>86.46 (9.26)</td>
</tr>
<tr>
<td>Adults</td>
<td>81.04 (9.09)</td>
<td>88.12 (5.67)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

Table 2.3 indicates the relative difficulty of identifying particular target words, in descending rank order. Note, for example, that MOON (rank = 40.00) was the most difficult target to identify out of this sample of items, with TREE (rank = 2.13) being the easiest. To determine whether listeners found the same targets difficult, the average rank across age groups was correlated with the rankings by individual age groups. As shown in Table 2.4, these strong positive correlations reflect the fact that all listeners, irrespective of age, experienced comparable difficulty with the same target words.
Table 2.3: Average difficulty of targets in descending rank order.

<table>
<thead>
<tr>
<th>Target</th>
<th>Average Rank</th>
<th>Target</th>
<th>Average Rank</th>
<th>Target</th>
<th>Average Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>MOON</td>
<td>40.00</td>
<td>WHEEL</td>
<td>24.75</td>
<td>HOUSE</td>
<td>9.88</td>
</tr>
<tr>
<td>BOOK</td>
<td>38.13</td>
<td>CLOCK</td>
<td>23.63</td>
<td>CAT</td>
<td>9.75</td>
</tr>
<tr>
<td>BALL</td>
<td>36.63</td>
<td>COW</td>
<td>21.63</td>
<td>GRAPES</td>
<td>9.63</td>
</tr>
<tr>
<td>FORK</td>
<td>36.50</td>
<td>BREAD</td>
<td>21.25</td>
<td>KEY</td>
<td>7.25</td>
</tr>
<tr>
<td>CORN</td>
<td>36.38</td>
<td>SPOON</td>
<td>21.13</td>
<td>SHOE</td>
<td>6.63</td>
</tr>
<tr>
<td>KNIFE</td>
<td>33.00</td>
<td>EAR</td>
<td>19.63</td>
<td>CHAIR</td>
<td>6.50</td>
</tr>
<tr>
<td>BELL</td>
<td>32.88</td>
<td>DOG</td>
<td>18.00</td>
<td>SUN</td>
<td>5.38</td>
</tr>
<tr>
<td>LEAF</td>
<td>31.63</td>
<td>DUCK</td>
<td>17.88</td>
<td>FISH</td>
<td>4.50</td>
</tr>
<tr>
<td>EYE</td>
<td>30.63</td>
<td>HAT</td>
<td>17.75</td>
<td>SNAKE</td>
<td>2.63</td>
</tr>
<tr>
<td>FLAG</td>
<td>29.75</td>
<td>CAKE</td>
<td>17.63</td>
<td>TREE</td>
<td>2.13</td>
</tr>
<tr>
<td>FROG</td>
<td>29.38</td>
<td>HORSE</td>
<td>17.13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BED</td>
<td>27.63</td>
<td>BUS</td>
<td>15.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HEART</td>
<td>26.75</td>
<td>HAND</td>
<td>14.88</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRUCK</td>
<td>26.00</td>
<td>KITE</td>
<td>13.50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PIG</td>
<td>25.50</td>
<td>PANTS</td>
<td>10.88</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Maximum rank is 40 (i.e., most difficult), minimum rank is 1 (i.e., easiest).
Table 2.4: Spearman correlation coefficients relating overall average ranking with rankings of individual age groups.

<table>
<thead>
<tr>
<th></th>
<th>5-year-olds</th>
<th>9-year-olds</th>
<th>11-year-olds</th>
<th>Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average rank</td>
<td>.93</td>
<td>.91</td>
<td>.95</td>
<td>.93</td>
</tr>
</tbody>
</table>

Note: All correlations are significant to the \( p = .01 \) level at \( \alpha = .05 \).

Discussion

When the accuracy of speech identification by 5-year-olds, 9-year-olds, 11-year-olds, and adults was equated in low noise, the addition of further noise (7 dB) had comparable consequences for all age groups. Note, however, that 5-year-olds still required SNRs (-28 and -35 dB at low- and high-noise levels, respectively) that were 5 dB more favorable than those of adults (-33 and -40 dB) to obtain comparable performance. These results are consistent with evidence that 5-year-olds’ and adults’ absolute and masked thresholds for narrow-band noise differ by approximately 5 dB (Trehub et al., 1988; Schneider et al., 1989). Moreover, Elliott et al. (1979), using a four-alternative picture-pointing response, found that 5-year-olds’ and adults’ identification thresholds for monosyllables presented in babble differed by approximately 5 dB. Comparable performance decrements from high noise, regardless of age, imply that children’s performance on the present task results from differences in stimulus audibility rather than nonsensory factors. In fact, the minimal impact of nonsensory factors attests to the utility of the present task, which featured low-context sentences, words familiar to the youngest participants, a four-alternative picture-pointing response, and visual reinforcement. Moreover, comparable difficulty with the same target words at all age levels attests further to the role of sensory factors and to the limited contribution of linguistic experience to the present findings. Thus, the findings support Nittrouer and Boothroyd’s (1990) contention that age-related differences in word
Identification in noisy backgrounds are largely due to perceptual factors, provided the task is equally suitable for all age levels.

Children and adults performed better in high noise when it followed the low-noise condition rather than preceding it. In other words, participants showed evidence of learning in the initial low-noise condition that generalized to the subsequent high-noise condition. It is interesting to note that the non-significant age trend in high-noise performance as a function of order parallels the non-significant age-related improvement in performance over the course of the low-noise condition. Experience in a relatively undemanding situation (low noise) may have allowed listeners to gain information, perhaps about the speaker's voice or other aspects of the task, that facilitated subsequent performance under more difficult circumstances (high noise). By contrast, experience in high noise showed no such transfer effects.

Although the SNR required for word recognition is affected by characteristics of the speaker's voice, the nature of the speech materials, and whether or not the materials are from an open or closed set (Miller, Heise, & Lichten, 1951; Sumby & Pollack, 1954), it was still surprising to find young adults recognizing 85% of the words at SNRs as low as -33 dB. The unusually low SNRs result, in part, from the spectral characteristics of the female voice in relation to the background babble. Recall that the speaker's power at the higher frequencies translated to a 12-dB advantage in SNR relative to the male speaker used in the SPIN task (Bilger et al., 1984). Moreover, the closed-set response (i.e., selecting from four alternatives) undoubtedly contributed to the high levels of performance at such low SNRs.

To ascertain the consequences of open-set responding, an independent sample of 20 adults (6 males, 14 females; ages 19-25) was tested on the same target words without the pictorial alternatives. After listening to each of the same sentences in the same babble background at -22 dB SNR or -18 dB SNR, adults identified the last word of the sentence by means of written responses. Listeners achieved 81.125% and 89.17% correct performance at -22 dB SNR and -18 dB SNR, respectively. Thus, the SNR that would yield 85% correct
performance with open-set responding is approximately -20 dB. In other words, the change from a closed set of four alternatives to an open set resulted in a 13-dB shift in SNR. Miller et al. (1951) found a comparable shift between a 256-word response set and a 4-word response set. The 13-dB advantage attributable to the four-alternative response-set coupled with the 12-dB high-frequency advantage for the female voice can account for the performance levels obtained in the context of the very low SNRs in the present experiment.

Although equivalent additions of noise had comparable effects on children and adults in the present task, such noise increments could have differential consequences in situations that accord a greater role to cognitive, or top-down, factors. For example, the use of high-predictability as well as low-predictability sentences would reveal whether noise interferes with children's ability to profit from contextual information. The present procedure could be used to investigate this question and others involving the identification of speech in quiet and in noise.

Children 5, 9, and 11 years of age required more favorable SNRs than young adults to achieve comparable accuracy on low-context sentences presented in background babble. Nevertheless, equivalent increases in noise level led to similar performance decrements for all age groups. The findings are consistent with the view that bottom-up (sensory) processing plays the primary role in children's and adults' perception of simple, low-context sentences in noisy backgrounds. The availability of a sensitive means for evaluating children's perception of speech in noise will make it possible to document the impact of noise in situations with varying cognitive demands. It particular, it will allow us to delineate the relative contribution of perceptual and cognitive factors to the identification of spoken messages varying in complexity and listening conditions.
Chapter 3

Children's identification of minimally contrastive words in noise

Introduction

Children are less accurate than adults at identifying speech in noisy backgrounds (Chapter 2; Elliott, 1979; Elliott et al., 1979; Nittrouer & Boothroyd, 1990), but the reasons for this situation are unclear. No doubt, their elevated auditory thresholds (e.g., Schneider et al., 1986, 1989) contribute to poor performance in noise. Moreover, noise may interfere with children's optimal allocation of attention (Mills, 1975) and their deployment of effective linguistic strategies (Elliott, 1979). In Chapter 2, we minimized the cognitive demands on children by presenting highly familiar words in simple sentential contexts, having listeners respond by means of picture-pointing, and using monosyllabic target words that were phonologically dissimilar from the alternatives. Under those circumstances, age-related differences in speech identification in noise paralleled age-related differences in auditory thresholds. In other words, nonsensory factors played a minimal role.

The situation may differ when the target words and foils are phonologically similar. Although children as young as 4 can discriminate familiar words differing by a single phoneme (Graham & House, 1971; Higgs & Hodson, 1978; Menary et al., 1982; Morgan, 1984), their performance is not error-free, even in quiet conditions. By contrast, adults are highly accurate at discriminating minimal pairs in quiet backgrounds (Graham & House, 1971). Interestingly, when isolated words are presented in noise, adults have more difficulty transcribing words with many "competitors" (i.e., similar-sounding alternatives) than they do with words from sparse phonological neighborhoods (Luce & Pisoni, 1998; Sommers, Kirk, & Pisoni, 1997).

Young children are presumed to use global word recognition strategies, in contrast to older children, who focus on individual phonetic segments (Jusczyk, 1992; Metsala, 1997; Treiman & Breaux, 1982; Walley, Smith, & Jusczyk, 1986; Walley, 1993; but see Gerken et
Partial sensory information about a word, including word-initial input, may not constitute a reliable basis for young children’s word recognition due to their rapidly expanding lexicon and their weaker lexical knowledge compared to that of adults (Walley, 1988). Fine-grained acoustic analysis, which involves attention to each phonetic segment, is required to identify words that have similar-sounding neighbors (Bradlow & Pisoni, 1999; Pisoni & Luce, 1987). Perhaps children’s global recognition strategy is responsible for a lexicon with fewer close phonological neighbors compared to that of older children (Charles-Luce & Luce, 1990, 1995). According to Metsala (1997), the acquisition of words from dense phonological neighborhoods leads children to restructure their lexicon in terms of phonetic segments, a process that extends into middle childhood. Presumably, neighborhood density would have less impact on forced-choice responding (e.g., selecting one of four pictures) than on tasks involving word generation (Sommers et al., 1997, but see Tanenhaus, Magnuson, Dahan, & Chambers, 2000).

