



ELSEVIER

Cognition 71 (1999) 73–108

COGNITION

Continuous processing in word recognition at 24 months

Daniel Swingley*, John P. Pinto, Anne Fernald

Department of Psychology, Jordan Hall, Stanford, University Stanford, CA 94305, USA

Received 12 May 1998; accepted 12 March 1999

Abstract

Speech processing in adults is continuous: as acoustic-phonetic information is heard, listeners' interpretation of the speech is updated incrementally. The present studies used a visual fixation technique to examine whether young children also interpret speech continuously. In Experiments 1 and 2, 24-month-old children looked at visual displays while hearing sentences. Sentences each contained a target word labeling one of the two displayed pictures. Children's latency to fixate the labeled picture was measured. Children's responses were delayed when the competing distractor picture's label overlapped phonetically with the target at onset (dog-doll), but not when the pictures' labels rhymed (ball-doll), showing that children monitored the speech stream incrementally for acoustic-phonetic information specifying the correct picture. In Experiment 3, adults' responses in the same task were found to be very similar to those of the 24-month-olds. This research shows that by 24 months, children can interpret speech continuously. © 1999 Elsevier Science B.V. All rights reserved

Keywords: Speech processing; Visual displays; Word processing; Continuous

1. Introduction

Little is known about the processes by which young children recognize words in speech. It is clear that infants in their first year store words in memory (Jusczyk and Aslin, 1995; Jusczyk and Hohne, 1997); furthermore, infants around 12 months old can associate spoken words and their referents (e.g. Huttenlocher, 1974; Benedict, 1979; Thomas et al., 1981). These findings suggest that a lexicon begins to develop toward the end of the first year of life. However, the basic characteristics of early

* Corresponding author. Max-Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands. Tel.: +31-24-352-1373; fax: +31-24-352-1213.
E-mail address: daniel.swingley@mpi.nl (D. Swingley)

lexical representations are not well understood. For example, it is not clear how well specified the sound forms of young children's first words are. Although infants seem capable of discriminating phonemic minimal pairs (cf. Aslin et al., 1998), some studies assessing young children's ability to distinguish minimally different words have suggested that young children's representations of words are inaccurate or underspecified (e.g. Shvachkin, 1973; Eilers and Oller, 1976; Stager and Werker, 1997).

The process by which young children map heard speech onto known words over the course of an utterance is not well understood either. Although the on-line use of acoustic-phonetic information is a central topic in research on word recognition in adults, developmental researchers have long been unable to measure word recognition in young children over the course of an utterance. Consequently, the origins of adults' speech processing abilities are obscure.

The studies reported here employed a visual fixation technique to assess young children's on-line use of speech information, and to shed light on the specificity of early lexical representations. These problems are closely related, because of the interdependence of representation and process in word recognition (e.g. Marslen-Wilson, 1993). If children's lexical representations are qualitatively different from adults', as a number of researchers have suggested, it is likely that the same sorts of processing strategies would not be equally effective in mapping speech to words. For example, Walley (1993) has suggested that children and adults might differ in their ability to use speech information *incrementally* over time in interpreting speech.

Incremental processing in speech understanding refers to the continuous use of acoustic-phonetic information. In word recognition, incremental processing is shown in adults' ability to make decisions about the identities of words as the words are heard. This is reflected in listeners' responses to ambiguous word onsets: as a sound pattern like 'gamb...' is heard, for example, listeners begin to consider words such as 'gamble' and 'gambit' as words that might be intended by the speaker. Studies using various experimental methods have indicated that in such cases, adult listeners do not need to hear words in their entirety before recognizing them (e.g. Marslen-Wilson and Welsh, 1978; Grosjean, 1980; Marslen-Wilson, 1984). Furthermore, evidence from studies using a cross-modal semantic priming technique suggests that multiple words are activated in parallel, as long as the speech heard is still consistent with these words (Zwitserslood, 1989; Zwitserslood and Schriefers, 1995; Moss et al., 1997).

In young children, however, it is not clear that word recognition follows the flow of speech information so closely. There are three reasons to suppose that young children might use a different procedure. First, continuous interpretation of speech may be possible only with the benefit of massive practice, both with individual words and with word recognition. Because children have had less practice, they may need to use slower, more effortful strategies. Such strategies might not have the incremental characteristics of adult word recognition.

Second, in current models of adult word recognition, segmentation of the speech signal is accomplished in part by matching portions of speech with words in the

lexicon (e.g. McClelland and Elman, 1986). Because young children know fewer words, they may be more likely than adults to encounter parts of sentences that do not correspond to known words, disrupting the segmentation process. If, consequently, children do not use the lexicon for on-line segmentation, they may not need to be able to rapidly activate lexical candidates from partial information, as adults do.

Finally, if young children's representations of the sound forms of words are qualitatively different from adults' representations, rapid and incremental processing of speech might not be effective for early word recognition (e.g. Treiman and Breaux, 1982; Walley, 1988; Charles-Luce and Luce, 1990). For example, if the linear sequence of the sounds that make up a word is not accurately represented in children's early lexicons, identifying words based on partial input would result in a great deal of spurious activation. Also, if the identification of phonetic segments is essential for adult-like word recognition, as has been proposed by several researchers (e.g. Pisoni and Luce, 1987), then the putatively non-segmental or 'holistic' representations of young children would be inadequate. Under these circumstances, the most efficient procedure might be to store heard speech in a buffer of a few syllables, and then attempt to identify known words in this portion of the input. Charles-Luce and Luce (1990) describe such a strategy, suggesting that young children may not 'perceive words in an exclusive left-to-right manner' (p. 206).

The alternative hypothesis is that young children's processing of speech in word recognition follows the continuous flow of acoustic-phonetic information, just as in adults. If true, this would suggest that developmental models of word recognition need not posit a qualitatively distinct procedure applied only early in language acquisition. Rather, as far as the temporal aspects of spoken word recognition are concerned, young children may perform much like adults with very small lexicons.

1.1. Adults' on-line use of information in speech

Incremental processing in adult word recognition has been demonstrated in experiments using lexical decision tasks (Marslen-Wilson, 1984), cross-modal priming (Zwitserslood, 1989; Moss et al., 1997), and a visual fixation task (Alloppenna et al., 1998). Marslen-Wilson (1984) had subjects make speeded responses to auditory targets, indicating whether each target was a word of English (the lexical decision task). Response latencies to non-words were essentially constant if measured from the location of the 'discrimination point' of each item, defined as the point at which that target diverged from all words in the language. For example, responses to words like 'trenker' (with a discrimination point at the /k/) came about 450 ms after the /k/, and responses to 'zawritude' (with a discrimination point at the /ɔ/) came about 450 ms after the /ɔ/. These results indicated that listeners were able to monitor each target continuously for consistency with one or more known words.

Incremental processing in word recognition was also shown in a study by Zwitserslood (1989), employing cross-modal semantic priming. Dutch subjects heard sentences culminating in a word fragment which was consistent with more than one word (e.g. kapit, which is the onset of both 'kapitein' (captain) and 'kapitaal'

(capital)). Lexical decision responses to targets presented at the offset of the word fragment were facilitated when targets were semantically related to 'kapitein' or 'kapitaal', relative to unrelated controls. This result indicated that semantic activation may precede the unique identification of words. These effects have been replicated in several experiments (Zwitserslood and Schriefers, 1995; Moss et al., 1997).

Finally, a recent study by Allopenna et al. (1998; see also Cooper, 1974) demonstrated continuous interpretation of speech in another on-line task. On each trial, adult subjects viewed a computer display showing four pictures of familiar objects arranged in a square. Subjects began each trial by fixating the center of the square. Then they followed spoken instructions involving manipulating the pictures with a computer mouse (e.g. 'take the beaker and place it above the square'). Subjects' visual fixations were monitored using an eyetracking system. On some test trials, two of the pictures could be named using words with identical onsets (e.g. beaker, beetle). When the spoken word matched one of these two pictures, subjects tended to fixate both members of this onset-overlapping pair more often than the unrelated distractors, as the early part of the target was spoken. Shortly after the target word became inconsistent with the unnamed onset-overlap picture, fixations to this picture sank to the baseline frequency. These results showed that listeners are able to monitor the speech signal incrementally to guide behavioral responses that depend on word meaning.

While the above results provide solid evidence for incremental processing, they do not imply that the speed or accuracy of word recognition is determined solely by the availability of acoustic-phonetic information in the signal. For example, numerous studies have shown that frequent words are recognized more quickly than infrequent words (e.g. Taft and Hambly, 1986). Furthermore, there is considerable debate about what counts as a match between a spoken word and a known lexical item. Under some circumstances, slightly mispronounced words can trigger lexical activation, even when the mispronunciations include phoneme substitutions (e.g. Connine et al., 1993; Marslen-Wilson et al., 1996). Thus, word recognition may be broadly described as an incremental process in which words are activated while they still match the signal; but a complete description of performance must include factors such as word frequency, and must specify what counts as a match (see e.g. Marslen-Wilson, 1993 for discussion).

1.2. Word recognition in young children

The goal of many developmental studies of early word recognition has been to determine when children are capable of distinguishing minimally different words. Several such studies used a method pioneered by Shvachkin (1973). In Shvachkin's study, children between 10 and 18 months old were taught monosyllabic names for three novel objects over a period of a few days. Two of the three names differed in only one segment. To evaluate learning of the object names, children were shown two objects with dissimilar names, and given the label for one of them. Once children could pick out the named objects under these conditions, they were

given a more difficult test in which the target was paired with its similarly-named partner. Success on this test was considered evidence for discrimination of the contrast distinguishing the similar pairs. Shvachkin found that children did not consistently discriminate several of the contrasts until well into their second year.

Garnica (1973) attempted a replication of Shvachkin's study with American children, using a substantially similar procedure. Subjects were 16 children, beginning the 4-month study at between 17 and 22 months of age. While the order of acquisition of the contrast discriminations varied somewhat from that reported by Shvachkin, the basic finding that many contrasts were not evident was the same. For example, in the first half of the study, only 20% of the subjects demonstrated consistent discrimination of the voiced/voiceless distinction.

Other studies, with slightly older children, have found better performance. Barton (1980) familiarized 20- to 24-month-olds with toys (correctly) labeled as a bear, a pear, a goat, and a coat. Children were tested using instructions to remove the objects from a bag one by one (i.e. 'Get out the coat. Now get out the pear.'). Barton reported that 'eight of the 10 children could do at least one of the discriminations' (p. 101). We might assume that adults would perform at ceiling without difficulty. In another study, Eilers and Oller (1976) contrasted real words with minimally different novel words (taught in the experimental session), and found that none of their 14 22- to 26-month-olds could distinguish 'fish' and 'thish' or 'mucky' and 'monkey' in a task involving retrieving a reward from underneath one of two objects.

These object choice experiments have been supplemented by studies asking children to make more abstract judgments of speech sounds. Graham and House (1971) tested three- and four-year-olds in a same/different judgment task using 240 (16 Cs times 15 Cs) spoken consonant contrasts, using consonants embedded in the frame /hə'Cʌdə/ (where 'C' represents the test consonant). The mean error rate on 'match' trials (i.e. children saying 'different' when the two words were the same) was 3%; the error rate on 'mismatch' trials (i.e. children saying 'same' when the words were different) was 14%. The most difficult contrasts involved fricatives (/ʃ/-/tʃ/, /ʒ/-/s/, /s/-/z/), sonorants (/l/-/r/, /m/-/n/), and, surprisingly, /p/-/t/; the error rate for each of these oppositions was above 20%.

Using a word-monitoring task, Gerken et al. (1995) found that four-year-olds made many false alarms to words minimally different from targets. Children were taught to press a button if a spoken word (presented in list form) matched a reference word target ('little' or 'lick'). Over four experiments, miss rates ranged from 2% to 10%; false alarm rates to minimally different probes ranged from 21% to 56%, and false alarms to control probes like 'cookie' ranged from 3 to 6%. Probes varied from targets by either one or two features (e.g. place, manner). The high false alarm rate showed children's willingness to consider near-neighbor non-words as equivalents of real words.