Stimulus variability may pose further difficulty for young children. Indeed, preschoolers identify words spoken by a single talker more accurately than those spoken by multiple talkers (Oliver-Ryalls & Pisoni, 1997). Target words that differ from foils in unpredictable ways may pose more difficulty for children than those that differ from foils in predictable ways. For example, monosyllabic CVC words that are minimally contrastive can differ in their initial, medial, or final phoneme. One would expect children to perform better if all word pairs in a set differed in the same way, say, in their initial phoneme.

The position of contrastive elements also contributes to the difficulty of the discrimination. Listeners of all ages experience more difficulty discriminating word-final than word-initial contrasts, even in quiet backgrounds (Hnath-Chisolm et al., 1998; Menary et al., 1982; Morgan, 1984; Redford & Diehl, 1999). They are thought to accord greater attention to word beginnings than to word endings because the former occur earlier, are less redundant, and are less affected by coarticulation (Nootboom, 1981). Moreover, the initial segment of a monosyllabic word is of longer duration and of greater energy than its final
segment (Redford & Diehl, 1999), which should yield higher performance on word-initial than on word-final contrasts. Word-initial information also plays a dominant role in most theories of speech recognition (e.g., Marslen-Wilson, 1987; Norris, 1994; Taft & Forster, 1975) and becomes increasingly important in the development of lexical and articulatory abilities (Brooks & MacWhinney, 2000; Wijnen, 1992). Indeed, 5-year-olds detect word-initial mispronunciations more accurately than 4-year-olds, but both age groups exhibit comparable performance on word-final mispronunciations (Walley, 1987). Although Walley (1987) concedes that 4- and 5-year-olds give priority to word-initial information, they may do so only for highly constrained situations that provide cues to word identity. Young children's lesser attention to word endings has parallels in production: Common misarticulations by preschoolers include the omission of final, unvoiced consonants, such as ba for bat, and the devoicing of final voiced consonants, such as bak instead of bag (Anisfeld, 1984). Coarticulatory factors have a greater effect on the perception of synthetic CVC monosyllables by young children compared to older listeners (Nittroer & Studdert-Kennedy, 1987). Adverse listening environments, such as those with noisy backgrounds, could further disrupt young children's discrimination of word-final contrasts.

The purpose of the present study was to compare 5-year-olds', 9-year-olds', and adults' discrimination of minimal pairs in noise. A prerecorded female voice presented in multitalker babble instructed listeners to touch one of four pictures. On some trials, the target picture (e.g., nail) was presented with an alternative differing in its initial phoneme (e.g., pail). On other trials, one of the alternatives differed in its final phoneme (e.g., cap/cat). Still other trials had maximally contrastive alternatives (e.g., boat/cake/house/mop). To investigate the contribution of stimulus uncertainty to performance, trial types (i.e., initial contrast, final contrast, or maximal contrast) were randomized or blocked. Noise, which consisted of background babble, was set at levels that had yielded approximately 85% performance on maximally contrastive trials for each age group in Chapter 2. Signal-to-noise
ratios (SNRs) selected in this way controlled for age-related differences in stimulus audibility.

Listeners, irrespective of age, were expected to experience more difficulty identifying target words when presented with phonologically similar foils than with maximally contrastive foils. Because the initial phoneme is louder and longer (i.e., less susceptible to masking) than the final phoneme (Redford & Diehl, 1999), word-initial contrasts were expected to be identified more accurately than word-final contrasts. If listeners benefit from lower stimulus variability, they should perform better on targets presented in blocked order compared to those in mixed order. At the very least, listeners in the blocked condition were expected to select the target or its phonologically similar foil (e.g., confusing cap with cat but not with bowl or house) more often than those who received a mixed order of trials.

Despite the adjustment of SNRs to compensate for age differences in stimulus audibility, young children were expected to perform more poorly than older listeners on minimal pair trials because of their limited phonetic segmentation skills (Jusczyk, 1992; Metsala, 1997; Treiman & Breaux, 1982; Walley et al., 1986; Walley, 1993). Moreover, 5-year-olds were expected to exhibit disproportionate difficulty on word-final contrasts because of the adverse consequences of coarticulatory factors (Nittrouer & Studdert-Kennedy, 1987). Finally, 5-year-olds were expected to be especially disadvantaged by stimulus variability, resulting from a mixed presentation of trial types.

Method

Participants.

The participants were 72 5-year-olds ($M = 5;3$; range 5;0-5;6), 72 9-year-olds ($M = 9;3$; range 8;10-9;8), and 72 adults ($M = 21.7$; range 18-25). Equal numbers of females and males were included in each age group. None of the children had experienced frequent ear infections or pressure equalizing tubes. All participants were native speakers of English with the exception of 7 adults, all of whom attended English-speaking elementary schools in North America. No participant had a cold on the day of testing. An additional 16 5-year-
olds and 3 9-year-olds were tested, but were excluded due to experimenter error (n = 4),
failure to complete the entire protocol (n = 10), or inattentiveness (n = 5).

Stimuli and Apparatus.

Testing occurred in a double-wall sound-attenuating booth, 3 x 2.8 x 2 m in size.
Participants were seated facing a non-glare touch screen monitor (Goldstar 1465DLs) 33 cm
x 33 cm. Loudspeakers (KEF Model 101) were 45° to the left and right of the participant
(distance of 70 cm) at approximate ear level. All sentences, which were spoken by the same
young woman, were digitized at a rate of 20 kHz by means of a 16-bit Tucker Davis (DD1)
analog-to-digital converter. The babble portion of SPIN forms used by Pichora-Fuller et al.
(1995) was similarly digitized and stored. Sentence files and babble files were converted to
analog form using Tucker-Davis digital-to-analog converters under the control of a computer
with a Pentium processor. Sentence and babble amplitudes were controlled separately using
programmable attenuators. After mixing, the combined signals were amplified (SAE 2600)
and presented over loudspeakers located inside the testing booth. Sound-field levels were
determined in the absence of the listener with a Bruel and Kjaer ½-in microphone.

Speech stimuli consisted of the prompt, “Touch the ____,” and a target word (e.g.,
ball). Target words consisted of 60 monosyllabic CVC nouns for the test phase and an
additional 20 monosyllabic nouns for the training phase. A subset of test items (n = 40)
consisted of minimal pairs, differing only in their initial or their final phoneme (see Table
3.1). The degree of confusability between critical contrasts (cf. Miller & Nicely, 1954) was
not significantly different for initial- and final-contrast pairs, t(19) = -.40, p = .69. Because
word familiarity is positively related to word intelligibility (Elliott et al., 1983; Metsala,
1997; Rosenweig & Postman, 1957), we used three measures of word frequency – Thorndike
and Lorge’s general word frequency count (1944), Thorndike and Lorge’s (1944) juvenile
count, and Kucera and Francis’s (1967) word count (see Table 1) – to confirm that maximal-,
initial-, and final-contrast targets did not differ in familiarity. The comparison between
initial- and maximal-contrast trials approached statistical significance on Thorndike and
Lorge's juvenile count, $t(33) = -1.809$, $p = .08$. The remaining comparisons were not significantly different, $p > .1$ in all cases.

Table 3.1: Word frequencies for initial-, final-, and maximal-contrast trials in log units using three measures of familiarity.

<table>
<thead>
<tr>
<th>IC Targets</th>
<th>FC Targets</th>
<th>MC Targets</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TL</td>
<td>TL-J</td>
</tr>
<tr>
<td>BEAR</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>BELL</td>
<td>2.81</td>
<td>2.54</td>
</tr>
<tr>
<td>BONE</td>
<td>2.85</td>
<td>2.65</td>
</tr>
<tr>
<td>BOOK</td>
<td>3.00</td>
<td>N/A</td>
</tr>
<tr>
<td>CAN</td>
<td>3.00</td>
<td>3.00</td>
</tr>
<tr>
<td>CANE</td>
<td>2.20</td>
<td>1.41</td>
</tr>
<tr>
<td>CHAIR</td>
<td>2.85</td>
<td>2.85</td>
</tr>
<tr>
<td>FAN</td>
<td>2.45</td>
<td>2.26</td>
</tr>
<tr>
<td>HOOK</td>
<td>2.43</td>
<td>2.43</td>
</tr>
<tr>
<td>HOSE</td>
<td>2.06</td>
<td>1.46</td>
</tr>
<tr>
<td>LOCK</td>
<td>2.58</td>
<td>2.44</td>
</tr>
<tr>
<td>NAIL</td>
<td>2.57</td>
<td>2.39</td>
</tr>
<tr>
<td>NOSE</td>
<td>2.85</td>
<td>2.69</td>
</tr>
<tr>
<td>PAIL</td>
<td>2.34</td>
<td>N/A</td>
</tr>
<tr>
<td>PHONE</td>
<td>1.95</td>
<td>0.00</td>
</tr>
<tr>
<td>RAIN</td>
<td>3.00</td>
<td>N/A</td>
</tr>
<tr>
<td>SEAL</td>
<td>2.49</td>
<td>2.34</td>
</tr>
<tr>
<td>SHELL</td>
<td>2.61</td>
<td>2.61</td>
</tr>
<tr>
<td>SOCK</td>
<td>1.95</td>
<td>1.15</td>
</tr>
<tr>
<td>WHEEL</td>
<td>2.85</td>
<td>2.62</td>
</tr>
</tbody>
</table>

Average 2.59 2.23 1.44 2.53 2.36 1.35 2.70 2.63 1.61
St. Dev. .35 .78 .63 .36 .52 .39 .44 .51 .69

Note: TL stands for Thorndike and Lorge (1944) general frequency count; TL-J stands for Thorndike and Lorge's (1944) juvenile count; and KF stands for Kucera and Francis's (1967) word count. Independent t-tests revealed no significant differences among the three trial types for each word frequency measure, $p > .05$ in all cases.