Finally, in a recent study, Stager and Werker (1997) found that 14-month-olds do not differentiate minimal pairs (such as 'bih' and 'dih') when taught as novel words referring to objects, even though 14-month-olds can discriminate these pairs when they are not used referentially. In one experiment, infants were habituated to a

checkerboard pattern and repeated presentations of a novel word ('bih'). When infants' fixation of the checkerboard declined to a threshold, indicating habituation, the sound pattern was changed to 'dih.' Infants' increases in looking time indicated discrimination of the syllables. In a second experiment, infants were habituated to the same sound sequence, but this time the visual display showed a moving novel object rather than a checkerboard. In this case infants failed to recover from habituation when the sound pattern was changed to 'dih.' Stager and Werker argued that this failure was due to the fact that infants were attempting to link the sound pattern and the object, i.e. they were learning a new word. These results suggested that the critical difference between infant discrimination tasks (in which infants successfully discriminate minimal pairs) and previous assessments of older children's discrimination abilities (in which children sometimes fail to discriminate minimal pairs) may be *reference*: when linking words and objects, children may not attend to fine phonetic detail. This lack of attention might result in underspecified representations across the lexicon, and not only for novel words.

Many of these studies have been interpreted as evidence that children's representations of words are underspecified or indeterminate in some way that prevents adult-like performance in recognition. Given these findings, it is not unreasonable to suppose that adults and children might differ in their use of acoustic-phonetic information for word recognition (e.g. Walley, 1993). For example, children might recognize words only after hearing the entire word, or might activate words only after an accumulated 'buffer' of a few syllables.

The alternative hypothesis is that children's representations of the sound forms of familiar words, and the means by which children recognize words, are not qualitatively different from adult representations and mechanisms. Children's performance, which is generally inferior to what would be expected from adults, does not necessarily demonstrate representational differences between children and adults; perhaps children can distinguish similar words, for example, but fail to do so as consistently as adults because of lapses of attention, or for other non-linguistic reasons. On this account, the fact that children are more likely to fail on difficult contrasts (such as /m/ vs. /n/) than easy ones (e.g. /m/ vs. /k/) is not surprising, given that adults, too, are more likely to confuse similar phonemes than dissimilar phonemes (e.g. Miller and Nicely, 1955).

The experiments reported below tested young children's ability to use speech information over time to recognize familiar words. A visual fixation procedure was used. In two studies, 24-month-olds were shown pictures of objects whose names overlapped at onset (e.g. 'doggie,' 'doll'), rhymed (e.g. 'duck,' 'truck'), or did not overlap (e.g. 'doggie,' 'tree'). On each trial, one picture was named in a sentence ('Look at the doggie.'). Analyses of children's fixation patterns in each condition permitted assessment of the timing of children's differentiation of similar and dissimilar words. In a third study, adults were tested using the same pictures and sentences. The results show that children, like adults, have representations of lexical form that permit word recognition from partial information, and that even very young children are able to match speech to these representations incrementally.

2. Experiment 1

Most previous studies of word recognition in young children have required pointing or other manual selection responses. Such responses are too slow, and probably too variable, to be useful in addressing the questions posed here. The visual fixation procedure avoids these problems. Previous research has shown that uninstructed adults rapidly fixate the referents of spoken words (Cooper, 1974), and more recent work has demonstrated the viability of fixation measurement to study on-line language comprehension (e.g. Tanenhaus et al., 1996). Visual fixations have been used fruitfully to study language processing in young children as well, using a method developed by Golinkoff et al. (1987).

The method used in the current studies was developed by Fernald et al. (1998) to assess the time course of word recognition in infants (see Swingley et al., 1998, for a detailed description). In this procedure, infants are seated on their parent's lap facing two computer monitors that each display a picture of a common object. On each trial, the pictures are shown in silence for 4 s before a prerecorded utterance is played from a central loudspeaker, labeling one of the displayed objects (the 'target'). The pictures continue to be displayed for about 5 s, during which infants typically fixate the labeled picture. Blank (black) screens separate trials, with an inter-trial interval of 1 s. Infants' fixations are measured from videotapes recorded during the procedure. These data are used to determine infants' proportion of fixation to the labeled visual stimuli in response to the utterances, and infants' response latencies in orienting to the target pictures.

Fernald et al. (1998) found that 24-month-old children shifted their fixation from the distractor pictures to the targets about 680 ms after the onset of the spoken target words. The mean duration of the target words was 775 ms; thus, infants responded before the offsets of the target words. However, because the length of the target words was partly due to elongation of the final segment of each word (e.g. the /i/ of 'doggie'), it is still possible that children responded only after exposure to the entire phonemic content of the word. For this reason, the results of Fernald et al. do not conclusively demonstrate that 24-month-olds are capable of recognizing a word based on partial specification of the word, or that children track the information in speech continuously in an adult-like manner.

Experiment 1 was designed to test these hypotheses. 24-month-old children were presented with four pairs of pictures. In two pairs, the names overlapped at onset (doggie-doll; tree-truck), while in the other two pairs, there was no overlap (doggie-tree; doll-truck). These word pairs were selected based on several criteria: the words had to be known by most children at 24 months; they had to be picturable; and they had to overlap at onset by more than one phoneme. The two overlapping pairs used were the only pairs meeting these criteria. On each trial, one member of each pair (the target) was named in a sentence. If children are capable of continuous interpretation of speech, responses should be faster in the no-overlap case than the overlap case, because in the no-overlap case, the acoustic-phonetic information specifying only one of the two pictures is available at the onset of the word, whereas in the overlap case, the information specifying the target picture is not available until

late in the word. Furthermore, the duration of the delay in the overlap case should approximate the duration of the overlap in the target words themselves. On the other hand, if word recognition depends on hearing a word in its entirety, or storing an arbitrary number of syllables, there is no reason to expect a difference between the overlap and no-overlap conditions.

2.1. *Methods*

In Experiment 1, 24-month-old children performed in the preferential looking task. The target words to be recognized were ‘doggie’, ‘doll’, ‘tree’, and ‘truck’. The 16 test trials were preceded by a block of four ‘ostensive’ trials on which only one picture was shown and labeled. These ostensive trials were included to ensure that the children would consider each picture as intended (e.g. to prevent children from thinking the truck was a car). On each test trial, the target word was presented sentence-finally, in two sentences: ‘Where’s the [target]? See the [target]?’. On ‘onset-overlap’ trials, the names of the displayed pictures had the same initial phonemes: doggie and doll, tree and truck. On ‘no-overlap’ trials, the names did not share initial phonemes: doggie and tree, doll and truck. On each trial, only one of the pictures was labeled. Children’s visual fixations were recorded and coded off-line by coders who noted the timing of stimulus onsets and changes in fixation.

2.1.1. *Subjects*

The 32 subjects were 24 months of age, with a range of 104–109 weeks (mean 106). Half were girls. All were from monolingual English environments. Low-birth-weight (under 5.5 lb) and premature (over 4 weeks) children were excluded. All subjects had all of the target words in their productive vocabularies, as assessed by parental report. An additional 49 subjects were tested but were not retained in the final sample. Of these, 30 did not complete the procedure: they failed to fixate either picture within 3 s of the offset of the target word on three or more trials (or two, if those involved the same test pair).¹ Other subjects were excluded because the child did not know the four target words (by parental report, on the MacArthur CDI: 9); parents did not contribute a CDI (4); experimenter error occurred (3); the child did not meet the monolingual environment prerequisite (2); or the parent interfered during the procedure (1).

2.1.2. *Visual stimuli*

The visual stimuli were similarly-sized digitized photographs or realistic paintings of objects on a gray background, presented on 15-inch Macintosh color monitors. The objects were a yellow ‘Tonka’ dump truck, a plastic doll, a leafy tree, and a stuffed-animal dog. Pictured objects appeared on each side four times, twice as target and twice as distractor, in the test trials.

¹This attrition rate was about 25% higher than we have found in previous studies using similarly stringent criteria; it may be due, in part, to the relatively large number of trials.

2.1.3. Auditory stimuli

The speech stimuli were recorded by a female native speaker in a soundproofed room on a Revox B77 reel-to-reel tape recorder. Her speaking rate was slow and in a moderately 'infant-directed' register. Several tokens of each type were recorded. Tokens were selected for use in the experiment on the basis of their naturalness and the degree to which the onset-overlap pairs had similar beginning sounds. To create the critical 'Where's the...' stimulus sentences, target words were spliced onto the same recorded token of 'Where's the...'. This carrier token, from a recording of 'Where's the doggie?' was not itself the source of the 'doggie' target; thus, all 'Where's the...' sentences were spliced. This splicing procedure ensured that children could not have had any information specifying the target word until the word itself began. 'See the...' sentences were not spliced, because they always came after the first presentation of the target word. The ostensive utterances labeled the displayed picture three times: 'That's a [target]. [Target.] Do you see the [target]?'

The 'Where's the' carrier phrase was 714 ms long. Target word lengths in the 'where's the' sentences were as follows: doggie, 973 ms; doll, 961 ms; tree, 949 ms; truck, 935 ms. Note that these durations include the duration of the silent stop closure at the onset of each target word; the closure duration ranged from 131 ms to 139 ms. These words are long for adult conversation, but are not unusual for speech to young children, especially in labeling contexts, in which parents typically 'highlight' words prosodically (Fernald and Mazzie, 1991). The tape was digitized at a sampling rate of 22 kHz. Measurements of the formant frequencies of the initial vowels of the target words are presented in Appendix A.

2.1.4. Preliminary evaluation of auditory stimuli

To assess how much the onset-overlap words actually overlapped, a gating method was used with adult subjects. In gating (e.g. Grosjean, 1980), subjects hear successively longer and longer fragments of a word, and are asked to identify the word at each step.

For the present gating task, two sets of partial versions of each sentence were made. Each item in both sets contained the carrier 'Where's the...'. In one set, for each sentence the first item was interrupted at the burst of the stop consonant beginning the target word; the second item was interrupted 40 ms later; the third after 80 ms, and so on in 40 ms increments, for a total 16 items for each word. On a given trial, the subject would therefore have 16 (increasingly easy) chances to identify the target. The second set of stimuli was created the same way, but starting 20 ms after the burst, and continuing in 40 ms increments (20, 60, 100, etc.). After each 'gate,' subjects tried to identify the word, and indicated their confidence by choosing a number from one (uncertain) to 10 (certain).

Subjects were told in advance what the four possible words were. Subjects judged each of the four sentences either two or three times, randomly ordered; thus, each subject had 10 trials of 16 sentence fragments each.

Subjects were also instructed to identify the moment when they could tell whether the word was a 'tr' word or a 'd' word. If, upon having done so, they felt that they had no basis for further judgment (e.g. differentiating 'tree' from 'truck') they were

Table 1

Results from adult gating assessment of materials from Experiment 1. Zero ms here refers to the onset of the stop closure, not the burst.

Target word	1st correct response (ms)	Isolation point (ms)	Confidence >6 (ms)
Doll	346	372	459
Doggie	390	407	509
Tree	239	239	289
Truck	233	234	300
mean do–	369	389	483
mean tr–	236	236	294

to refrain from guessing which it was until they felt they had some basis for judgment, however minute. This modification of the traditional gating procedure permitted assessment of when the ‘tr’ words and ‘d’ words were discriminable, and was also intended to limit the occurrence of pure guesses.

Ten adults participated for course credit or a small payment. Six of the adults had participated in Experiment 3 and as a result were familiar with the utterances; the other four were played each sentence once before beginning the gating task. Subjects were randomly assigned to the 0, 40, 80... set or the 20, 60, 100... set.

Inspection of the data revealed that subjects always correctly identified which onset a word had as soon as minimal information was available – after hearing 40 ms (in the first set) or 20 ms (in the second). Thus, starting from the onset of the burst of the initial stop consonant, as little as 20 ms of each word was required to permit categorization as a ‘tr’ word or a ‘d’ word. In principle, then, once this much of a word was heard, children in the fixation task would have had sufficient acoustic-phonetic information to guide their responses on the no-overlap trials. Subjects in the gating task did not guess ‘tr’ or ‘d’ without having heard at least a fragment of the target (recall that the carrier phrases, having been spliced, were acoustically identical). Note that in subsequent analyses, we have defined the onset of the target as the onset of the initial stop closure, not the onset of the burst. Thus, strictly speaking, all four targets overlap to some degree, because the closure was equivalent in all four targets. The gating results show that the no-overlap words diverge immediately following the burst.

To gauge when children have heard speech information minimally sufficient to make the comparable distinction in the onset-overlap pairs, we may consider when subjects in the gating task first correctly identified the target word, and the point at which subjects identified the word and never subsequently changed their minds (the ‘isolation point’: Grosjean, 1980). Both measures are displayed in Table 1, along with the points at which subjects expressed a confidence above 6/10.²

The ‘first correct response’ and isolation point measures suggest that the informa-

²Confidence-based gating measures have been found to correlate with response latencies in other tasks (Tyler and Wessels, 1983). However, our goal here is to estimate when information was first available in the signal, and not to predict response latencies themselves in the fixation task. Consequently, in the following discussion we use measures that do not depend on confidence estimation to predict differences between conditions.

tion sufficient for uniquely identifying each word was available after about 370–390 ms in the do–overlap words, and about 240 ms in the tr–overlap words. Recall that the ‘tr’ versus ‘d’ distinction was available after as little as 150 ms (including 135 ms for the stop closure), suggesting that on no-overlap trials the information relevant for making the correct choice began virtually at the onset of the burst. Consequently, if children process speech continuously, fixation responses should be delayed in the onset-overlap cases by about $(240 - 150 = 90)$ ms for ‘tr’ targets, and $(390 - 150 = 240)$ ms for ‘d’ targets.

2.1.5. Apparatus and procedure

The experiment was conducted in a room containing a three-sided cloth-walled booth measuring $1 \times 1.2 \times 2$ m tall. The parent sat on a chair in the open end of the booth, holding her child on her lap facing two computer monitors (horizontally separated by about 60 cm) at the closed end of the booth. Speech stimuli were delivered through a concealed speaker on the floor directly ahead of the child. The child’s looking was monitored and recorded with a video camera positioned at a circular hole in the back of the booth slightly above the loudspeaker. A curtain hung from the side walls between the child and the parent, obstructing the parent’s view of the monitors. One experimenter controlled the procedure from an adjacent room.

The parent and child were led into the testing room by a second experimenter, who discussed the study and played with the child. The parent signed a consent form while the experimenter attempted to help the baby feel at ease. Parents had been requested to bring a completed MacArthur Communicative Development Inventory (Words and Sentences CDI; Fenson et al., 1994); this was collected. The parent was then seated in the booth, with the child on her lap. As the curtain was lowered the second experimenter presented a pair of identical pictures of shoes, while verbally encouraging the child to attend to them. This warm-up period served to familiarize the child with the loudspeaker and the position of the monitors. The first trial began when the child was engaged with the shoe-pictures and after the second experimenter slipped out of the room.

The experiment began with the four ‘ostensive’ trials on which only one picture was displayed: the dog (on the left), the doll (on the right), the tree (left), then the truck (right). Each picture was labeled with three ostensive sentences. The order of targets and their side of presentation in these trials was invariant across test orders.

Sixteen test trials followed. On each trial the presentation of the pictures preceded the speech stimulus by 4 s. This allowed the children to note the locations and identities of the pictures, and provided baseline information on each child’s preferences for individual pictures. The test phase of each trial was the 5-s period after the onset of the target word in the first sentence. Trials were separated by a 1-s pause, during which the screens were black. The entire 21-trial procedure lasted about 3.5 min.

Eight stimulus orders were created. Four of these orders were left/right reflections of the other four (excepting the ostensive trials). Condition (overlap/no-overlap) and

target side were counterbalanced by target. Each order was presented to an equal number of boys and girls.

2.1.6. Coding

During recording, videotapes of the children were time-stamped with a digital stopwatch identifying each video frame (33 ms intervals). This enabled coders to make accurate measurements of looking times to the left and right pictures by examining, frame by frame, each change in the location of children's fixations. Coding in each study was done by several highly trained coders who were blind to the condition or target side on each trial. Coders' judgments were then coordinated with information about target side and the timing of the speech stimulus, using custom software. Reliability over this data set was established by having second coders re-code a randomly selected block of four trials for each of 10 randomly selected subjects. Response latencies measured by the two coders differed by zero or one frame (33 ms) for 95.8% of latencies.

2.2. Results

Before analyzing the timing of children's shifts in fixation, it is necessary to be sure that these shifts were responses to the utterance, and not merely spontaneous refixations. To verify that this was the case, a series of analyses examined the proportion of time the children spent looking at the target, compared with the total time they spent looking at both pictures. These proportions were calculated for each of four 1-s 'windows' beginning from the onset of the target word. If children did look at the target picture in response to the label, these proportions should be greater than 0.5 (with the possible exception of the first window, during which the target word was spoken). Over the 32 children, target fixation proportions were as follows: window 1, 0.502; window 2, 0.824; window 3, 0.789; window 4, 0.696. In windows 2, 3 and 4, children fixated the target picture more than would be expected by chance (0.50; all $t(31) \geq 6.0$; all $P < 0.0001$). Performance was comparable to 24-month-olds' performance in our previous research using this method. It is clear that children's fixation responses were related to the speech stimulus; specifically, children tended to fixate the labeled target picture.

Each of the 32 children was exposed to 16 test trials, producing a total of 512 trials. At the onset of the spoken target word on each trial, a child may have been fixating the target picture, the distractor, or neither (where children may have been shifting between the pictures, or distracted; such trials compose 24% of the total). For trials on which children initially fixated the target or distractor, there are four possible outcomes. Children may: initially fixate the target, but shift to the distractor (T → D); initially fixate the target, and persist in fixating the target (T–T); initially fixate the distractor, but shift to the target (D → T); or initially fixate the distractor, and persist in fixating the distractor (D–D). In response latency analyses, we are concerned with D → T responses.

Not all D → T shifts are properly considered responses to the speech signal. For example, because the motor programming of eye movements is not instantaneous,

extremely fast shifts cannot plausibly be considered responses to the target word (having been initiated before the word began). Here we will define shifts occurring 200 ms or less after the onset of the target word as too fast to be responses to the signal. Research in younger infants has provided a range of estimates of the *minimum* latency to initiate a saccade in response to a peripheral target. The fastest estimate is about 133 ms (Canfield et al., 1997), while the most commonly-used estimate is 233 ms (Haith et al., 1988; Canfield and Haith, 1991; Dougherty and Haith, 1997). Response latencies of instructed, attentive adults in similar tasks average about 200 ms (e.g. Saslow, 1967).

Our use of 233 ms as the fastest RT is conservative (excludes relatively few trials) for two reasons. First, while minimum saccade latencies may be as small as 133 ms, mean latencies are certainly larger. While we are aware of no estimates of these measures at 24 months, mean saccade latencies at 12 months (with a minimum of about 167 ms, estimated from Canfield et al., (1997; p. 44)) are close to 300 ms, in a task involving making a saccade when a peripheral stimulus appears.

Second, these estimates come from tasks in which the currently fixated stimulus disappears before (or when) the peripheral stimulus appears (the ‘gap task’). Research with adults and infants shows that saccades are slower when the central stimulus remains visible even as the peripheral stimulus appears (the ‘overlap task;’ Fischer and Weber, 1993; Hood and Atkinson, 1993). In our task, of course, the distractor and target pictures remain visible throughout each trial, suggesting that minimum RTs in our task are greater than estimates derived from gap tasks. Consequently, it is possible that by adopting a minimum RT of 233 ms we are including a few fast shifts that are not responses to the target words, but are merely noise; however, until more data are available on 24-month-olds’ saccade response latencies in overlap tasks, we prefer a conservative exclusion criterion. In any event, in the present studies the magnitude of the effects observed, and the pattern of statistical significance, are unaffected by the choice of any lower-bound from 133 ms to 267 ms.

At the other end of the continuum, only the D → T shifts occurring within 2 s of the onset of the target word were included in analyses. Previous research with 24-month-olds has suggested that the few shifts occurring beyond this cutoff are likely to be spontaneous refixations due to flagging interest in the fixated picture (see Swingley et al., 1998, for discussion). Again, the exact criterion is not important; across all subjects there were only four D → T shifts between 1.5 and 2.5 s. The 2-s criterion also ensured that none of the shifts included in the analyses was a response to the second presentation of the target word (recall that two sentences were uttered on each trial, labeling the same picture). Thus, the response latency was defined as the number of ms between the onset of the target word (defined as the onset of the stop closure of the first phoneme), and the beginning of the child’s shift to the target picture, provided that this was between 233 and 2000 ms (inclusive). Note that this does not include the duration of the actual eye movement from one picture to the other.

Each child had two trials of each type (small capitals indicate the target, lower case the distractor): two DOGGIE-doll trials, two DOLL-dog trials, two DOGGIE-tree,

and so on. Because a response latency was only generated when children happened to fixate the distractor at target onset (which occurred on 42.8% of trials) before shifting, mean RTs for each child could not always be generated for every trial type. For example, an analysis of ‘d-’ overlap and no-overlap trials would include 18 of the 32 subjects. Consequently, the latency analyses were completed over the pool of trials produced, without regard to which subject contributed each trial. However, in every analysis the pattern of results by subjects, for those subjects contributing the relevant RTs, was equivalent to the pattern of results over trials; for example, over the 18 subjects mentioned above, the delay effect was highly significant ($t(17) = 4.7$, $P < 0.001$).

Recall that incremental interpretation of the speech signal would be indicated by delays in responding to a given target word when the distractor overlapped with the target (‘onset-overlap trials’), relative to when the distractor did not overlap (‘no-overlap trials’). As shown in Fig. 1, the predicted delay was observed in three of the four items.

An ANOVA of response latencies by condition (overlap, no-overlap) and target (doggie, doll, tree, truck) yielded main effects of condition ($F(1,168) = 19.7$, $P < 0.0001$) and target ($F(3,168) = 7.0$, $P < 0.001$), qualified by a target by condition interaction ($F(3,168) = 3.2$, $P < 0.025$). These effects were examined using one-tailed t -tests, as follows.

The response latency may be seen as a measure of the time required to reject the currently fixated object as a match for the spoken word. In these terms, it took 272 ms longer to reject the doll than the tree when hearing ‘Where’s the doggie?’ and 336 ms longer to reject the dog than the truck upon hearing ‘Where’s the doll?’ Both of these differences were significantly greater than zero (dog as target, $t(43) = 3.0$, $P < 0.003$; doll as target, $t(41) = 3.6$, $P < 0.001$), one-tailed. Recall that according to the gating pretest done with adults, a delay of about 240 ms was expected, based on duration of the overlap between the tokens of ‘doggie’ and ‘doll.’

A similar result was found in one of the tr-overlap comparisons. The comparable

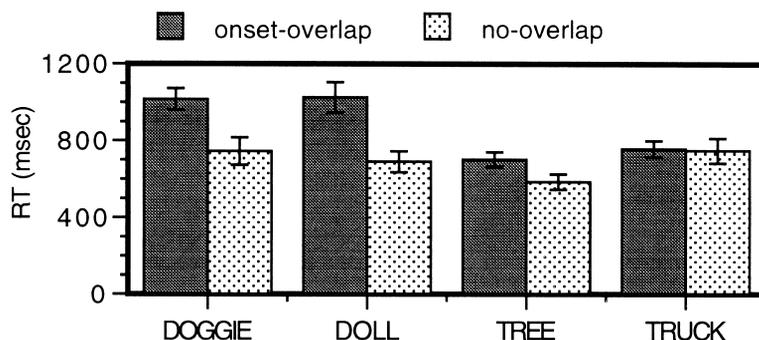


Fig. 1. Children’s mean response latencies to each item. Labels on the x-axis refer to the spoken word. In each case, the darker bar refers to the onset-overlap trials, and the lighter bar refers to the no-overlap trials. For example, for the pair of bars labeled ‘doggie,’ the dark bar refers to the dog-doll condition, and the light bar refers to the dog-tree condition. Error bars are standard errors.

response delays were 115 ms for ‘tree’ and 9 ms for ‘truck.’ Only the ‘tree’ result was significant ($t(43) = 2.0, P < 0.03$); the ‘truck’ result was not ($t(41) = 0.1, ns$). We will return to this pattern of results in Experiment 3, where we show that adults produce the same pattern of results with these stimuli.