The sentences had a fundamental frequency of approximately 252 Hz and an average duration of 1289 ms. Independent t-tests confirmed that initial-, final-, and maximal-contrast
sentences did not differ significantly in fundamental frequency or duration, t(38) < 1.5 in all cases (see Table 3.2). Sentences were presented at approximately 44 dB (A scale). Root-mean-square (RMS) values were calculated and adjusted such that each sentence was presented at an equal RMS value following the procedure described in Schneider et al. (2000). Each age group was tested at the SNR that had yielded approximately 85% correct performance in Chapter 2: -28 dB for 5-year-olds, -30 dB for 9-year-olds, and -33 dB for adults. SNRs for the training phase were made 5 dB more favorable than those in the low-noise conditions: -23 dB, -25 dB, and -28 dB for 5-year-olds, 9-year-olds, and adults, respectively.

Table 3.2: Fundamental frequency and duration for initial- (IC), final- (FC), and maximal-contrast (MC) sentences.

<table>
<thead>
<tr>
<th></th>
<th>Frequency (Hz)</th>
<th></th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Range</td>
</tr>
<tr>
<td>IC</td>
<td>251.11</td>
<td>12.84</td>
<td>232.37-274.37</td>
</tr>
<tr>
<td>FC</td>
<td>251.36</td>
<td>10.20</td>
<td>237.40-279.18</td>
</tr>
<tr>
<td>MC</td>
<td>253.74</td>
<td>18.99</td>
<td>213.10-306.84</td>
</tr>
</tbody>
</table>

Note: Independent t-tests confirmed no significant differences between trial types for fundamental frequency (p > .6 in all cases) or for duration (p > .1 in all cases).

Visual stimuli consisted of 80 black-and-white line drawings of familiar, concrete objects from the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997), Snodgrass and Vanderwart (1980), and a local artist. A post-test was devised to ensure that 5-year-olds'
performance on the speech-in-noise task was not due to the unfamiliarity of some pictures. After performing the speech-in-noise task, the experimenter showed children 4-alternative pictorial arrays (printed on 8 ½ - 11 in. sheets of paper) that corresponded to trials in which the target was misidentified. In quiet conditions, the experimenter instructed children to “Touch the [target].” For inclusion in the study, children could not misidentify more than one picture during the post-test. Only two additional 5-year-olds (one boy, one girl) were excluded from the sample for failing to reach this criterion.

Procedure.

All participants were tested individually. A trial, which was initiated by means of a button box located inside the testing booth, consisted of the simultaneous presentation of vocal stimuli (sentence and noise) and visual stimuli. Multitalker babble began with the onset of a sentence and terminated when a sentence ended. Auditory stimuli were accompanied by an array of four different images, one appearing in each corner of the touch screen. The locations of targets and foils were randomly selected on each trial. Feedback for correct performance consisted of the target picture flashing in the middle of the screen. Incorrect selections resulted in the screen going blank.

Participants were presented with three trial types. Trials in which all four alternatives did not share any common phonemes (e.g., house, boat, cake, mop) were termed maximal-contrast trials (n = 20). Initial-contrast trials (n = 20) consisted of a minimal pair differing in the initial phoneme (e.g., hook, book), and two foils that were phonologically dissimilar from the pair (e.g., rope, sun). By contrast, final-contrast trials (n = 20) consisted of a minimal pair differing in the final phoneme (e.g., doll, dog), and two foils that were phonologically dissimilar from the pair (e.g., bike, chair). Each member of a minimal pair served as a target word during the test session. A picture could not appear as a foil on the trial immediately following its presentation as a target. Half of the participants received a mixed presentation of trial types; the remaining participants received a blocked presentation. Trials were randomly ordered in both presentation conditions. Participants in the mixed condition took a
short break after completing half of the trials, and participants in the blocked condition rested briefly after each block of trials. Equal numbers of each trial type were presented to 5-year-olds for a total of 60 trials. To discourage use of an exclusionary strategy, older listeners performed additional initial-contrast (n = 40) and final-contrast (n = 40) trials for a total of 100 trials. Participants in each age group were counterbalanced as closely as possible across sex and order of lists.

The instructions were tailored to the age of participants. The experimenter explained to 5-year-old children that if they only hear part of a word, they should choose the picture that sounds similar to what they hear (e.g., “If you hear ‘irt’ and there are pictures of a spoon, frog, shirt, and tree on the screen, you should pick the shirt since ‘shirt’ sounds the most like ‘irt’.”). Older children and adults were told that the pictures were identifiable by basic-level terms. For example, a picture of a shirt would be identified by the word “shirt,” not “button-down” or “clothing.” No other explicit strategies were provided.

The test session consisted of a training phase and a test phase. In the training phase, listeners had to correctly identify 4 consecutive targets within 20 trials or obtain at least 50% correct performance (10 out of 20 correct selections). On average, listeners achieved the training criterion in 7.77 trials. Only three listeners (one 5-year-old, one 9-year-old, and one adult) required all 20 trials. After reaching the criterion, participants proceeded immediately to the test phase. Adults and older children initiated trials at their preferred pace. The experimenter initiated trials for 5-year-olds when she judged them to be ready and attentive. She remained in the testing booth during the entire session, offering verbal reinforcement and encouragement when appropriate. An additional motivational technique was used with the 5-year-olds. After every four trials, children received a colored sticker to place in an “incomplete” black-and-white picture.

Results

Because 9-year-olds’ and adults’ performance may have improved with additional initial- and final-contrast trials, performance on the first presentation of a target was
compared with performance on its subsequent appearance. Using percent correct as the dependent variable, a repeated-measures ANOVA with target order and trial type as within-subject variables and presentation (mixed vs. blocked) and age as between-subjects variables revealed that listeners improved significantly on the second presentation of a target, $F(1, 190) = 10.93$, $p < .001$ ($M = 69.85\%$ and $M = 72.43\%$, for the first and second occurrence, respectively). This effect was qualified by an order x trial interaction, $F(1, 140) = 5.16$, $p = .036$. Although the improvement for final-contrast trials was statistically significant, $t(143) = 4.02$, $p < .001$, the improvement observed for initial-contrast trials was not, $t(143) = 1.00$, $p = .316$. Nevertheless, older listeners’ performance on the first presentation of initial- and final-contrast trials was used in the main analysis. Thus, each listener’s word identification performance is based on 60 trials (equal numbers of initial-, final-, and maximal-contrast trials).

Preliminary analysis revealed that presentation mode (blocked or mixed) did not affect performance significantly, $F(1, 210) = .001$, $p = .975$, and did not interact with age, $F(2, 210) = 1.55$, $p > .2$. Thus, presentation mode was excluded from the main analysis.

Figure 3.1 depicts the performance of 5-year-olds, 9-year-olds, and adults on maximal-, initial-, and final-contrast trials. A repeated-measures ANOVA with trial type as a within-subject variable and age as a between-subjects variable revealed a main effect of trial type, $F(1.99, 423.43) = 279.68$, $p < .00001$. (Degrees of freedom were adjusted using Greenhouse-Geisser corrections because the data were not homogenous for final-contrast trials.) Maximal-contrast trials ($M = 82.85\%;$ $SD = 10.46$) were identified more accurately than initial-contrast trials ($M = 73.68\%;$ $SD = 10.23$), $t(215) = 10.64$, $p < .001$, and final-contrast trials ($M = 62.73\%;$ $SD = 11.27$), $t(215) = 23.95$, $p < .001$. Moreover, listeners performed significantly better on initial- than on final-contrast trials, $t(215) = 12.20$, $p < .001$. A main effect of age was observed, $F(2, 213) = 20.10$, $p < .001$, reflecting 5-year-olds’ poorer performance relative to that of 9-year-olds, Games-Howell $p < .001$, and adults, Games-Howell $p < .001$ (see Table 3.3). Furthermore, there was a two-way interaction between trial
type and age, $F(3.98, 423.48) = 4.20, p = .002$. Although 5-year-olds’ performance on maximal- and final-contrast trials was significantly different from that of 9-year-olds and adults—$F(2, 213) = 26.55, p < .001$, and $F(2, 213) = 10.14, p < .001$, respectively—listeners performed comparably on initial-contrast trials, $F(2, 213) = 1.83, p = .16$.

Figure 3.1: Performance of 5-year-olds, 9-year-olds, and adults on maximal-, initial-, and final-contrast trials.
Table 3.3: Performance (percent correct) of 5-year-olds, 9-year-olds, and adults on maximal-contrast (MC), initial-contrast (IC), and final-contrast (FC) trials.

<table>
<thead>
<tr>
<th></th>
<th>MC Trials</th>
<th>IC Trials</th>
<th>FC Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td>76.39 (10.18)</td>
<td>71.81 (10.95)</td>
<td>58.06 (12.38)</td>
</tr>
<tr>
<td>9-year-olds</td>
<td>84.93 (8.94)</td>
<td>74.58 (9.45)</td>
<td>65.35 (9.05)</td>
</tr>
<tr>
<td>Adults</td>
<td>87.22 (9.03)</td>
<td>74.65 (10.12)</td>
<td>64.79 (10.76)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

To ascertain whether listeners chose similar-sounding foils when they erred, the percentage of minimally contrastive errors on initial- and final-change trials was compared using a repeated-measures ANOVA with trial type as a within-subject factor and age as a between-subjects factor. [As in the main analysis, presentation mode, whether mixed or blocked, did not significantly affect performance, \( F(1, 210) = 2.41, p = .122 \), and did not interact with age, \( F(2, 210) = 1.59, p = .207 \).] Listeners were more likely to select phonologically similar foils on final-contrast (\( M = 73.32\%; SD = 18.11 \)) than on initial-contrast trials (\( M = 49.41\%; SD = 24.29 \), \( F(1, 213) = 159.12, p < .0001 \). One-sample t-tests using 33% as the test mean (corresponding to a 1-in-3 chance of choosing any foil) confirmed that similar-sounding (initial- and final-contrast) foils were selected significantly more often than were phonologically unrelated foils, \( t(215) = 9.73, p < .001 \), and \( t(215) = 32.46, p < .0001 \), for initial- and final-contrast trials, respectively. Even 5-year-olds selected similar- rather than dissimilar-sounding alternatives when they erred, \( t(71) = 4.04, p < .001 \), and \( t(71) = 15.57, p < .001 \), for initial- and final-contrast trials, respectively. However, 9-year-olds (\( M = 65.64\%; SD = 16.27 \)) and adults (\( M = 63.05\%; SD = 15.11 \)) were more likely to select phonologically similar foils than 5-year-olds were (\( M = 55.40\%; SD = 15.32 \), \( F(2, 213) = 8.35, p < .001 \). An age x trial interaction approached conventional levels of
significance, $F(1, 213) = 2.85$, $p = .06$, reflecting a greater age difference for final- than for initial-contrast trials (see Table 3.4).

Table 3.4: Percentage of errors committed by selecting a similar-sounding foil.