The response latency analyses considered only the first shift in fixation after the onset of the target word, and not subsequent shifts; and only considered those trials on which children initially fixated the distractor. A fuller picture of children’s responses over time is shown in Fig. 2. Each graph shows responses by condition for the /d/ and /tr/ targets. The upper two graphs show the onset-overlap conditions, and the lower graphs show the no-overlap conditions. For example, the lower-left graph shows responses to the DOGGIE-tree and DOLL-truck trials. In each of these graphs, the curve that finishes on top (squares) is calculated from the trials on which the children initially fixated the distractor. Each point on this curve shows, for that 33-ms interval, on what proportion of trials children were looking at the target. (Given perfect performance, this curve would go to 1.0 and form a plateau.) The

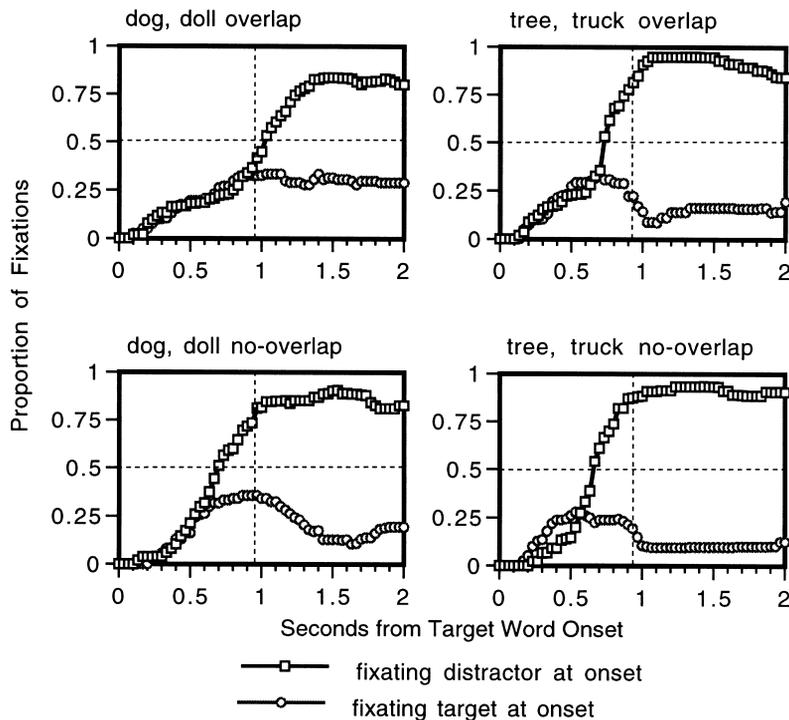


Fig. 2. Results of Experiment 1, showing children’s responses over time. Each graph shows responses collapsed over /d/ targets (left graphs) or /tr/ targets (right graphs). Responses on overlap trials are shown in the upper graphs, no-overlap on the lower graphs. In each graph, the upper curve (squares) shows results only from the trials on which the child fixated the distractor at word onset (time zero on the x-axis); the lower curve (circles) shows trials on which the child already fixated the target at word onset. The y-axis represents the proportion of trials on which at that moment children fixated the picture *other than* the one they had initially fixated. The dashed vertical lines indicate the location of the offset of the target word.

curve that finishes below (circles) is calculated from the trials on which children initially fixated the target. Each point shows the proportion of trials on which children have left the target to look at the distractor. (Given perfect performance, this curve would go to zero and stay there.) Note that the plots are not cumulative functions; the proportions displayed are calculated over the relevant set of trials at each 33-ms interval. The number of trials averaged for each curve depends on the number of children fixating the relevant picture at the target word's onset; for these graphs, this ranged from 17 to 36 trials (mean 24).

The graphs reveal that in all four cases, the curves diverge some time after the onset of the target word — about 1000 ms in the dog/doll overlap case; 700 ms in the truck/tree overlap case; and about 600 ms in the no-overlap cases. Before the divergence points, spontaneous refixation to the target (D → T) and to the distractor (T → D) occurs at a similar rate; after the divergence points, correct shifts outnumber incorrect shifts.

We interpret children's delay in responding to the onset-overlap items as evidence for continuous interpretation of the speech signal. An alternative interpretation is that the delays in the dog/doll case were due to greater visual or conceptual similarity on the part of the dog and the doll, resulting in delays in determining which picture was the referent of the target word. However, recall that children were given 4 s to look at the pictures before the utterance began. On the vast majority of trials, children did look at both pictures during this period. It would seem that this would provide an adequate opportunity for children to identify each picture, especially given that they had been exposed to the same pictures in the preliminary 'ostensive' trials. Further evidence against this interpretation comes from previous research using these pictures, but calling the doll 'baby.' In such cases, responses to the 'baby' and 'doggie' pictures (presented together) were not slower than their responses to other items (e.g. Swingley et al., 1996). If the pictures are difficult to interpret, this delay should be observed regardless of their labels. If the 'doll' label for the baby/doll picture was confusing, 'doll' as an item should lead to consistent delays (which are not found: see Experiment 2); and if the children believe that the baby/doll picture is actually a baby, there should be no onset-overlap effect in the DOGGIE-doll item of Experiment 1. Thus, the delay in the onset-overlap case cannot be attributed to difficulty in interpreting the pictures. Rather, the delay was due to the timing of the speech information relevant to the task: when the correct picture could be determined from the onset of the target word, children shifted relatively quickly; when the correct picture could only be determined from word-medial information, children shifted more slowly.

An equal number of boys and girls participated in the experiment. Girls were slightly faster (mean RT 740 ms, boys 830 ms) but this effect was marginal in a 2-tailed *t*-test using subject means ($t(30) = 1.89, P = 0.07$) and in a two-way ANOVA over trials with sex and condition (overlap, no-overlap) as factors ($F(1,172) = 3.87, P = 0.051$). There was no interaction between sex and condition. Inspection of the boys' and girls' mean RTs for each of the eight trial types revealed that in seven of the eight cases, girls were slightly faster than boys. Thus, girls appeared to hold a weak but consistent advantage.

Parents assessed their children's vocabularies using the MacArthur Communicative Development Inventory (Words and Sentences: Fenson et al., 1994). Analysis of production-vocabulary scores (range: 145–652 words; median 461) revealed a significant negative correlation between vocabulary size and mean RT (Pearson $r = -0.428$; d.f. = 30, $P = 0.014$). However, because children did not produce equal numbers of onset-overlap (slow) and no-overlap (fast) responses, comparisons between children must be made separately for each condition, to avoid 'penalizing' children who happened to contribute more RTs on overlap trials. When analyzed by condition, correlations between vocabulary size and mean RT were still negative, but no longer significant (onset-overlap $r = -0.327$, d.f. = 28, $P = 0.078$; no-overlap $r = -0.195$, d.f. = 30, $P = 0.29$). Thus, the possibility that vocabulary size might be a determinant of processing speed (or vice versa) was not convincingly supported by these data.

2.3. Discussion

Children showed significant delays in responding to 'doggie' and 'doll' when these targets had to be differentiated from overlapping alternatives. Smaller delays were found, though inconsistently, for the items for which smaller effects were predicted. The results suggest that children monitored the sentences continuously, awaited the arrival of the phonetic material distinguishing the alternative lexical possibilities, and responded when the phonetic evidence specified a particular word.

It is also possible, however, that the children in all cases waited until they had heard most of each word, and then responded; but these responses were delayed in the dog/doll case because these words are more similar to each other. In contrast to the 'incremental' interpretation described above, an 'overall similarity' interpretation might be that the heard word, and the child's representation of that word, are compared much as two images might be compared, with no essentially incremental or temporal component. On this view, comparing two similar word representations takes longer because of some difficulty in discriminating them, and this difficulty is reflected in greater latencies.

This similarity-based interpretation of Experiment 1 comports with the view that early lexical representations are 'holistic,' or non-segmental, and that young children, therefore, compare words using broad similarity criteria rather than by an incremental matching process. The argument that such differences persist in children older than 24 months is based primarily on tasks involving explicit judgments (Treiman and Baron, 1981; Treiman and Breaux, 1982; Walley, 1988). For example, in one study, adults were more likely than children to consider /bis/ and /bun/ more similar to each other than to /diz/. Similar results were obtained in an analysis of errors made by children and adults when tested for immediate recall of three newly-taught monosyllabic nonce labels: adults were relatively more likely than children to confuse words that matched in their initial consonant, as opposed to words that were more similar overall (Treiman and Breaux, 1982). If similarity judgments or encoding strategies reflect the characteristics of the lexical representations used in recog-

dition, even school-aged children might be expected to recognize words according to a non-incremental, similarity-based procedure.

Experiment 1 does not decisively favor one interpretation over another, but the speed of the children's responses on the control trials suggests that only a small portion of the spoken word needs to be heard before it can serve as a guide to which picture is being labeled. As Fig. 2 shows, in the no-overlap cases, the point at which correct responses begin to outnumber incorrect responses is at about 600 ms. This indicates that children are capable of using information quite early to guide their choices, especially when we consider the fact that eye movements cannot be made instantaneously. For this reason, it seems most parsimonious to conclude that children's responses were roughly synchronized to the timing of the information present in the signal: when the information permitting a choice of pictures was available, the response came about 1/2 s later.

The 'similarity' hypothesis, that delays in responses are due to difficulty in resolving perceptually similar words, predicts that the temporal location of the overlapping material in competing words should not matter; if children do not interpret speech continuously, but as whole, 'simultaneous' objects, such temporal considerations should have no influence. Consequently, a delay in responding should be found when the overlapping words are rhymes. Experiment 2 addresses this question. At the same time, Experiment 2 assesses whether children's representations of minimally different words are sufficiently detailed to permit accurate identification of those words.

3. Experiment 2

Experiment 2 tested whether children are able to rapidly (and correctly) distinguish words that differ only at onset. If children can rapidly distinguish rhyming words, the non-incremental 'similarity' interpretation of the results of Experiment 1 may be ruled out.

Previous studies of children's ability to distinguish minimally different words have suggested that children's phonological representations of words are not as well-specified as adults' representations. This has been tested in procedures requiring the overt selection of an object or picture from a set of at least two. Generally speaking, children's performance in such tasks has not been impressive (e.g. Garnica, 1973). Results from such studies contrast with results from infant perception studies involving minimal task demands (e.g. Jusczyk and Aslin, 1995). The present experiment examines children's ability to recognize similar words under conditions that minimize task demands.

Experiment 2 was, in a sense, a mirror image of Experiment 1. In Experiment 2, the overlapping words rhymed. The test pairs were 'ball' and 'doll,' 'duck' and 'truck' (rhyme pairs); and 'doll' and 'truck,' 'ball'; and 'duck' (control pairs). If children failed to distinguish 'ball' and 'doll' (which differ in the place of articulation of the initial consonant) or if children confused 'truck' and 'duck' (which differ in several features), this result would suggest that 24-month-olds are, indeed, unable

to reliably differentiate words differing only in their onsets, when those words are used referentially.

Another possibility is that children would perform equally well with rhymes and controls, indicating no difficulty with similar-sounding words. This finding would rule out an explanation of the results of the first experiment based on the overall similarity of the onset-overlap words. If children do not have trouble distinguishing ‘ball’ and ‘doll,’ there is no reason to believe that children should have trouble distinguishing ‘doggie’ and ‘doll’ based on their overall similarity. Also, this result would be consistent with experimental demonstrations of the abilities of young infants to discriminate minimal pairs (e.g. Aslin et al., 1998).

The preferential looking procedure permits the assessment of a final possibility: that children might perform well on the rhymes by tending to look at the target more than the distractor, but show a delay in their responses. This would indicate that children do have some trouble distinguishing very similar words, but that this trouble need not prevent children from eventually identifying the words correctly.

3.1. *Methods*

Subjects were 32 24-month-olds, ranging from 104 to 109 weeks of age (mean 108). Half were girls. All were from monolingual English environments, and low-birthweight and premature infants were excluded. All subjects knew all four target words, as assessed by parental report. An additional 24 subjects were run but were not retained in the final sample: 18 did not complete the procedure; 3 were excluded because of parental interference during the procedure; and 3 were revealed at the test session (by the parent) to not know all four target words.

The only procedural changes involved the pictures (the dog was replaced with a ball, the tree with a toy duck) and the speech stimuli. The recording procedure was the same; all new sentences were recorded. The speaking rate was somewhat faster than in the first experiment, resulting in shorter target words: ball, 585 ms; doll, 577 ms; duck, 502 ms; truck, 542 ms. Note that these measurements include the stop closure forming the onset of the initial stops; the closure durations were about 60 ms, except for ‘truck,’ which had a closure duration of 114 ms. The carrier phrases of the first sentence were not spliced in Experiment 2, because attempts at splicing revealed that coarticulation at the word ‘the’ betrayed the source token’s onset consonant, resulting in unnatural-sounding ‘ball’ utterances. Nevertheless, the carriers were all very similar. Measurements of the formant frequencies of the initial vowels of the target words are presented in Appendix A.

Coding was done by a pool of several trained coders. Six subjects were partially re-coded by a coder other than each subject’s original coder. Re-coders coded only the first shifts in fixation initiated after the onset of the first stimulus sentence of each trial. Response latencies measured by the two coders differed by zero or one frame (33 ms) for 94.4% of latencies.

3.1.1. *Preliminary evaluation of stimuli*

To verify that the target words were distinguishable shortly after their onsets, a

gating task was again used to evaluate the stimuli. The procedure was as in Experiment 1, except that for each trial, subjects were told which two of the four words were possible. This was necessary because pilot results showed that if subjects had to consider all four words, they tended to confuse ‘doll’ and ‘duck’ in early gates; these words were not paired in the fixation experiment. Thus, subjects evaluated each word with respect to its no-overlap partner and its rhyme partner on separate trials. Because this modification doubled the number of trial types (from four to eight), the number of trials was increased to 24. Also, because the carrier phrases were not spliced, coarticulatory information may have been informative before the onsets of the target words. To assess the utility of this information, each series of gated words began a few gates before the onsets of the targets. As in Experiment 1, two sets of gated stimuli were generated. Here, one set began 140 ms before the burst of each target’s initial consonant (–140, –100, –60...) and the other began 120 ms before each target’s first burst (–120, –80, –40...). Once again, the task was to identify each spoken target. Eight adults participated for a small payment.

Table 2 shows the mean point at which subjects first correctly identified each target; the point after which subjects never changed their minds (isolation point); and the point at which subjects’ confidence ratings exceeded 6/10.

The fact that adults were occasionally able to identify the words before their onsets, and nearly always before their initial consonants’ bursts, suggests that coarticulatory information present in the carrier phrase (‘Where’s the...’) was often sufficient to rule out the distractor. With the exception of the ‘duck’ case, however, the difference between the rhymes and non-rhymes was small (about 20 ms). These results suggest a refinement of our prediction of the 24-month-olds’ fixation results for the ‘duck’ item. If children are capable of processing speech incrementally, and if children can make use of pre-burst acoustic-phonetic information as a guide to lexical identity, we would expect a delay of about 80 ms in the DUCK-truck case relative to the DUCK-ball case ($85 - 7 = 78$). If young children do not use this information, response latencies would not be expected to differ by condition for any item pairs. On the other hand, if the onset-overlap delays of Experiment 1 were due to the overall similarity of the overlapping items, large rhyme delays should be obtained. Finally, regardless of children’s possible sensitivity to subtle coarticula-

Table 2

Results from gating assessment of materials from Experiment 2. Zero ms here refers to the onset of the stop closure, not the burst.

Target-distractor	1st correct response (ms)	Isolation point (ms)	Confidence >6 (ms)
BALL-doll	–20	–20	128
BALL-duck	–35	–35	107
DOLL-ball	–19	–19	101
DOLL-truck	1	1	123
DUCK-truck	85	97	182
DUCK-ball	7	7	122
TRUCK-duck	87	87	176
TRUCK-doll	59	59	164

tory information, if children are incapable of distinguishing words that differ minimally, their proportion of fixation to the target pictures should be lower in the rhyme cases than the no-overlap cases.

3.2. Results

As in Experiment 1, target fixation proportions were calculated for four 1-s ‘windows’ beginning at the onset of the target words. Over the 32 children, target fixation proportions were as follows: window 1, 0.572; window 2, 0.824; window 3, 0.791; window 4, 0.643. In each window, children fixated the target significantly more than expected by chance (0.50: all $t(31) \geq 3.0$; all $P < 0.01$). Recall that about half of window 1 includes the spoken target word; this fact accounts for the lower proportions in that window. Performance was comparable with 24-month-olds’ performance in Experiment 1.

Whether the names for the displayed pictures rhymed or not had no impact on children’s tendency to fixate the labeled pictures. As shown in Fig. 3, across the first 4 s after the onset of the target word, children were equally likely to fixate the target picture whether the distractor rhymed with the target or not. A repeated measures ANOVA of proportions of fixation to the target by condition (rhyme, no-overlap) and window yielded a main effect of window ($F(3,93) = 24.9$, $P < 0.001$), but no effect of condition ($F(1,31) = 0.005$, $P > 0.50$) and no interactions. Paired t -test analyses by trials at the level of item type (e.g. ball-doll vs. ball-duck) also failed to reveal any hints of a rhyme effect.

An ANOVA of response latencies by condition (rhyme, no-overlap) and target (ball, doll, duck, truck) yielded a main effect only for target ($F(3,165) = 3.4$, $P = 0.02$), with no interaction and no effect of condition. Item-by-item examination of mean response latencies revealed that children’s responses were not delayed by the rhyme distractors (Fig. 4). None of the within-target differences was significant by trials (largest $t(36) = 1.30$, $P > 0.10$, one-tailed). In addition, because 30/32 children provided RTs for at least one trial of each condition, it was possible to evaluate the rhyme effect within subjects (for these 30). Again, this effect was not

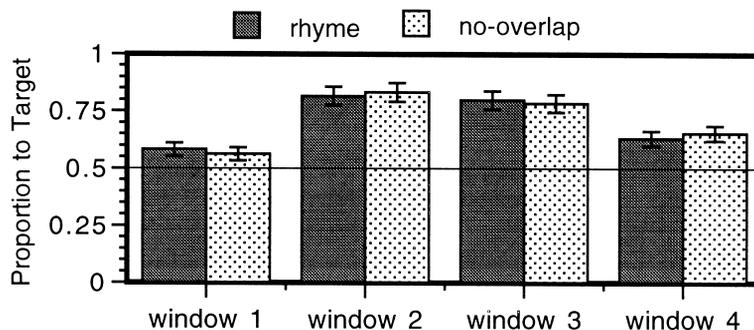


Fig. 3. Results of Experiment 2, showing children’s mean proportion of fixation to the target picture over four 1-s windows. The first window began at the onset of the target word. Error bars are standard errors.

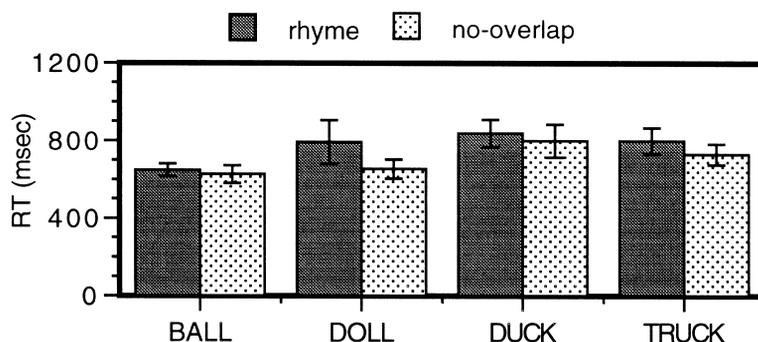


Fig. 4. Children's mean response latencies to each item in Experiment 2. Labels on the x-axis refer to the spoken word. In each case, the darker bar refers to the rhyme trials, and the lighter bar refers to the no-overlap trials. Error bars are standard errors.

significant ($t(29) = 1.6$, $P = 0.12$). The results show that distinguishing the rhyme pairs was not difficult, relative to the no-overlap pairs. The lack of a delay in the duck-truck case suggests that children did not use the coarticulatory information that permitted the adults in the gating pretest to make earlier guesses about 'duck' in the no-overlap case than the rhyme case.

Children's fixation responses over time are shown in Fig. 5. The pattern of fixations is similar in all four cases; there is no evidence that children's responses were affected by whether the distractor rhymed with the target or not. The number of trials averaged for each line depends on the number of children fixating the relevant picture at the target word's onset; for these graphs, this ranged from 21 to 31 trials (mean 26).

An equal number of girls and boys participated in the experiment. Boys were slightly faster (mean RTs by subjects: 732 ms for boys, 760 ms for girls) but this difference was not significant ($t(30) = 0.42$, $P > 0.50$). Sex and condition did not interact in a two-way repeated measures ANOVA of response latencies ($F(1,28) = 1.2$, $P > 0.25$).³

As in Experiment 1, children's productive vocabulary sizes were estimated using the MacArthur CDI (Fenson et al., 1994) to assess the relationship between lexical development and response latency. Counts of production vocabulary (range: 41–651; median 370) were negatively correlated with response latency ($r = -0.477$, $P = 0.005$). This effect held for both the rhyme pairs and the no-overlap pairs ($r = -0.438$, -0.397 , both $P < 0.03$).

3.3. Discussion

The fact that children were not delayed in recognizing very similar words like 'ball' and 'doll' rules out the 'similarity' account of the results of Experiment 1,

³This analysis was missing one girl and one boy, who did not contribute RTs in both conditions.

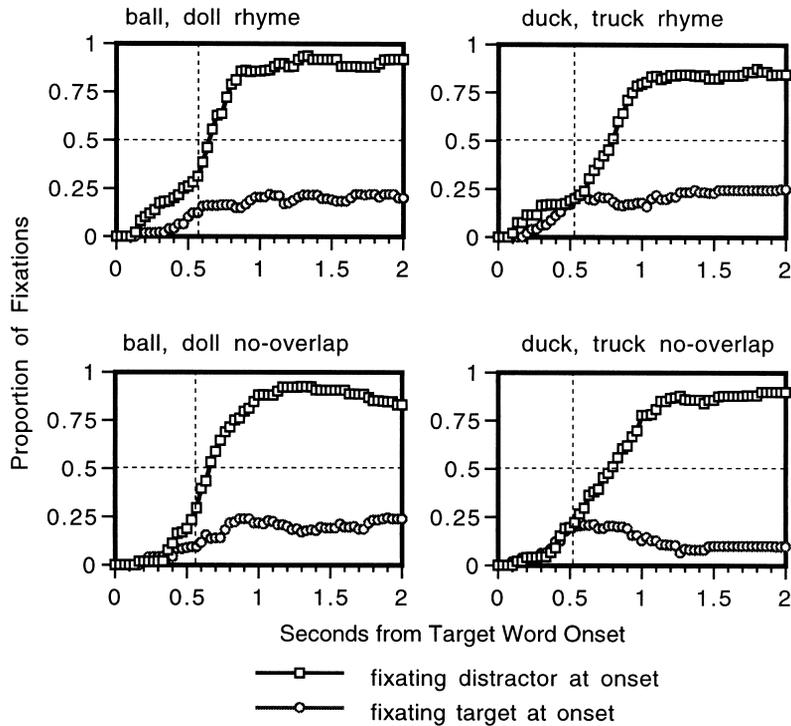


Fig. 5. Results of Experiment 2, showing children's responses over time. Each graph shows responses collapsed over 'all' targets (left graphs) or 'uck' targets (right graphs). Responses on rhyme trials are shown in the upper graphs; responses on no-overlap trials are shown in the lower graphs. In each graph, the upper curve (squares) shows trials on which children initially fixated the distractor; the lower curve (circles) shows trials on which children initially fixated the target.

namely the view that the delay evident in children's responses to the doggie/doll pair was simply a consequence of 'doggie' and 'doll' sharing an overall similarity that raised processing difficulties. The rhyme pairs of Experiment 2 were more similar than the onset-overlap pairs of Experiment 1, but did not result in any processing delays. Taken together, the first two experiments show that 24-month-olds can interpret familiar words incrementally.