<table>
<thead>
<tr>
<th></th>
<th>IC Trials</th>
<th>FC Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td>44.54 (23.78)</td>
<td>66.26 (17.87)</td>
</tr>
<tr>
<td>9-year-olds</td>
<td>55.74 (25.44)</td>
<td>75.54 (17.11)</td>
</tr>
<tr>
<td>Adults</td>
<td>47.95 (22.50)</td>
<td>78.16 (17.35)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

The likelihood of selecting a target or its similar-sounding foil (i.e., correct choice or a reasonable alternative) provides further perspectives on the process of word identification. As shown in Table 3.5, the percentage of trials on which targets or phonologically similar foils were selected was greater on final-contrast trials ($M = 89.65\%$, $SD = 8.13$) than on initial-contrast trials ($M = 86.64\%$, $SD = 7.99$). A Kruskal-Wallis test confirmed that older listeners selected either of two phonologically similar words (i.e., target or foil) more often than 5-year-olds did for initial-contrast trials, $\chi^2(2) = 12.42$, $p = .002$, and final-contrast trials, $\chi^2(2) = 24.70$, $p < .001$.

Table 3.5: The likelihood (%) of selecting a target or its minimally contrastive foil.

<table>
<thead>
<tr>
<th></th>
<th>IC Trials</th>
<th>FC Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td>84.10 (8.89)</td>
<td>85.49 (9.35)</td>
</tr>
<tr>
<td>9-year-olds</td>
<td>88.82 (7.62)</td>
<td>91.39 (6.40)</td>
</tr>
<tr>
<td>Adults</td>
<td>87.15 (6.65)</td>
<td>92.22 (6.55)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.
To confirm young children’s ability to discriminate the initial- and final-contrast pairs in quiet conditions, an independent sample of 5-year-olds (n = 17; 8 boys, 9 girls) received a randomized order of initial- and final-contrast trials. Children performed near ceiling, with a slight advantage for initial-contrast trials (M = 99.71%) over final-contrast trials (M = 92.06%; SD = 7.08).

Discussion

Children and adults were required to identify words in noise by selecting one of four pictures. To compensate for age-related changes in stimulus audibility, signal-to-noise ratios (SNRs) were adjusted accordingly. Thus, age differences would reflect the contribution of other factors such as attention, auditory processing, phonological knowledge, or problem-solving ability. There were a number of similarities across age. Children and adults identified words more accurately when the alternatives were maximally contrastive (e.g., ball vs. house) than when they were phonologically similar (e.g., hose vs. nose). Moreover, listeners of all ages performed more poorly when phonologically similar alternatives contrasted in their final segments (e.g., cap, cat) than in their initial segments (e.g., nail, pail). Finally, greater stimulus variability—presenting trials in a mixed order rather than a blocked order—did not produce performance decrements for any age group.

Age-related differences were also apparent. Children of 5 performed more poorly than did 9-year-olds and adults, who did not differ from one another. In particular, 5-year-olds performed more poorly than older listeners when the foils differed in their final segments and when the foils were maximally contrastive. Identification accuracy was comparable across age, however, when the foils differed in their initial segments.

What are the implications of these findings? Do adult-child differences stem from different word recognition strategies? If 5-year-olds used more holistic than analytic processing strategies (Jusczyk, 1992; Metsala, 1997; Treiman & Breaux, 1982; Walley et al., 1986; Walley, 1993), they would have distributed their attention evenly over a word rather than focusing on a particular segment. Hence, 5-year-olds should have performed more
poorly than 9-year-olds and adults on trials with minimal pairs. At all age levels, however, listeners performed comparably on words differing in initial position. Previous research supports the notion that young children attend to the initial portions of words. In optimal listening situations, 5-year-olds more readily detect word-initial than word-final mispronunciations in a story (Cole & Perfetti, 1980; Walley, 1987).

In the present study, 5-year-olds’ errors also revealed their phonetic segmentation skills. Young children chose similar-sounding alternatives more often than chance when they erred on initial- and on final-contrast trials. Moreover, 5-year-olds selected phonologically similar foils more frequently on final- than on initial-contrast trials. This pattern of responding implies that young children accorded priority to word-initial input. Word onset information uniquely specifies a target word presented with an initial-contrast foil (nail/pail), but it does not differentiate between words differing only in their final portions (dog/doll). The marked differences in 5-year-olds’ performance between initial- and final-contrast trials could not have resulted from an even distribution of attention over the target word, or holistic strategy (e.g., Metsala, 1997; Walley, 1993).

Further evidence of early analytic skill arises from investigations of rhyming and alliteration. Preschoolers show some ability to detect rhyme and alliteration, both of which require understanding that the initial sound of a word is distinct from the sounds that follow (Knafle, 1974; Lenel & Cantor, 1981; Maclean, Bryant, & Bradley, 1987). Interestingly, Lenel and Cantor (1981) found that 5-year-olds experienced more difficulty identifying one of two words that rhymed with a target when a consonant common to the target and the non-rhyming foil was in initial position (toes, tent, nose) than when it was in final position (pin, sun, gun). In other words, young children seemed to recognize the similarity between the initial sounds of the target and the non-rhyming alternative. Research on young children’s perception of subsyllabic structure supports this notion (Kirtley, Bryant, MacLean, & Bradley, 1989). In the case of CVC monosyllabic words, even prereaders are sensitive to changes that occur in a “natural” subsyllabic unit, such as the onset. They experience
difficulty, however, with changes in the final sound, which is part of the rime (i.e., the vowel and the final consonant).

Young children do not exhibit effective processing of word-initial information in all contexts (Walley, 1987, 1988). The presumption is that contextual constraints facilitate the deployment of processing resources to word-initial information. For example, 5-year-olds detect word-initial mispronunciations more accurately than word-final mispronunciations when the target words are embedded in familiar nursery rhymes (Cole, 1981) or in stories rich in contextual cues (Cole & Perfetti, 1980; Walley, 1987). In the present study, however, the sentences provided relatively few contextual cues (Touch the house), decreasing the likelihood that young children’s focus on word onset was influenced by contextual constraints.

Although our findings indicate that children engage in analytic speech processing, they also indicate that young children’s processing is less refined than that of older listeners. For example, young children experienced disproportionate difficulty with final-contrast trials, which is consistent with comparable difficulties in quiet conditions (Hnath-Chisolm et al., 1998; Menary et al., 1982; Morgan, 1984). Indeed, an additional sample of 5-year-olds tested in the present study performed more poorly on final-contrast trials than on initial-contrast trials when both were presented in quiet. Why do children experience such difficulty with word-final contrasts? The masking of coarticulatory cues in the final portion of words seems to have greater impact on young children than on older listeners (Nittrouer & Studdert-Kennedy, 1987). Young children’s reliance on coarticulatory cues implies that their phonological representations of words may involve larger units than those of adults (Gerken et al., 1995). Kirtley et al. (1989), for example, found that 5-year-olds have difficulty separating vowels from final consonants. Put another way, words contrasting in elements of the rime (i.e., final-contrast pairs) are not as easily identified as those contrasting in the onset (i.e., initial-contrast pairs).
Although 5-year-olds chose similar-sounding alternatives more often than chance, older listeners selected a target or its phonologically similar counterpart more frequently than did 5-year-olds (see Tables 3.4 & 3.5). If adult and child listeners had similar phonological representations of words, then targets or similar-sounding foils would be selected at comparable rates regardless of age. Interestingly, 5-year-olds' and adults' selections were more comparable on initial- than on final-contrast trials, implying that young children's lexicons may be organized into subsyllabic units but not into discrete phonetic units (Kirtley et al., 1989).

Surprisingly, maximal-contrast trials posed greater difficulty than expected for 5-year-olds. In Chapter 2 no age differences were observed on trials with maximally contrastive alternatives, even though listeners were tested at the same SNRs as those in the present study (i.e., SNRs adjusted to equalize age-related differences in stimulus audibility). What accounts for the differences between 5-year-olds and older listeners and the absence of such differences in Chapter 2? As shown in Table 3.6, 5-year-olds' performance on maximally contrastive trials in the first block of trials was comparable to that of 9-year-olds and adults, $F(2, 33) = .80, p = .46^1$. Also note that performance in the first block of the present study was comparable to performance on the first 20 trials of Chapter 2. In the present study, 9-year-olds and adults showed improved performance on maximally contrastive trials in later trial blocks, indicating transfer of learning from earlier to later trials. By contrast, the 5-year-olds did not show comparable gains on later trials. This finding contrasts with previous evidence of similar transfer for older and younger listeners over the test session (Chapter 2; see Table 3.6). In the present study, the difficulty that young children experienced with minimal pairs may have interfered with transfer of learning to maximally contrastive trials. Older listeners in the present study showed greater

---

1 Note that the age difference observed on final-contrast trials cannot be explained by transfer of learning. A 3 x 3 ANOVA revealed no significant interaction between age and block, $F(4, 99) = .50, p = .74$. 

46
improvement than those in Chapter 2, an effect that may be attributed to the greater number of trials in the present study.

Table 3.6: Age-related differences in performance (percent correct) on the maximally contrastive block of trials as a function of when the block occurred.

<table>
<thead>
<tr>
<th></th>
<th>Present study</th>
<th>Chapter 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1st block</td>
<td>2nd block</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>80.83 (9.49)</td>
<td>81.04 (8.97)</td>
</tr>
<tr>
<td>9-year-olds</td>
<td>80.42 (8.38)</td>
<td>82.71 (9.09)</td>
</tr>
<tr>
<td>Adults</td>
<td>85.00 (11.28)</td>
<td>81.04 (9.09)</td>
</tr>
<tr>
<td></td>
<td>3rd block</td>
<td>2nd half</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>75.00 (9.77)</td>
<td>85.21 (7.44)</td>
</tr>
<tr>
<td>9-year-olds</td>
<td>87.50 (9.17)</td>
<td>84.58 (7.65)</td>
</tr>
<tr>
<td>Adults</td>
<td>89.58 (7.22)</td>
<td>88.12 (5.67)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

As noted, common processes were also evident across age. First, listeners identified target words more accurately when they were presented with maximally contrastive foils than with phonologically similar foils. One might argue that maximal-contrast targets were more identifiable than initial- and final-contrast targets for reasons other the presence of phonologically similar foils. Note, however, that the three trial types did not differ in word frequency (see Table 3.1) or in fundamental frequency or duration (see Table 3.2). Moreover, there is no indication that neighborhood density has consequences for performance in forced-choice contexts (Sommers et al., 1997; but see Tanenhaus et al., 2000). Thus, it is likely that poorer identification of minimal pairs was entirely due to the inclusion of phonologically similar foils. Second, listeners performed more accurately on initial- than on final-contrast trials. Word onsets contain richer acoustic information compared with final segments (Redford & Diehl, 1999), which would contribute to listeners' difficulty with final-contrast trials. In line with previous research (e.g., Marslen-Wilson, 1987; Nooteboom, 1981; Norris, 1994; Taft & Forster, 1975), listeners use initial segments to
determine word identity. Thus, they experience greater difficulty when word onset does not uniquely specify a target (i.e., final-contrast trials), than when it does (i.e., initial-contrast trials).

In sum, the presence of minimal pairs produced similarities and differences in younger and older listeners' perception of speech in background babble. Once age-related differences in stimulus audibility were minimized, young children experienced disproportionate difficulty identifying words contrasting in their final phonemes, but they performed much like older listeners on word-initial contrasts. Young children also exhibited poorer than expected identification of words that were maximally contrastive from foils, perhaps because of the inclusion of more difficult items in the same test session. Emerging segmentation skills may be reflected in young children’s relatively good performance on initial-contrast trials.
Chapter 4
Children's use of contextual cues in degraded listening environments

Introduction

When identifying words in spoken utterances, listeners typically monitor contextual cues as well as acoustic-phonetic information. At times, acoustic-phonetic cues are insufficient for the identification of particular words because of limitations in the signal (e.g., poor articulation on the part of the speaker) or listening environment (e.g., the presence of competing sounds or reverberation). In such circumstances, contextual cues—lexical, semantic, and syntactic—can help listeners decode a message. For example, upon hearing the word “eating,” listeners expect a food-related noun to follow. Although adults derive considerable benefit from such contextual cues, especially in challenging listening environments (e.g., Bilger et al., 1984; Dubno et al., 2000; Kalikow et al., 1977; Pichora-Fuller et al., 1995), the extent to which children do so is unclear.

A number of investigators have examined children's use of semantic contextual cues to decode an altered signal. For example, Cole and Perfetti (1980) presented mispronounced words in high-predictability and low-predictability sentences within a story. They found that 4-year-olds were better at detecting mispronounced words in high-predictability than in low-predictability sentences, indicating that very young children profit from such contextual cues. Craig et al. (1993) presented their target words in high- or low-predictability contexts. They altered the to-be-identified word by eliminating the later portions of the word. In such “gating tasks” (Grosjean, 1980, 1985), participants engage in repeated attempts to identify the target word from successively longer portions, or gates. Craig et al. (1993) found that 8- to 10-year-old children required shorter gates (i.e., less acoustic information) to identify target words in high-predictability than in low-predictability sentences. By contrast, 5- to 7-year-olds required as much acoustic information with high-predictability as with low-predictability contexts (Craig et al., 1993). Because of young children’s limited phonetic
representation skills (Walley, 1988, 1993), the gating task may underestimate their ability to use semantic cues.

Tyler and Marslen-Wilson (1981) examined children’s use of semantic contextual cues by altering the semantic adequacy of sentences. Target words were presented in semantically appropriate (“John has to go back home”) or anomalous (“John had to sit on the shop”) utterances. Children 5-10 years of age more readily detected the target words in semantically appropriate than in anomalous sentences, confirming their ability to profit from the semantic context.

Other investigators have focused on word identification in adverse listening conditions involving background noise. Nittrouer and Boothroyd (1990) presented short sentences varying in semantic adequacy in a background of spectrally-matched noise. Listeners’ task was to repeat these sentences verbatim. Although semantic cues facilitated performance, 4- to 6-year-olds did not benefit from such cues to the same extent as did young and elderly adults. Elliott (1979) investigated children’s use of contextual cues by means of the Speech Perception in Noise, or SPIN, task (Bilger et al., 1984; Kalikow et al., 1977). Participants in the SPIN task are required to repeat the final word of high- and low-predictability sentences presented in multitalker babble. Elliott (1979) found that children 9-13 years of age derived less benefit from contextual cues in high-predictability sentences than did older listeners. Interestingly, the ability to profit from semantic contextual cues (as assessed by the SPIN task) continues to improve into older adulthood (Pichora-Fuller et al., 1995; but see Dubno et al., 2000). Pichora-Fuller et al. (1995) suggest that elderly adults may compensate for a declining perceptual system by relying on contextual cues to a greater extent than younger listeners.

In principle, children whose perceptual system is not fully developed could profit from heavy reliance on contextual cues in noisy backgrounds. The available evidence indicates (e.g., Elliott, 1979; Nittrouer & Boothroyd, 1990), however, that they do not do so. Why not? One would expect young children to perform more poorly than adults on difficult
listening tasks simply on the basis of their higher thresholds in quiet and in noise (e.g., Schneider et al., 1986, 1989). For example, 5-year-old children required signal-to-noise ratios (SNRs) that were 5 dB more favorable than those of adults to achieve comparable identification in simple low-context sentences (e.g., “Touch the dog”; Chapter 2). If SNRs are not adjusted to compensate for age-related sensory differences, then it is impossible to separate the contributions of sensory and cognitive factors. In addition, the test materials in Nitttrouer and Boothroyd (1990) and Elliott (1979) were less than optimal for young children. For example, some of Nitttrouer and Boothroyd’s (1990) high-predictability sentences (e.g., “Tough guys sound mean”; “Dull paint won’t shine”) would have confused many, if not all, young children. Moreover, the requirement of repeating the final word of a test sentence (Elliott, 1979), or the entire sentence (Nitttrouer & Boothroyd, 1990), poses disproportionate difficulty for young children (Elliott et al., 1979). In short, meaningful comparisons of children’s and adults’ use of contextual cues in noise depend on comparable task demands and listening conditions across age.

In the present study, we evaluated the contribution of contextual cues to word recognition in noise by 5- and 9-year-old children and adults. To this end, we devised age-appropriate low- and high-context sentences in which the target word was a highly familiar monosyllabic noun. Pilot testing revealed that 5-year-old children were able to complete high-context sentence stems (e.g., My wagon has a broken ____ when presented with pictorial alternatives (e.g., wheel, tree, brush, cat) in quiet. Task difficulty was minimized by using a four-alternative picture-pointing response (following Chapter 2). To minimize differences in stimulus audibility across age, we established SNRs that yielded comparable performance levels (78%) on low-context sentences (e.g., We looked at the bread). If older children and adults are better able to profit from contextual cues than are younger children, then we would expect a statistical interaction between age and context. By contrast, the absence of such an interaction would be consistent with comparable use of contextual cues across age. A secondary goal was to evaluate the widely held view that noise has more
adverse perceptual consequences on young children than it does on older listeners (Elliott, 1979; Mills, 1975). Although the noise increments in Chapter 2 had comparable effects across age, the test sentences were structurally simpler than those in the current study. Accordingly, we tested another group of listeners at each age level with 2 dB of additional noise. As before, an interaction between age and noise level would be expected if the noise increment had disproportionate consequences on younger, but not older listeners.

Method

Participants.

The participants were 48 5-year-olds (range: 5;0-5;6, M = 5;3 years), 48 9-year-olds (range: 9;0-9;6, M = 9;3 years), and 48 adults (range: 19-25, M = 20.7 years), with equal numbers of males and females in each age group. None of the children had experienced frequent ear infections or pressure-equalizing tubes, and none had colds on the day of testing. All listeners were monolingual speakers of English, with the exception of three adults, who had learned English by 6 years of age and attended elementary school in North America. Additional participants were excluded because of experimenter error (one 5-year-old, two 9-year-olds, one adult), failure to complete the test session (one 5-year-old), failure to meet the training criterion (four 5-year-olds), failure to follow instructions (one 5-year-old, one 9-year-old), inattentiveness (three 5-year-olds, one 9-year-old), or apprehensiveness (two 5-year-olds).

Stimuli and Apparatus.

Testing occurred in a double-wall sound-attenuating booth, 3 x 2.8 x 2 m in size. Participants were seated facing a non-glare touch screen monitor (Goldstar 1465DLs) 33 cm x 33 cm. Loudspeakers (KEF Model 101) were 45° to the left and right of the participant (distance of 70 cm) at approximate ear level. Sentence files and babble files were converted to analog form using Tucker-Davis digital-to-analog converters under the control of a computer with a Pentium processor. Sentence and babble amplitudes were controlled separately by means of programmable attenuators. After mixing, the combined signals were
amplified (SAE 2600) and presented over loudspeakers located inside the testing booth. Sound-field levels were determined in the absence of the listener by means of a Bruel and Kjaer ½-in microphone.

A large corpus (n = 188) of high-context sentences with concrete, monosyllabic nouns in sentence-final position was generated. Sentences were grammatically simple, short (5 to 10 syllables in length), and familiar to 5-year-olds in their words and situations. To assess the degree to which high-context stems primed final target words, an independent sample of 361 adults (289 women, 72 men; mean age = 22.3 years; range 18-54) completed subsets of high-context sentence stems (e.g., Mice like to eat ____ ) with the word that first came to mind. Each participant completed approximately 47 sentence stems by means of a paper-and-pencil task, yielding approximately 90 observations per sentence. The data from 19 additional adults were excluded for failing to complete over 90% of the stems (4 women, 1 man), learning English after age 14 (10 women, 2 men), or providing blatantly foolish answers (2 men). From the initial corpus, 60 sentences were selected to serve as practice (n = 20) and test (n = 40) items. The mean number of syllables in high-context sentences was 7.28. On average, adults completed each high-context stem with the target word 52.5% of the time (range 3-95%). Low-context sentences (n = 60, 40 test items, 20 practice items) were generated with these target words. Low-context sentences could take one of five forms:

(1) “[Pronoun] looked at the [target]”; (2) “[Pronoun] talked about the [target]”; (3) “[Pronoun] read about the [target]”; (4) “[Pronoun] heard about the [target]”; and (5) “[Pronoun] pointed at the [target].” Thus, low-context sentences contained 5 or 6 syllables (M = 5.77). High- and low-context sentences are listed in the Appendix.