The fact that children did not have measurable difficulties in distinguishing minimally different words contrasts with previous research using less familiar words. Such studies have found that children under about 24 months have more difficulty differentiating similar words than dissimilar words, and have been interpreted as showing that children do not have fully-specified lexical representations. It is possible that young children's representations of well-known, familiar words are fully specified, while their representations of newly learned words are not. This would be consistent with previous research using novel words (e.g. Shvachkin, 1973; Garnica, 1973; Eilers and Oller, 1976; Stager and Werker, 1997). If this is correct, children do not lack the basic capacities for accurate representation of lexical form; they are

simply slow to learn the forms, relative to adults (if we assume that adults quickly form accurate representations of new words).

The results found here agree with the findings of Barton (1978) with older 2-year-olds, who performed quite well in an object selection task involving minimal pairs. Experiment 2 supplements these findings by extending them to younger children and assessing for the first time children's response latencies in identifying phonetically similar words.

Experiment 2 only tested one minimal-pair contrast in one pair of words. Whether this finding may be generalized to other contrasts in the language is open to question. Evidence from consonant confusions in adults indicates that the /b-/d/ contrast is not among the most difficult (Miller and Nicely, 1955). It is possible that children's performance would have been worse on other contrasts. It also may be the case that children's representations of the initial sounds of 'ball' and 'doll' are well-specified precisely because children know both words: the need to differentiate them may force children to attend to the relevant distinctions (e.g. Jusczyk, 1986; Charles-Luce and Luce, 1990).

The failure to find a rhyme delay for the DUCK-truck item relative to the DUCK-ball item, despite the clear delay in the adult gating test, is consistent with a lack of early sensitivity to cues to word identity appearing before the burst of the initial stop consonant. Of course, it is also possible that the lack of delay may be attributed to differences between the gating and fixation tasks, or to possible idiosyncrasies of these particular items. One way to address these issues is to examine adult performance in the visual fixation task using the same stimuli that were used in Experiments 1 and 2.

4. Experiment 3

Experiment 3 used adults as subjects to provide a standard against which the children's results could be compared. Adult data in the same task may be useful in two ways: to evaluate the gating task as a predictor of adult and child response latencies; and to estimate differences in overall speed of performance between children and adults.

In Experiments 1 and 2, we predicted that children's responses would be made, on average, some fixed amount of time after the onset of information specifying the target words (as assessed using gating). While this prediction was largely upheld, it did not take into account several aspects of the experimental situation that might be expected to influence response latencies: the target word's frequency or its age of acquisition, the representativeness of the particular token of the target used, the relative attractiveness of the pictures, and so on. Where the gating data and the children's fixation results did not match well (i.e. in the failure to find the predicted onset-overlap delay for 'truck') adult subjects might show whether this discrepancy was due to a difference in the ways children and adults process speech, or some other factor not accounted for in our predictions.

Data from adults would also be useful in evaluating an alternative explanation of

the results from Experiments 1 and 2.⁴ If children's representations of the first parts of words are more robust than their representations of later parts of words, it is conceivable that children might not interpret speech incrementally, but still show delayed responding for words that are similar at onset (in which the critical phonetic information is late in the words) but not for words that are similar at offset (in which the critical information arrives early). On this 'onset salience' account, a close match between adult and child patterns of response would be coincidental, because there is no reason to expect that the effects of differential salience across words (causing delay effects in children) would be numerically similar to the effects of ambiguous acoustic-phonetic onset information (causing delay effects in adults). Alternatively, if adults and children respond similarly, a simple model could account for the results: differences in response latencies are determined by the temporal structure of the relevant acoustic-phonetic information in the speech signal.

Finally, previous research has indicated that 24-month-olds are faster in this task than 18-month-olds, who are in turn faster than 15-month-olds (Fernald et al., 1998). Testing adults in the same task permits evaluation of the eventual outcome of this developmental progression. It is not necessarily the case that response latencies here directly measure the speed with which subjects process words, because other factors such as motivation, strategy, or attentiveness might intrude; nevertheless, such data may provide the best available comparative estimates of adults' and children's processing speed in word recognition.

4.1. *Methods*

Subjects were 14 university students who participated for course credit. None was familiar with the stimuli and none was told the purpose of the experiment.

Visual and auditory stimuli were identical to those used in Experiments 1 and 2, except that on each trial only the first ('Where's the...') sentence was presented. Trials were shortened, relative to the child experiments: adults had about 2 s to fixate the pictures before the utterance began, and trials ended about 2 s after the utterance ended.

Subjects were tested in two groups, Group A ($n = 8$) and Group B ($n = 6$). Subjects in Group A were given 89 trials, including between four and six trials of each type from Experiments 1 and 2. Note that the doll/truck pair was presented in both studies; thus, there were two versions each of the DOLL-truck and TRUCK-doll types. Subjects in Group B received 78 trials of the same 16 types as subjects in A, and an additional eight 'mismatch' trials on which the spoken target word did not match either of the pictures. Those trials were not relevant to the current study and will not be discussed further. Trials were run in two randomized blocks of approximately equal length, and subjects were given 1 min to rest between blocks. Subjects in Group A went on to complete the gating task after the looking task; subjects in Group B did not.

Preceding the test blocks, subjects went through six 'ostensive' trials, naming the

⁴We thank an anonymous reviewer for suggesting this possibility.

six targets. The utterance used was ‘That’s a [target],’ taken from the ostensive sentences of Experiments 1 and 2. Adults, like children, were given the ostensive sentences so that they would be aware of the name to be used for each picture.

Children in Experiments 1 and 2 had not been explicitly instructed, except inasmuch as the sentences on each trial asked them to look at particular pictures. Adults were given the following verbal instructions:

On each trial, a picture will appear on each screen. Scan them to identify them, and then choose one to look at. In a moment you will hear a sentence saying ‘Where’s the...’ and then one of the pictures will be labeled. Your task is to look at the labeled object. This is not a race: you should try to be accurate first, and fast second. Once you have chosen a picture after hearing the word, keep looking at it until the trial is over.

The instructions were intended to persuade subjects to participate in the fixation task for the entire test period. Pilot testing showed that uninstructed subjects tended to develop idiosyncratic games and strategies.

Videotapes were coded using the same procedures as in the previous experiments.

4.2. Results

On all 1180 trials, subjects found the target, so that by about 1500 ms, target fixation was at 100%. Incorrect shifts were rare. Of 553 trials on which subjects were fixating the target at word onset, subjects incorrectly shifted to the distractor only 31 times (5.6%).

For the items from Experiment 1 (onset overlap), adults showed the same pattern of results as the 24-month-olds. Fig. 6 shows the mean response latencies for all D → T shifts, split by trial type, for the items from Experiment 1 (top graph) and Experiment 2 (bottom graph). The predicted onset-overlap delays for ‘doggie’ (172 ms) and ‘doll’ (251 ms) were found. A delay (46 ms) in the TREE-truck item was evident, but there was only a slight delay in the TRUCK-tree item (21 ms). Comparison with Fig. 1 reveals a close correspondence between adults’ and 24-month-olds’ performance. For the items from Experiment 2 (rhyme), adults showed a pattern of results similar to that of the 24-month-olds. Unlike the children, adults showed a rhyme delay in the DOLL-ball, DOLL-truck pair.

Separate two-way ANOVAs were carried out for the stimuli from Experiments 1 and 2. The former ANOVA yielded main effects of condition ($F(1,277) = 32$, $P < 0.0001$), target ($F(3,277) = 9.7$, $P < 0.0001$), and a target by condition interaction ($F(3,277) = 6.6$, $P < 0.0005$). The latter ANOVA also yielded main effects of condition ($F(1,247) = 7.5$, $P < 0.01$), target ($F(3,247) = 5.8$, $P < 0.001$), and a target by condition interaction ($F(3,247) = 2.8$, $P < 0.05$). These effects were explored using paired *t*-tests over subjects (unpaired analyses over trials rather than over subjects yielded the same pattern of results).

Not all subjects contributed a D → T response for all trial types. Consequently, between zero and five (mean 2.1) subjects were excluded from each within-subject analysis of each item. Paired one-tailed *t*-tests by subjects revealed that the onset-

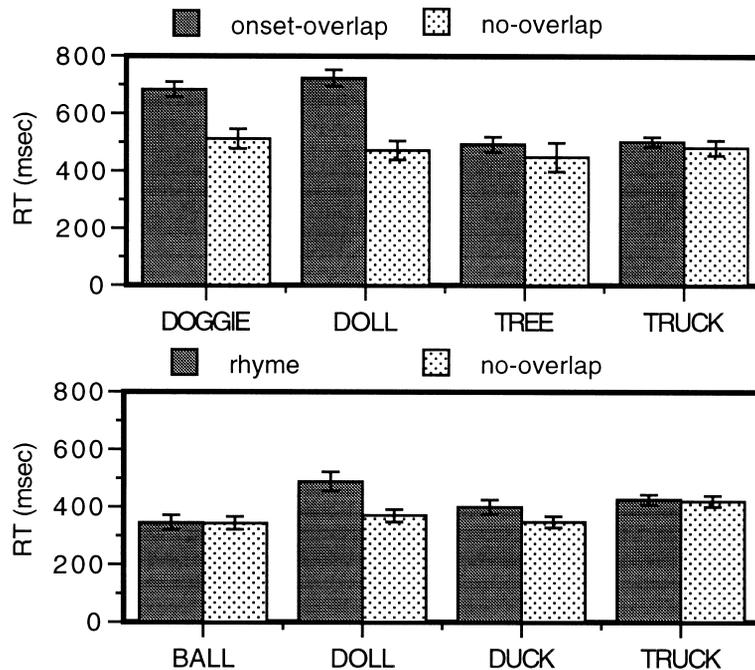


Fig. 6. Adults' mean response latencies to each item. The upper graph shows responses to the stimuli from Experiment 1; the lower graph shows responses to the stimuli from Experiment 2. Labels on the x-axis refer to the spoken word. In each case, the darker bar refers to the overlap trials, and the lighter bar refers to the no-overlap trials. Means are over trials, not over subjects. Error bars are standard errors.

overlap delays for DOGGIE and for DOLL were significantly different from zero (doggie, $t(13) = 4.4$, $P < 0.0005$; doll, $t(12) = 7.3$, $P < 0.0001$). A significant delay for the rhyme item DOLL was also found ($t(10) = 5.1$, $P < 0.001$), as well as a smaller delay for rhyme item DUCK ($t(9) = 2.0$, $P = 0.037$). No other effects were found.

Adults were faster when responding to the sentences from Experiment 2, which had been spoken at a faster rate. If all of the no-overlap pairs from the two sets of stimuli are compared, RTs to the stimuli from Experiment 2 were 106 ms faster, significant in a paired comparison by subjects ($t(13) = 4.3$, $P < 0.001$, two-tailed). It seems likely that this difference may be partially accounted for by the carrier phrases, which were spliced in Experiment 1 but not in Experiment 2. As the Experiment-2 stimulus gating results revealed, targets were sometimes identifiable before their onsets. If adults were able to use this information in making their responses on-line, slightly faster responses to the Experiment 2 stimuli would be expected.

It is also possible that the faster target words conveyed more information in their initial 250 ms or so, permitting greater (or earlier) confidence about lexical identity. However, an informal comparison of the first 250 ms of the target words of each

experiment suggests that this explanation is unlikely to be true: the initial 250 ms of each target clearly specifies the onset consonant(s) and the long steady-state portions of the vowels, and nothing more, for both sets of target words. The two stimulus sets do not appear to be differentially informative over the portion of the targets that evidently determined adults' responses. Therefore it is improbable that adults responded faster to the more rapidly spoken words because they were more informative shortly after onset.