A four-alternative, forced-choice task was used to assess 5-year-olds’ ability to predict target words from the high-context stems. Subsets (n = 20) of prerecorded high-context sentence stems were presented at 44 dB (A) to an independent sample of 29 5-year-olds (14 girls, 15 boys; M = 5;3 years). Two additional children (one girl and one boy) were excluded because of inattentiveness. Sentence stems were accompanied by four pictures.
appearing in different corners of a touch screen. The four alternatives included the target and three foils that depicted words semantically and phonologically dissimilar from the target word. Visual feedback in the form of a flashing picture was provided for correct responses. The experimenter explained to children that they would hear a lady tell them something, but that the computer would "chop off" the final word. Their task was to select the picture of what the lady was supposed to say. To ensure that they understood the task, children performed two practice trials with photocopied arrays of pictures from the practice plates of the Picture Peabody Vocabulary Test (Dunn & Dunn, 1997). For example, the experimenter presented pictures of a table, doll, car, and man and said, "We put dinner on the _____." Children selected the target word on 98.8% of test trials, indicating that the high-context sentence stems cued the target test words.

High- and low-context sentences were produced by a vocally-trained woman and digitized at a rate of 20 kHz using a 16-bit Tucker Davis (DD1) analog-to-digital converter. The babble portion of SPIN forms used by Pichora-Fuller et al. (1995) was similarly digitized and stored. Sentences had an average fundamental frequency of 231 Hz and an average duration of 1968 ms. As shown in Table 4.1, the duration of high-context sentences was greater than that of low-context sentences, \( t(78) = -5.56, p < .001 \), but the duration of target words in low-context sentences was longer than that of high-context sentences, \( t(78) = 2.60, p = .01 \). It is likely that the longer durations of low-context target words reflect the speaker's intuitive compensation for low redundancy in the sentence (e.g., Fowler, 1988; Lieberman, 1963). The mean fundamental frequency of target words did not differ as a function of contextual condition, \( t(78) = 1.159, p = .25 \). Similarly, the mean fundamental frequency of entire low- and high-context sentences did not differ significantly, \( t(78) = .05, p = .96 \). Sentences were presented at approximately 44 dB (A). Root-mean-square (RMS) values were calculated and adjusted such that each sentence was presented at an equal intensity following the procedure described in Schneider et al. (2000). SNR was varied by adjusting the level of babble (\( F_0: 185 \) Hz). Pilot-testing established the SNR at which each
age group achieved approximately 78% correct performance on low-context sentences: -24 dB for 5-year-olds, -27 dB for 9-year-olds, and -30 dB for adults. These levels were designated low-noise conditions. High-noise conditions were created for each age group by decreasing the SNR in the low-noise condition by 2 dB (by adding 2 dB of noise), resulting in SNRs of -26 dB for 5-year-olds, -29 dB for 9-year-olds, and -32 dB for adults. SNRs for the training phase were set 5 dB lower than those in the low-noise conditions: -19 dB, -22 dB, and -25 dB for 5-year-olds, 9-year-olds, and adults, respectively.

Table 4.1: Fundamental frequency and duration of low- and high-context sentences.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sentence</td>
</tr>
<tr>
<td><strong>Low Context</strong></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>231.24</td>
</tr>
<tr>
<td>SD</td>
<td>13.39</td>
</tr>
<tr>
<td>Range</td>
<td>198.30-257.65</td>
</tr>
<tr>
<td><strong>High Context</strong></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>231.12</td>
</tr>
<tr>
<td>SD</td>
<td>9.36</td>
</tr>
<tr>
<td>Range</td>
<td>206.92-251.59</td>
</tr>
</tbody>
</table>

Visual stimuli consisted of 60 black-and-white line drawings of familiar, concrete objects corresponding to the target items. Images were gathered from various sources, including Cywowicz, Friedman, Rothstein, and Snodgrass (1997), Snodgrass and Vanderwart (1980), the Peabody Picture Vocabulary Test (Dunn & Dunn, 1997), and a local artist.
Procedure.

All participants were tested individually. A trial, which was initiated by means of a button box located inside the testing booth, consisted of the simultaneous presentation of vocal stimuli (sentence and noise) and visual stimuli. Sentences in noise were accompanied by an array of four different images, one appearing in each corner of the touch screen. The multitalker babble began with the onset of the sentence and terminated when the sentence ended. The visual array included the target image and three designated foils that were phonetically and semantically dissimilar from the target. The locations of targets and foils were randomly selected on each trial. Reaction times accurate to the millisecond were recorded automatically. Feedback for correct performance consisted of the target flashing in the middle of the screen. Incorrect selections resulted in the screen going blank.

The instructions were tailored to the age of participants. Listeners were told that the last word of the sentence would correspond to one of the pictures on the screen. The experimenters explained to 5-year-old children that they should choose the picture that matched the last word that the lady said. Moreover, 5-year-olds were told that if they only hear part of a word, they should choose the picture that sounds similar to what they hear (e.g., "If you hear 'irt' and there are pictures of a can, plate, shirt, and boat on the screen, you should pick the shirt because 'shirt' sounds the most like 'irt'."). Older children and adults were told that the pictures were identifiable by the most basic term. For example, a picture of a shirt would be identified by the word "shirt," not "button-down" or "clothing." No other strategies were suggested.

The test session consisted of a training phase and a test phase. All participants were required to meet a training criterion of 4 consecutive correct responses within a maximum of 20 trials consisting of randomly ordered high- and low-context sentences. On average, listeners achieved the training criterion in 7.4 trials. After reaching the criterion, participants proceeded to the test phase, which consisted of 40 randomly ordered sentences (20 low-context, 20 high-context) presented in either low or high noise. Target items could occur
only once within a test session, necessitating two lists (A and B), each containing 40 sentences. The presentation of lists and noise levels was fully counterbalanced across age groups and gender.

Adults and 9-year-old children initiated trials at their preferred pace. The experimenter initiated trials for 5-year-olds when she judged them to be ready and attentive. The experimenter remained in the testing booth during the entire session for children, offering periodic verbal reinforcement and encouragement that was unrelated to their performance. To maintain the interest of 5-year-olds, they received a colored sticker after every four trials, which they placed in an "incomplete" black-and-white picture. At the end of 40 trials, children had completed the picture.

Results

Figure 4.1 illustrates performance at each noise level as a function of context and age. Because preliminary analyses revealed no difference in performance as a function of list (A or B), this factor was omitted from subsequent analyses. Despite attempts to control for differences in stimulus audibility on low-context sentences, a one-way ANOVA with age as the independent variable analyses revealed a marginal effect of age, $F(2, 69) = 2.776$, $p = .07$. Post-hoc Tukey analyses, however, did not achieve conventional levels of significance ($p > .1$ for all comparisons). A 2 x 3 ANOVA with context as a within-subject factor and age as a between-subjects factor revealed that high-context sentences were identified more accurately than low-context sentences, $F(1, 69) = 79.94$, $p < .001$, and that 5-year-olds performed more poorly than 9-year-olds and adults, $F(2, 69) = 3.34$, $p = .041$ (see Table 4.2). As illustrated in Figure 4.1, there was no interaction between age and context, $F(2, 69) = 1.76$, $p = .18$. 

57
Figure 4.1: Performance of 5-year-olds, 9-year-olds, and adults on low- and high-context sentences in both levels of noise. Error bars represent standard error of the mean.
Table 4.2: Performance (percent correct) of 5-year-olds, 9-year-olds, and adults on low- and high-context sentences in both levels of noise.

<table>
<thead>
<tr>
<th></th>
<th>Low noise</th>
<th>High noise</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low context</td>
<td>High context</td>
</tr>
<tr>
<td>5-year-olds</td>
<td>76.59 (9.51)</td>
<td>86.67 (9.17)</td>
</tr>
<tr>
<td>9-year-olds</td>
<td>76.46 (9.61)</td>
<td>91.04 (6.59)</td>
</tr>
<tr>
<td>Adults</td>
<td>82.08 (9.20)</td>
<td>91.25 (7.26)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

In high noise, however, performance on low-context sentences varied significantly with age, F(2, 69) = 3.09, p = .052. Post-hoc Tukey comparisons of 5-year-olds with older listeners approached conventional levels of significance, p = .058 and p = .087 for 5- and 9-year-olds, and 5-year-olds and adults, respectively (see Table 4.2). A 2 x 3 ANOVA with context as a within-subject factor and age as a between-subjects factor revealed main effects of context, F(1, 69) = 81.36, p < .001, and of age, F(2, 69) = 3.92, p = .024, but no interaction between context and age.

To compare performance at both noise levels, a 2 x 2 x 3 repeated-measures ANOVA, with context as a within-subject factor and noise level and age as between-subjects factors, revealed a main effect of context, F(1, 138) = 160.01, p < .0001, and of age, F(2, 138) = 6.59, p = .002, reflecting 5-year-olds’ poorer performance (M = 75.83%, SD = 10.83) than that of adults (M = 81.72%, SD = 9.14), Tukey HSD, p = .012. As expected, listeners performed better in low noise (M = 83.99%, SD = 7.04) than in high noise (M = 74.48%; SD = 9.62), F(1, 138) = 49.77, p < .001. There were no two-way or higher order interactions.

To gain further perspective on context effects, we examined reaction times to words correctly identified in low- and high-context sentences across age. The means and standard deviations of individual listeners on low- and high-context sentences were calculated.
separately, excluding trials on which reaction times exceeded that individual’s mean by 2 standard deviations. Discarded trials for low-context items represented 6.26% (SD = 3.87), 5.32% (SD = 3.77), and 5.23% (SD = 3.12) of all correct trials for 5-year-olds, 9-year-olds, and adults, respectively. For high-context items, discarded trials represented 6.70% (SD = 3.34), 6.33% (SD = 4.58), and 5.88% (SD = 5.20) of all correct trials for 5-year-olds, 9-year-olds, and adults, respectively. A repeated-measures ANOVA indicated that these discard rates did not differ across context, F(1, 138) = 2.24, p > .1, age, F(1, 138) = 1.28, p > .2, or noise, F(1, 138) = 1.06, p > .3. No interactions between the factors achieved conventional levels of significance. To achieve homogeneity of variance, two 5-year-old girls who exhibited very high latencies were excluded from the subsequent analysis. Table 4.3 reports average latencies in each age group as a function of context. A 2 x 2 x 2 x 3 repeated-measures ANOVA, with mean latency as the dependent variable, context as a within-subject factor, and age and noise as between-subjects factors revealed that target words in high-context sentences (M = 1.361s, SD = .573) were identified more rapidly than those in low-context sentences (M = 1.695s, SD = .573), F(1, 130) = 89.40, p < .0001. Moreover, latencies in low noise (M = 1.449s, SD = .516) were shorter than those in high noise (M = 1.670s, SD = .544), F(1, 130) = 4.132, p = .04. As shown in Table 4.3, 9-year-olds and adults responded more rapidly than 5-year-olds, F(2, 130) = 7.84, p = .001, but there was no age x context interaction.
Table 4.3: Average latencies in seconds for low- and high-context sentences as a function of age.