In contrast to these adult results, 24-month-olds did not respond more quickly to the no-overlap items of Experiment 2 (mean no-overlap RT: 713 ms) than those of Experiment 1 (692 ms). Perhaps adults, but not children, used coarticulatory information present in the carrier phrases of Experiment 2 to identify the words quickly.

The carrier phrases may also help explain the difference between the DUCK-truck and DUCK-ball adult RTs. Adults responded to 'duck' more slowly when it was paired with 'truck' than when it was paired with 'ball' (by subjects, a 65-ms effect). Recall that subjects in the gating task also identified 'duck' later when it was paired with 'truck' than when paired with 'ball,' an effect of about 80 ms. In other words, the apparent rhyme delay seen in the 'duck' case is simply a consequence of subjects' ability to rule out 'ball' sooner than 'truck' based on the carrier phrases.

The same explanation does not hold for the rhyme delay seen for 'doll.' Adults were (by subjects) 117 ms slower in responding to 'doll' with the rhyme distractor 'ball' than with the no-overlap distractor 'truck.' The gating results predicted, at most, a 20 ms effect in the other direction. The most likely explanation is that the token of 'doll' used in the experiment did not match some subjects' representations of that word; for them, the words 'doll' and 'ball' do not rhyme, and the vowel of the experimental token of 'doll' (/ and /o/) matched 'ball' better than 'doll.' This account is corroborated by examination of adults' errors: trials on which subjects initially fixating the target shifted away. Of the 31 such errors experiment-wide, 21 involved the word 'doll,' suggesting a relatively poor fit between the tokens and adults' representations. Of course, while this account is consistent with the data, further research would be required to conclusively establish such error effects (which were seen here on a small proportion of trials, and which were not replicated in the 24-month-old data).

Finally, adults were shown to be faster than children. If we compare response latencies from the Experiment 1 stimuli (in which there is no question of coarticulatory information in the spliced carriers), adults responded in 558 ms, on average, while 24-month-olds responded in 785 ms. This 227 ms difference was statistically significant ($t(44) = 5.4$, $P < 0.0001$). Considering only the no-overlap trials from the same set of stimuli, the difference between adults and children was nearly the same (206 ms).

4.3. Discussion

Adults' responses were very similar to children's responses, both in the overall pattern of results, and in the magnitude of the differences between conditions. For the words that overlapped at onset for about 200 ms, corresponding delays were

found in response latencies at both ages. Words overlapping less at onset produced either small delays, or no delay, in both adults and children.

It is not clear why subjects in both age groups rejected the tree picture as quickly as the dog picture on ‘truck’ trials, when the gating test predicted a difference of about 100 ms. Whether this lack of effect is due to noise, genuine differences between gating and fixation as response measures, or an interaction between the particular pictures and speech samples used in the current studies, cannot be determined from the available data. However, at present the important point to note is that the adults and children behaved similarly on the onset-overlap comparisons.

Two unpredicted differences between adults and children were discovered. First, adults, but not children, responded more quickly to the stimuli from Experiment 2 than the stimuli from Experiment 1, suggesting that adults were better able to take advantage of the cues to the target words’ identities present in the carrier phrase. While infants are sensitive to certain coarticulatory effects (Fowler et al., 1990), children’s ability to extract lexically relevant information from subtle coarticulatory cues is unknown. Adults, however, are extremely sensitive to such information (e.g. Martin and Bunnell, 1982), and this sensitivity has important consequences for the recognition of words (Warren and Marslen-Wilson, 1987; Marslen-Wilson and Warren, 1994).

Second, adults (but not children) were shown to be delayed in responding to ‘doll’ when it was paired with the ball. It seems likely that this difference was due to our use of what may have been a non-prototypical vowel for the word ‘doll.’ This vowel did match the prototypical vowel for ‘ball,’ leading to brief confusion in the DOLL-ball items. This effect was not found in the gating pretest, suggesting that subjects in the gating task were able to focus their attention on the initial consonants, and did not consider the ‘doll’ vowel to be decisive.

One purpose of Experiment 3 was to evaluate absolute differences in speed between adults and 24-month-olds. Results showed that adults were indeed faster than 24-month-olds, but the differences were modest: about 215 ms. With the current data we may only speculate about the sources of the difference. First, it is possible that adults were more attentive than children and more inclined to respond as quickly as possible. Many children appeared to enjoy the task and often seemed to be treating it as a game, but this is no guarantee that they were attempting to respond as soon as they were able to.

Another possibility is that part of the difference between children and adults was due to developmental differences in the speed with which eye movements may be programmed. However, such differences are likely to be slight, given the early maturation of the oculomotor system (e.g. Haith et al., 1988). As described previously, in tasks requiring saccadic responses to peripheral targets, mean RTs at 12 months are around 300 ms, while instructed adults average around 200 ms. It seems likely, then, that less than half of the 215 ms developmental difference in latency found in the present studies should be attributed to oculomotor speed.

We believe that the difference in mean response latency between 24 months and adulthood is most likely to be a reflection of adults’ greater experience in word recognition. It would be surprising, indeed, if 18 years of practice with spoken

language did not result in some improvement in performance. Our data do not allow us to specify the locus of this improvement. The task involves recognizing a word, comparing it with a picture, and deciding whether to shift in fixation or not. Adults could be faster and more accurate in any or all of these processes.

The fact that 24-month-olds were slower than adults is consistent with claims that children need to hear more of a word to recognize it than adults do (Huttenlocher, 1984; Walley, 1988; Charles-Luce and Luce, 1990). However, the close correspondence between the adults' and children's results suggests that subjects in both age groups operate with similar representations of the sound forms of familiar words, and that these representations are activated in a continuous manner as acoustic-phonetic information arrives.

5. General discussion

One of the fundamental assumptions of current models of word recognition is that lexical activation is continuously modulated by information in the speech signal. The purpose of the present research was to determine whether children in the early stages of language acquisition also exhibit continuous interpretation of the speech signal, and to test the specificity of the lexical representations children use for on-line word recognition. In Experiment 1, 24-month-olds' fixation responses were delayed when a spoken word was, for a brief period, equally compatible with two pictures. The delay persisted until the spoken word was no longer compatible with one of the pictures; then children responded rapidly. In Experiment 2, 24-month-olds showed no delay in responding when the names for the pictures were rhymes, showing that children had no difficulty resolving similar words. Finally, Experiment 3 largely replicated these results in a sample of adults. Taken together, the striking similarities between adults' and children's performance on this task show that by 24 months, children process speech incrementally, as adults do, and represent the sound forms of words with greater specificity than previously shown in more demanding tasks.

We do not claim that lexical processing in infants is as sensitive and robust as it is in adults. Every effort was made to minimize the demands of the task: targets were familiar words; these words were presented sentence-finally in slow, clear speech; and the two possible referents were displayed in advance. In the vast majority of studies demonstrating adults' skill in using acoustic-phonetic information to recognize words in speech, contextual cues to lexical identity have been minimal or absent, meaning that subjects have had to consider their entire lexicons as potential experimental items (in a lexical decision task, for example). As reviewed in Section 1, children often perform poorly relative to adults on more difficult tasks designed to assess lexical discrimination, particularly when these tasks involve abstract judgments (e.g. Gerken et al., 1995; Walley et al., 1995) or use newly-learned words (e.g. Shvachkin, 1973). It remains to be seen whether the abilities demonstrated here generalize to situations in which the possible referents of the sentence are not so clearly specified, or to less familiar words.

However, given the present results, poor linguistic performance in other tasks cannot be explained by assuming that young children lack the basic representational or procedural capacities that are necessary for adult-like performance in word recognition, at least in terms of the use of speech information over time. In other words, it is clear that word recognition in the daily lives of young children rarely involves a two-alternative discrimination (or even a single go/no-go decision). But poor performance in other tasks or situations cannot be attributed to a processing architecture that does not work incrementally, or to a representational scheme incapable of rapidly differentiating similar words.

Several studies showing only mediocre performance in the differentiation of minimal-pair words used words that had recently been taught (e.g. Shvachkin, 1973). Though we draw different conclusions about children's perceptual abilities, our results using familiar words are not inconsistent with these previous findings. It is reasonable to suppose that 24-month-olds require considerable exposure to words to achieve the standards of performance seen here, and it is likely that adults would require less training. If so, this might be a consequence of a specifically linguistic ability that is not available to young children, or a consequence of more general learning abilities present only in adults. For example, adults learning new words may mentally spell them out, rehearse them, or otherwise make conscious efforts at learning. Children are less likely than adults to use such strategies (e.g. Flavell et al., 1966). Research using the present task, but with novel or more recently learned words, could shed light on these issues.

The current studies do not demonstrate conclusively that children's representations of the tested words were accurate. For example, a child whose representations of 'doll' and 'dog' started with /k/ might nevertheless be able to perform reasonably well in these tasks. The ball/doll comparison serves only as a test for sensitivity to place of articulation in /b/ and /d/ onsets. However, proposals that children's early word representations are inaccurate usually assume that the inaccuracies are a matter of vagueness or underspecification, rather than precise but incorrect specification. Though we believe that 24-month-olds' speed and accuracy in the fixation task makes this sort of vagueness unlikely, the current results do not rule it out.

The results of Experiments 1 and 2 are consistent with a growing body of research revealing considerable language-specific knowledge in infants, and with evidence of infants' ability to distinguish minimally different consonants (see e.g. Gerken, 1994; Goodman and Nusbaum, 1994; Jusczyk, 1997). Children's unimpressive performance in more difficult comprehension tasks involving minimal pairs (and, usually, newly-learned words) had created the appearance of a developmental paradox in which infants discriminate the sounds of their language, but children fail to show this ability consistently. However, we found no evidence that children and adults differ in their ability to distinguish similar well-known words.

Children's incremental interpretation of speech renders them susceptible to the spurious activation of words that are temporarily consistent with the speech signal. Children hearing 'dog' will momentarily activate 'doll;' 'candle' will activate 'candy.' In the latter case, if the child is to learn 'candle,' it is important that the early activation for 'candy' does not prevent detection of the mismatch between the

expected /di/ and the actual /dl/. If children were over-confident of lexical identity on the basis of partial initial information, it could prevent the detection of, and therefore the learning of, some novel words. Such effects are not found in adults, who are capable of detecting mispronounced words whose onsets match known words. Nevertheless, the possibility that children might not share this ability has led to the suggestion that young children might be better off without incremental word recognition (Walley, 1993). Given that children do appear to interpret speech incrementally, however, we might well inquire whether novel words whose onsets are similar to those of known words are more difficult for children to learn.

Such words are learned, of course, and some are learned quite early, as shown by children's use of words varying in their suffixes (e.g. Brown, 1973). If activation of an anticipated 'car' somehow swamped detection of a final -s, children would not learn 'cars.' Morphological variations of a given root are obviously a poor test case for the learning of onset-overlap words, not least because 'cars' is not really a novel word in the same sense as, say, 'cart.' However, the suffixing example raises the possibility that incremental processing actually helps children recognize the connection between the various morphological forms of words. If the overall similarity of words governed word recognition, 'washing' and 'washed' might be perceived as quite different. On an incremental account, hearing 'wash...' should activate the semantics of 'wash,' possibly before the suffix is even encountered. This might serve to enhance, rather than obscure, detection of the following affix. Research using mispronunciation-detection has found that children seem to devote greater attention to phonemic detail in highly predictable sentential contexts (Cole and Perfetti, 1980); the same mechanism could aid in the discovery of suffixes, even if at the expense of unrelated but similar words like 'dog' and 'doll.'

Incremental processing is one of several well-established characteristics of word recognition in adults, who also show processing advantages for more frequent words, facilitation of recently-heard words, interlexical competition, and semantic priming, among other effects. Young children may differ from adults in any of these facets of language processing. However, our results show that there are important similarities in the processing problems faced by both adults and very young children, and suggest that the solutions to these problems may be similar as well. A challenge for research on early speech perception will be to understand the development of these similarities between adults and children, and to determine their extent. The challenge this poses for models of word recognition is to account for on-line speech processing even as language and the lexicon are being acquired.