<table>
<thead>
<tr>
<th>Age</th>
<th>Low-Context</th>
<th>High-context</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year-olds</td>
<td>1.883 (.595)</td>
<td>1.623 (.625)</td>
</tr>
<tr>
<td>9-year-olds</td>
<td>1.672 (.639)</td>
<td>1.254 (.512)</td>
</tr>
<tr>
<td>Adults</td>
<td>1.534 (.420)</td>
<td>1.217 (.500)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are in parentheses.

Discussion

Children 5 and 9 years of age and adults listened to high- and low-context sentences in noise (multitalker babble) at levels that were adjusted to equalize signal audibility across age. Under these circumstances, listeners of all ages identified the target words more accurately and more quickly in high-context sentences than in low-context sentences and at lower noise levels than at higher noise levels. The youngest listeners had gains in identification accuracy and reductions in latency that were comparable to those of older listeners.

Favorable effects of context on auditory word recognition are in line with previous findings of speech identification in noise by children (Elliott, 1979; Nittrouer & Boothroyd, 1990) and adults (e.g., Bilger et al., 1984; Boothroyd & Nittrouer, 1988; Kalikow et al., 1977). Nevertheless, comparable gains in identification accuracy across age are at odds with the prevailing view that young children use contextual information less effectively than do older children, who, in turn, use such information less effectively than adults (Elliott, 1979; Nittrouer & Boothroyd, 1990). Instead, our findings are consistent with the notion of similar lexical organization in 5-year-olds and older listeners (Cirrin, 1984). Thus, there is every reason to believe that the contextual cues in the test sentences primed similar semantic networks in 5-year-olds, 9-year-olds, and adults.
It is likely that the principal source of discrepancies between the present findings and previous findings (Elliott, 1979; Nittrouer & Boothroyd, 1990) is methodological. Open-set responding (e.g., repeating the final word of test sentences), which is known to pose disproportionate difficulty for young children (Elliott et al., 1979; Geffner et al., 1996), was used in previous studies (e.g., Elliott, 1979; Nittrouer & Boothroyd, 1990), unlike the closed-set responding (picture-pointing) in the present investigation. Moreover, some of the high-predictability sentences in previous research included semantic contextual cues that would be ambiguous or incomprehensible to young children. For example, metaphorical expressions (e.g., He played a game of cat and mouse) such as those used by Elliott (1979) are not mastered until well into the school years (Winner, 1988). Few researchers have provided feedback, which is known to enhance children’s performance (Smith & Hodgson, 1970). Most critical, however, is the typical failure to control for age-related differences in stimulus audibility. When all age groups are tested at the same signal-to-noise level, it is difficult to separate the contributions of bottom-up and top-down factors.

In the present study, reduced latencies to identify words in high-context sentences imply that 5-year-olds, like adults, engage in interpretive processing as each message unfolds. Thus, words appearing early in a sentence constrain listeners’ expectations of what words will follow, leading to faster identification of the final words of high-context sentences. Nevertheless, 5-year-olds’ response latencies were higher than those of older listeners for low- as well as high-context sentences. In other words, older listeners processed speech in noise faster than did young listeners. More elaborate semantic networks in older listeners may account for their increased speed of semantic processing.

The addition of 2 dB of noise had comparable effects on performance irrespective of age, disconfirming claims that young children are especially disadvantaged by increases in noise (Mills, 1975). In Chapter 2, a 5-dB difference in SNR between 5-year-olds and adults—the difference in absolute and masked thresholds between these age groups (Schneider et al., 1986, 1989)—led to comparable word identification accuracy with very simple test sentences.
(e.g., Touch the dog; Touch the key). Pilot testing with the more complex sentences in the present investigation (e.g., Mom talked about the fish; I went to the pond and caught a fish) indicated that a 6-dB difference in SNR equalized 5-year-olds' and adults' performance in low noise. Ultimately, that SNR difference failed to equalize performance levels across age, especially at the higher noise level. What accounts for this difference between studies? All sentences in Chapter 2 were identical except for the final monosyllabic word, which means that they were equivalent in number of syllables and roughly equivalent in duration. By contrast, sentences in the present investigation varied in the number of syllables, overall duration, syntactic structure (At the soccer game, I waved my flag), and availability of contextual cues. Such variability may pose differential difficulty for 5-year-olds, just as multiple talkers interfere with preschoolers' ability to identify words in noise (Oliver-Ryalls & Pisoni, 1997). With increasing age, and corresponding growth in linguistic and cognitive ability, children are likely to become more skilled at ignoring prominent but irrelevant cues (Bialystok, 1993; Morton & Trehub, in press).

In contrast to the 12% benefit from context that was evident in the present study, Kalikow et al. (1977) reported high-context gains as great at 60% in his adult sample. Note, however, that Kalikow et al. (1977) used open-set responding. For the adults in Chapter 2, the SNR needed to maintain a particular performance level for an open-set task was 13 dB higher than that needed in a closed-set task. Even with low-context sentences, listeners could benefit from cues contained in the array of pictures. If they hear the sound /b/, for example, they could restrict their choices to depicted words. Thus, the closed response may account for the modest but highly reliable performance gains from semantic cues in the test sentences.

In short, our findings indicate that, under adverse listening conditions, young children benefit from semantic contextual cues to the same extent as do older children and adults. It remains to be determined whether 5-year-olds show comparable enhancement when the contextual cues are embedded in more complex discourse.
Chapter 5
General Discussion

The studies in this collection examined some of the factors involved in children's perception of speech under adverse listening conditions. The use of adverse listening conditions in these circumstances was expected to shed light on word recognition processes that might not be apparent in optimal listening conditions. Because the procedures, materials, and task demands of previous investigations may have placed young children at a disadvantage relative to older listeners (e.g., Elliott, 1979; Elliott et al., 1979; Larson, Petersen, & Jacquot, 1974; Nittrouer & Boothroyd, 1990), age-appropriate methods and signal-to-noise levels were used in the present investigation to compare children's and adults' identification of words in a background of multitalker babble.

The picture-pointing task developed for this research minimized cognitive demands on young listeners by using familiar, monosyllabic nouns, closed-set responses, and corrective feedback. On the basis of children's lesser auditory sensitivity (Schneider et al., 1986, 1989), it was not surprising that they required more favorable SNRs to achieve comparable performance levels to those of adults. Of particular interest, however, was the finding that equivalent additions of noise had similar consequences for listeners of all ages. This was the case when the words and foils were maximally contrastive and when the words were embedded in low-context sentences (Chapter 2). When the target words and foils were minimally contrastive, young children's performance approached that of adults for critical contrasts in word-initial position, but not for contrasts in word-final position (Chapter 3). Finally, the provision of semantic cues led to comparable gains in performance for 5-year-olds, 9-year-olds, and adults (Chapter 4). Nevertheless, variations in sentence structure necessitated more favorable SNRs for children than were required for the simpler sentences in Chapter 2.
In a general sense, what do these findings indicate about children’s recognition of words in noisy backgrounds? Returning to Cole and Jakimik’s (1980) first assumption that speech recognition is achieved by integrating auditory and linguistic information, the present findings indicate that young children can integrate auditory cues and linguistic knowledge despite the distracting effects of noise. There are numerous claims about the adverse consequences of noise for young listeners (e.g., Mills, 1975) and its interference with the coordination of linguistic and acoustic-phonetic information (Elliott, 1979). Although 5-year-olds required more favorable listening conditions than did older children and adults, noise did not prevent young children from attending to the target signal or using their knowledge of language to determine word identity.

Recall Cole and Jakimik’s (1980) second assumption that word-initial information activates word candidates, with successive information eliminating inappropriate word candidates until a single word is specified (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987). Young children require more word-initial information to identify words than do older children (Elliott et al., 1987; Walley, 1987), which is consistent with the notion of less analytic processing by young listeners (Walley, 1993). When target words are presented with maximally contrastive alternatives, as in Chapter 2, correct identification can be achieved by a global matching process. Focusing on particular portions of words, however, is critical for distinguishing between similar-sounding alternatives (Bradlow & Pisoni, 1999). The findings of Chapter 3 indicate that 5-year-olds can use word-initial information to identify targets presented with minimally contrastive alternatives, and noise does not interfere with their ability to do so. These findings are consistent with 5-year-olds’ ability to identify rhyming words (Knafle, 1973, 1974) and their difficulty doing so when nonrhyming foils have the same initial sound as the target (Lenel & Cantor, 1981). Nevertheless, 5-year-olds’ difficulty with word-final contrasts indicates that their processing of fine-grained acoustic information is less differentiated than it is in older listeners. Indeed, 5-year-olds’ performance in noise parallels their reported difficulties with word-final differences in
optimal listening conditions (Hnath-Chisofm et al., 1998; Menary et al., 1982; Morgan, 1984).

Cole and Jakimik’s (1980) third assumption is that words are recognized sequentially, with preceding semantic and syntactic information affecting subsequent word identification. Contrary to previous reports (e.g., Elliott, 1979; Nittrouer & Boothroyd, 1990), noise did not impair 5-year-olds’ ability to use such contextual cues (Chapter 4). In fact, 5-year-olds and adults derived comparable benefit from semantic cues.

In sum, the results of the present investigation indicate that noise does not interfere with 5-year-olds’ ability to integrate acoustic-phonetic information with linguistic knowledge, to use word-initial information, or to benefit from semantic contextual cues. It should be noted, however, that these tasks were designed to minimize the cognitive and linguistic demands on listeners. With the exception of the noise background itself, the conditions approximated the best-case scenario for evaluating young children’s identification of words. No doubt, young children would experience considerably greater difficulty in situations involving less familiar words, more complex sentence structures, and open-set responses (e.g., Elliott, 1979; Nittrouer & Boothroyd, 1990). Arguably, these conditions are closer to the challenges of everyday life than are the favorable conditions of the present investigation. Nevertheless, understanding how children perform in the most supportive environment is essential for pinpointing the sources of difficulty in more challenging situations.

The present research suggests a number of promising directions for future research. Among the most important is a more complete picture of the course of development of speech recognition in noise. In particular, children younger than those in the present investigation may experience disproportionate difficulty perceiving speech in noise because of underspecified phonological representations as well as lesser linguistic and world knowledge. Presumably, one would expect more global, or holistic, processing strategies in younger children, including particular difficulties with minimally contrastive words. A
number of theoretical formulations (e.g., Best, 1994; Studdert-Kennedy, 1986) include some links between articulatory and acoustic events. It follows, then, that articulation skill should contribute to word identification in noise, a proposal that has not been evaluated to date.