Acknowledgements

This work was supported by grants from the National Institutes of Child Health and Human Development (NIH) Grant MH41511 (AF) and the National Science Foundation (D.S., J.P.P.). Portions of the research reported here were presented at the Boston University Conference on Language Development (1997) and at the International Conference on Infant Studies (1998). The authors would like to extend

Table 3
Formant frequencies of the (first) vowel of each target word.

Target	F_1 (Hz)	F_2 (Hz)	F_3 (Hz)
<i>Experiment 1</i>			
doggie (/ɔ/)	743	1054	2869
doll (/ɔ/)	738	1037	2813
tree (/i/)	316	2760	3235
truck (/ʌ/)	720	1389	2778
<i>Experiment 2</i>			
ball (/ɔ/)	597	984	2831
doll (/ɔ/)	615	984	2753
duck (/ʌ/)	720	1512	2444
truck (/ʌ/)	760	1503	2575

their thanks to the students of the Center for Infant Studies at Stanford for their assistance. In addition, we are grateful to Ellen Markman, Herb Clark, and Anne Cutler for discussions about the research, and to Delphine Dahan and Michael Tanenhaus for comments on the manuscript.

Appendix A. Vowel frequencies

Kay Elemetrics CSL model 4300B hardware was used to measure the first three formant frequencies of the initial vowels of the target words in each experiment. The pitch of the vowels often varied considerably (as is common in infant-directed speech); analyses considered the most stable 100–120 ms of each vowel, judged by eye from spectrograms. Measurements were based on LPC analysis, and verified using spectrograms. As shown in Table 3, formant frequency values were similar in vowels that were intended to overlap in each experiment.

References

- Alloppenna, P.D., Magnuson, J.S., Tanenhaus, M.K., 1998. Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language* 38, 419–439.
- Aslin, R.N., Jusczyk, P.W., Pisoni, D.B., 1998. Speech and auditory processing during infancy: constraints on and precursors to language. In: Kuhn, D., Siegler, W.D.R. (Eds.), *Handbook of Child Psychology, Volume 2: Cognition, Perception, and Language*. Wiley, New York, pp. 147–254.
- Barton, D., 1978. The discrimination of minimally-different pairs of real words by children aged 2;3 to 2;11. In: Waterson, N., Snow, C. (Eds.), *The Development of Communication*. Wiley, New York, pp. 255–261.
- Barton, D., 1980. Phonemic perception in children. In: Yeni-Komshian, G.H., Kavanagh, J.F., Ferguson, C.A. (Eds.), *Child Phonology, Volume 2: Perception*. Academic Press, New York, pp. 97–116.
- Benedict, H., 1979. Early lexical development: comprehension and production. *Journal of Child Language* 6, 183–200.

- Brown, R., 1973. *A First Language*. Harvard University Press, Cambridge, MA.
- Canfield, R.L., Haith, M.M., 1991. Young infants' visual expectations for symmetric and asymmetric stimulus sequences. *Developmental Psychology* 27, 198–208.
- Canfield, R.L., Smith, E.G., Breznsnyak, M.P., Snow, K.L., 1997. Information processing through the first year of life. *Monographs of the Society for Research in Child Development*, 62, Serial Number 250.
- Charles-Luce, J., Luce, P.A., 1990. Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language* 17, 205–215.
- Cole, R.A., Perfetti, C.A., 1980. Listening for mispronunciations in a children's story: the use of context by children and adults. *Journal of Verbal Learning and Verbal Behaviour* 19, 297–315.
- Connine, C.M., Blasko, D.G., Titone, D., 1993. Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language* 32, 193–210.
- Cooper, R.M., 1974. The control of eye fixation by the meaning of spoken language. *Cognitive Psychology* 6, 84–107.
- Dougherty, T.M., Haith, M.M., 1997. Infant expectations and reaction time as predictors of childhood speed of processing and IQ. *Developmental Psychology* 33, 146–155.
- Eilers, R.E., Oller, M.K., 1976. The role of speech discrimination in developmental sound substitutions. *Journal of Child Language* 3, 319–329.
- Fenson, L., Dale, P.S., Reznick, J.S., Bates, E., Thal, D.J., Pethick, S.J., 1994. Variability in early communicative development. *Monographs of the Society for Research in Child Development* 59, Serial Number 242.
- Fernald, A., Mazzie, C., 1991. Prosody and focus in speech to infants and adult. *Developmental Psychology* 27, 209–221.
- Fernald, A., Pinto, J.P., Swingley, D., Weinberg, A., McRoberts, G.W., 1998. Rapid gains in speed of verbal processing by infants in the second year. *Psychological Science* 9, 228–231.
- Fischer, B., Weber, H., 1993. Express saccades and visual attention. *Behavioural and Brain Sciences* 16, 553–610.
- Flavell, J.H., Beach, D.R., Chinsky, J.M., 1966. Spontaneous verbal rehearsal in a memory task as a function of age. *Child Development* 37, 283–299.
- Fowler, C.A., Best, C.T., McRoberts, G.W., 1990. Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception and Psychophysics* 48, 559–570.
- Garnica, O.K., 1973. The development of phonemic speech perception. In: Moore, T.E. (Ed.), *Cognitive Development and the Acquisition of Language*. Academic Press, New York, pp. 215–222.
- Gerken, L., 1994. Child phonology: past research, present questions, future directions. In: Gernsbacher, M.A. (Ed.), *Handbook of Psycholinguistics*. Academic Press, San Diego, pp. 781–820.
- Gerken, L.A., Murphy, W.D., Aslin, R.N., 1995. Three- and four-year-olds' perceptual confusions for spoken words. *Perception and Psychophysics* 57, 475–486.
- Golinkoff, R.M., Hirsh-Pasek, K., Cauley, K.M., Gordon, L., 1987. The eyes have it: lexical and syntactic comprehension in a new paradigm. *Journal of Child Language* 14, 23–45.
- Goodman, J.C., Nusbaum, H.C., 1994. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. MIT Press, Cambridge, MA.
- Graham, L.W., House, A.S., 1971. Phonological oppositions in children: a perceptual study. *Journal of the Acoustical Society of America* 49, 559–566.
- Grosjean, F., 1980. Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 28, 267–283.
- Haith, M.M., Hazan, C., Goodman, G.S., 1988. Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Development* 59, 467–479.
- Hood, B.M., Atkinson, J., 1993. Disengaging visual attention in the infant and adult. *Infant Behaviour and Development* 16, 405–422.
- Huttenlocher, J., 1974. The origins of language comprehension. In: Solso, R.L. (Ed.), *Theories in Cognitive Psychology: The Loyola Symposium*. Erlbaum, Potomac, NJ, pp. 331–368.
- Huttenlocher, J., 1984. Word recognition and word production in children. In: Bouma, H., Bouwhuis, D.G. (Eds.), *Attention and Performance X: Control of Language Processes*. Erlbaum, Hillsdale, NJ, pp. 447–457.

- Jusczyk, P.W., 1986. Towards a model for the development of speech perception. In: Perkell, J., Klatt, D.H. (Eds.), *Invariance and Variability in Speech Processes*. Erlbaum, NJ, Hillsdale, pp. 1–19.
- Jusczyk, P.W., 1997. *The Discovery of Spoken Language*. MIT Press, Cambridge, MA.
- Jusczyk, P.W., Aslin, R.N., 1995. Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology* 29, 1–23.
- Jusczyk, P.W., Hohne, E.A., 1997. Infants' memory for spoken words. *Science* 277, 1984–1986.
- Marslen-Wilson, W., 1984. Function and process in spoken word recognition. In: Bouma, H., Bouwhuis, D.G. (Eds.), *Attention and Performance X: Control of Language Processes*. Erlbaum, Hillsdale, NJ, pp. 125–150.
- Marslen-Wilson, W., 1993. Issues of process and representation in lexical access. In: Altmann, G.T.M., Shillcock, R. (Eds.), *Cognitive Models of Speech Processing: The Second Sperlonga Meeting*. Erlbaum, Hillsdale, NJ, pp. 187–211.
- Marslen-Wilson, W., Moss, H.E., van Halen, S., 1996. Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 22, 1376–1392.
- Marslen-Wilson, W., Warren, P., 1994. Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review* 101, 653–675.
- Marslen-Wilson, W.D., Welsh, A., 1978. Processing interactions during word-recognition in continuous speech. *Cognitive Psychology* 10, 29–63.
- Martin, J.G., Bunnell, H.T., 1982. Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance* 8, 473–488.
- McClelland, J.L., Elman, J.L., 1986. The TRACE model of speech perception. *Cognitive Psychology* 18, 1–86.
- Miller, G.A., Nicely, P.E., 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27, 338–352.
- Moss, H.E., McCormick, S.F., Tyler, L.K., 1997. The time course of activation of semantic information during spoken word recognition. *Language and Cognitive Processes* 12, 695–731.
- Pisoni, D.B., Luce, P.A., 1987. Acoustic-phonetic representations in word recognition. *Cognition* 25, 21–52.
- Saslow, M.G., 1967. Latency for saccadic eye movement. *Journal of the Optical Society of America* 57, 1030–1033.
- Shvachkin, N.K., 1973. Original work published 1948. The development of phonemic speech perception in early childhood. In: Ferguson, C.A., Slobin, D.I. (Eds.), *Studies of Child Language Development*. Holt, Rinehart, and Winston, New York, pp. 91–127.
- Stager, C.L., Werker, J.F., 1997. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature* 388, 381–382.
- Swingley, D., Fernald, A., McRoberts, G.W., Pinto, J.P., 1996. Prosody, functors, and word recognition in young children. In: *Proceedings of the 20th Annual Boston Conference on Language Development*, Cascadilla Press, Boston, MA, pp. 760–767.
- Swingley, D., Pinto, J.P., Fernald, A., 1998. Assessing the speed and accuracy of word recognition in infants. In: Rovee-Collier, C., Lipsitt, L.P., Hayne, H. (Eds.), *Advances in Infancy Research*. Ablex, Stamford, CT, pp. 257–277.
- Taft, M., Hambly, G., 1986. Exploring the cohort model of spoken word recognition. *Cognition* 22, 259–282.
- Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., Sedivy, J.C., 1996. Using eye movements to study spoken language comprehension: evidence for visually mediated incremental interpretation. In: Inui, T., McClelland, J.L. (Eds.), *Attention and Performance 16: Information Integration in Perception and Communication*. MIT Press, Cambridge, MA, pp. 457–478.
- Thomas, D.G., Campos, J.J., Shucard, D.W., Ramsay, D.S., Shucard, J., 1981. Semantic comprehension in infancy: a signal detection analysis. *Child Development* 52, 798–803.
- Treiman, R., Baron, J., 1981. Segmental analysis ability: development and relation to reading ability. In: MacKinnon, G.E., Waller, T.G. (Eds.), *Reading Research: Advances in Theory and Practice*, Volume 3. Academic Press, New York, pp. 159–198.

- Treiman, R., Breaux, A.M., 1982. Common phoneme and overall similarity relations among spoken syllables: their use by children and adults. *Journal of Psycholinguistic Research* 11, 569–598.
- Tyler, L.K., Wessels, J., 1983. Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics* 34, 409–420.
- Walley, A.C., 1988. Spoken word recognition by young children and adults. *Cognitive Development* 3, 137–165.
- Walley, A.C., 1993. The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review* 13, 286–350.
- Walley, A.C., Michela, V.L., Wood, D.R., 1995. The gating paradigm: effects of presentation format on spoken word recognition by children and adults. *Perception and Psychophysics* 57, 343–351.
- Warren, P., Marslen-Wilson, W., 1987. Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics* 41, 262–275.
- Zwitserslood, P., 1989. The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition* 32, 25–64.
- Zwitserslood, P., Schriefers, H., 1995. Effects of sensory information and processing time in spoken-word recognition. *Language and Cognitive Processes* 10, 121–136.