It is also important to ascertain the consequences of less favorable listening situations than those used in the present investigation. The foregoing experiments involved a narrow range of noise levels. According to Pollack and Pickett (1958), high levels of noise overload the auditory systems of adults, resulting in disproportionate losses of intelligibility. The nature of the test voice (or voices) is also likely to affect children's performance. The well-modulated voice used in the present research provided more favorable acoustic cues for adults and children than did the male speaker in the SPIN task (see Chapter 2). Whether well-modulated voices are necessary to maintain children's attention in such tasks remains to be determined. Multiple speakers impair speech identification for adult listeners (e.g., Mullenix & Pisoni, 1990; Mullenix, Pisoni, & Martin, 1988; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994), but there are suggestions that the impairment is even greater for young children (Oliver Ryalls & Pisoni, 1997). The combination of noise and multiple talkers may be especially problematic for young children.

Although the present study revealed that children and adults derived comparable benefit from semantic cues (see Chapter 4), they did so in the context of relatively simple test stimuli and response formats. The closed set of alternatives was essential for separating perceptual from cognitive contributions. It is also necessary, however, to examine age-related differences in performance with open-set responding. With more complex sentences or stories (in noise), problems of attentional control (e.g., Bialystok, 1999) might become evident.

The specific task and materials developed for the present investigation could be applied productively to particular populations of interest. One such population is bilingual children. Bilingual adults experience greater difficulty identifying speech in noise than do monolinguals (Gat & Keith, 1978; Mayo et al., 1997; Oyama, 1978), the difficulty being
greater for later ages of second language acquisition (Mayo et al., 1997). Acquisition before age 6 or 7 is associated with ultimate native-like proficiency in grammatical (Johnson & Newport, 1989) and phonological (Flege, 1987; Flege & Fletcher, 1992; Oyama, 1976) aspects of the second language. There are suggestions that early bilinguals represent their two languages in common brain regions but that late bilinguals exhibit some spatial separation in these representations (Kim, Relkin, Lee, & Hirsch, 1997). Although the ultimate level of attainment is native-like in young second-language learners, such learners may face perceptual challenges relative to monolingual age-mates until they achieve some threshold of language proficiency. Thus, documenting age-related changes in young bilinguals' identification of speech in noise can provide invaluable information about maturational and experiential constraints on such skills.

Another population of interest is congenitally deaf children who receive auditory input from cochlear implants. Although implants provide very degraded spectral cues coupled with good temporal cues, many implanted children still manage to acquire spoken language solely on the basis of such input (e.g., Svirsky, Robbins, Kirk, Pisoni, & Miyamoto, 2000). There are numerous reports of speech perception in quiet on the part of young implant users (e.g., Chute, Hellman, Parisier, & Selesnick, 1990; Cowan et al., 1993; Dawson et al., 1992; Fryauf-Bertschy, Tyler, Kelsay, Gantz, & Woodworth, 1997; Meyer, Bertram, & Lenarz, 1995; Miyamoto, Kirk, Todd, Robbins, & Osberger, 1995; Waltzman et al., 1997), but little attention has been paid to the effects of noise. The few studies that have examined this issue provide gross comparisons of performance in quiet and in noise (e.g., Cowan et al., 1995), with little consideration of the factors that contribute to implanted listeners' difficulty. As with bilinguals, maturational and experiential constraints can be investigated in implanted children by examining performance as a function of duration of deafness prior to implantation and length of experience with the device.

Beyond the theoretical implications of the present investigation, there are practical implications for everyday speech recognition. School classrooms are known to have high
levels of ambient noise that are associated with increases in stress and decreases in attention, motivation, and academic performance (Hétu et al., 1990), effects that are more pronounced at younger ages (Hétu et al., 1990; Houtgast, 1981; Whaley & Hanson, 1984). Despite recommended classroom noise levels of 30-35 dB (A) (Berg, 1993; Crandell & Smaldino, 1994), actual noise levels range from 55-80 dB (A) (Bess, Sinclair, & Riggs, 1984; Hétu et al., 1990). Noise levels used in the present research were similar to those reported for classrooms, but classroom difficulties are compounded by reverberation, greater distance between the listener and the speaker (Finitzo-Herber, 1988), and fluctuating noise levels. In reverberant or noisy conditions, children’s consonant identification becomes adult-like by age 14; when noise and reverberation are both present, however, adult levels of performance are not observed until the late teenage years (Johnson, 2000).

The findings of the present investigation cannot be accounted for by current theories of speech perception and word recognition, which fail to capture the flexibility of the system. Although some models acknowledge multiple phonological, semantic, and syntactic influences on speech recognition (Liberman & Mattingly, 1985; Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986), they rely on pattern matching mechanisms that are largely immutable and cannot adequately account for developmental change (see Nusbaum & Goodman, 1994). Accurate speech perception requires listeners to recognize and direct their attention to the most informative aspects of the utterance (Goodman, Lee, & DeGroot, 1994). Noisy situations, in particular, demand such flexibility. The tasks developed for the present investigation are suitable for a wide age range. Thus, they have the potential to yield important information about auditory word recognition from the preschool period through old age. The resulting information about developmental changes in speech identification in noise should promote the refinement of theories that stress the flexibility and adaptability of speech processing.
<table>
<thead>
<tr>
<th>Practice</th>
<th>Low Context</th>
<th>High Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>He read about</td>
<td>Farm animals stay in</td>
<td>I bought Dad a leather belt.</td>
</tr>
<tr>
<td>the barn.</td>
<td>a barn.</td>
<td></td>
</tr>
<tr>
<td>Mom talked</td>
<td>I put the bird back</td>
<td>Mice like to eat cheese.</td>
</tr>
<tr>
<td>about the</td>
<td>in its cage.</td>
<td></td>
</tr>
<tr>
<td>belt.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>We looked at</td>
<td>Rain poured from the</td>
<td></td>
</tr>
<tr>
<td>the cage.</td>
<td>cloud.</td>
<td></td>
</tr>
<tr>
<td>He looked at</td>
<td>At the farm, I saw a</td>
<td>The king wore a gold crown.</td>
</tr>
<tr>
<td>the cheese.</td>
<td>cow.</td>
<td></td>
</tr>
<tr>
<td>We pointed at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the cloud.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>She read about</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the cow.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mom pointed at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the crown.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>He talked about</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the deer.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dad read about</td>
<td>Mike banged on a</td>
<td></td>
</tr>
<tr>
<td>the drum.</td>
<td>drum.</td>
<td></td>
</tr>
<tr>
<td>She pointed at</td>
<td>I was hot, so I turned</td>
<td></td>
</tr>
<tr>
<td>the fan.</td>
<td>on the fan.</td>
<td></td>
</tr>
<tr>
<td>She talked about</td>
<td>At the soccer game,</td>
<td></td>
</tr>
<tr>
<td>the flag.</td>
<td>I waved my flag.</td>
<td></td>
</tr>
<tr>
<td>We read about</td>
<td>To water the lawn,</td>
<td></td>
</tr>
<tr>
<td>the hose.</td>
<td>Dad used the hose.</td>
<td></td>
</tr>
<tr>
<td>Dad looked at</td>
<td>The bully punched my</td>
<td></td>
</tr>
<tr>
<td>the nose.</td>
<td>nose.</td>
<td></td>
</tr>
<tr>
<td>Dad looked at</td>
<td>I answered the phone.</td>
<td></td>
</tr>
<tr>
<td>the phone.</td>
<td>To hold cloth together, we use a pin.</td>
<td></td>
</tr>
<tr>
<td>Dad pointed at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the pin.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>He pointed at</td>
<td>I gave my mom a pretty rose.</td>
<td></td>
</tr>
<tr>
<td>the rose.</td>
<td>At the beach, I found a shell.</td>
<td></td>
</tr>
<tr>
<td>We talked about</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the shell.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>She looked at</td>
<td>I got bitten by a snake.</td>
<td></td>
</tr>
<tr>
<td>the snake.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mom read about</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the soap.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mom looked at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the sock.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>We pointed at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the bag.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mom pointed at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the ball.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>We read about</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the bed.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dad pointed at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the bee.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dad looked at</td>
<td></td>
<td></td>
</tr>
<tr>
<td>the book.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Mom looked at the **boots**.

We looked at the **bread**.

Mom read about the **broom**.

Mom pointed at the **brush**.

Mom looked at the **bus**.

Mom read about the **cake**.

She pointed at the **car**.

We pointed at the **cat**.

She talked about the **chair**.

She read about the **clock**.

He read about the **clown**.

Dad pointed at the **corn**.

He talked about the **cup**.

We read about the **doll**.

He talked about the **door**.

We talked about the **dress**.

Dad talked about the **duck**.

Mom talked about the **fish**.

Dad read about the **fork**.

She pointed at the **horse**.

She looked at the **house**.

She talked about the **key**.

He looked at the **kite**.

Dad read about the **net**.

Dad looked at the **pail**.

He read about the **pants**.

She pointed at the **pig**.

---

When it snows, I put on my **boots**.

Sandwiches are made with **bread**.

I cleaned the floor with a **broom**.

To untangle my hair, I use a **brush**.

Dad rides to work on the **bus**.

For dessert, we ate **cake**.

We drove to the store in our **car**.

The dog chased the **cat**.

I sat down on the **chair**.

I knew the time when I looked at the **clock**.

We laughed at the funny **clown**.

Farmers plant rows of **corn**.

I drink juice out of a **cup**.

The girl played with her **doll**.

Mom asked me to open the **door**.

She wore a pretty **dress**.

At the pond, I fed a **duck**.

I went to the pond and caught a **fish**.

I eat spaghetti with a **fork**.

I learned how to ride a **horse**.

Ann’s family lives in a **house**.

To open the door, Dad used a **key**.

I like to fly my **kite**.

Nick catches bugs with a **net**.

We carried the water in a **pail**.

I fell and ripped my **pants**.

The farmer fed the **pig**.
He looked at the pot.

He pointed at the shoe.

Dad talked about the skunk.

We looked at the snow.

She read about the star.

We talked about the tie.

Mom talked about the tree.

She looked at the wheel.

Mom cooks dinner in a pot.

I know how to tie a shoe.

An animal that smells bad is a skunk.

I like to play in the snow.

In the sky, I saw a bright star.

When Dad gets dressed up, he wears a tie.

A bird built its nest in our tree.

My wagon has a broken wheel.
References


78


85


