ELSEVIER

COGNITION

Cognition 76 (2000) 147–166

www.elsevier.com/locate/cognit

Spoken word recognition and lexical representation in very young children

Daniel Swingley*, Richard N. Aslin

Department of Brain and Cognitive Sciences, Meliora Hall, University of Rochester, Rochester, NY 14627, USA

Received 18 October 1999; received in revised form 2 February 2000; accepted 24 March 2000

Abstract

Although children's knowledge of the sound patterns of words has been a focus of debate for many years, little is known about the lexical representations very young children use in word recognition. In particular, researchers have questioned the degree of specificity encoded in early lexical representations. The current study addressed this issue by presenting 18–23-month-olds with object labels that were either correctly pronounced, or mispronounced. Mispronunciations involved replacement of one segment with a similar segment, as in 'baby–vaby'. Children heard sentences containing these words while viewing two pictures, one of which was the referent of the sentence. Analyses of children's eye movements showed that children recognized the spoken words in both conditions, but that recognition was significantly poorer when words were mispronounced. The effects of mispronunciation on recognition were unrelated to age or to spoken vocabulary size. The results suggest that children's representations of familiar words are phonetically well-specified, and that this specification may not be a consequence of the need to differentiate similar words in production. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Spoken word recognition; Lexical representation; Very young children

1. Introduction

When children learn to understand a spoken language, they must remember and

---

* Corresponding author. Present address: Max-Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands. Tel.: +31-24-352-1380; fax: +31-24-352-1213.

E-mail address: daniel.swingley@mpi.nl (D. Swingley).

PII: S0010-0277(00)00081-0

recognize the sound patterns of words. This is a non-trivial problem because distinct instances of the same word can have very different acoustic manifestations. The realization of a word varies with talker, rate, context, and several other factors. Through a process that is not well understood, mature speakers of a language are capable of recognizing words rapidly and efficiently in spite of this variation. It is usually assumed that the lexicon of a fluent speaker (and hearer) contains a representation of each word in an abstract, idealized form which is matched to heard speech; good matches result in rapid lexical activation and word recognition.

Because words vary across languages, lexical forms must be learned, and success in word recognition across speakers and contexts depends upon learning lexical representations and developing the matching process that links speech to these representations during word recognition. Very little is known about very young children's ability to do this, apart from the fact that children do recognize (and eventually say) words. Descriptions of early lexical representations have reflected a tension between experimental results showing that infants appropriately discriminate the sounds of the languages they hear (reviewed in Aslin, Jusczyk & Pisoni, 1998), and results with older children suggesting that in spite of these discrimination abilities, early representations of words are not specified in full phonetic detail (as discussed below). Several researchers have pointed out that for young children who know few similar-sounding words, full phonetic specification may not be necessary; perhaps children benefit from the sparseness of their lexicons by encoding only the detail necessary to distinguish known words (Charles-Luce & Luce, 1990; Jusczyk, 1986; Walley, 1993).

Most research testing the specificity of early lexical representations has assessed the discrimination of words differing by a single phonetic segment, or 'minimal pairs'. This research originated with a study by Shvachkin (1948/1973), who taught Russian children monosyllabic nonsense words (such as 'dak' and 'gak') for novel objects, and then tested children's ability to tell the words apart by presenting the objects and requesting one of them. Children ranging in age from 10 to 24 months failed to consistently discriminate words differing by any one of several consonant contrasts. Garnica (1973), in a similar study of American children between 17 and 22 months old, also found inconsistent word-discrimination performance. For example, in the first half of her 4-month study, only one in five children performed above chance in distinguishing words differing in the voicing of the initial consonant. Eilers and Oller (1976) also reported highly variable performance in 22–26-month-olds, noting that some contrasts were considerably more difficult than others.

However, results from studies using newly-taught words might not reflect children's representations of more familiar words. Barton (1976, 1978), using a manual picture selection task, found that children between 27 and 35 months of age were largely successful at differentiating words like 'bear' and 'pear' if children had used both words productively without prompting (success rate 89.2%). Barton noted that performance was poor when the words had to be taught (success rate 47.7%; figures from Barton, 1976, p. 63). Thus, failures to discriminate contrasts in new words cannot be assumed to characterize knowledge of familiar words.

Unfortunately, children younger than those studied by Barton tend to know very few minimal-pair words, which has largely thwarted further attempts to use this picture-pointing method to examine 1-year-olds' representations of familiar words. As a consequence, the existing literature on early lexical representations relies on children's speech production, and on studies of the perceptual abilities of infants. These two sources of evidence provide an incomplete picture. First, the nature of the relationship between children's productive and receptive lexical representations is obscure. For example, when children mispronounce a word, this does not necessarily imply ignorance of how the word *should* sound (e.g. Dodd, 1975; Kiparsky & Menn, 1977; Smith, 1973, 1978). Second, even when children correctly pronounce a word, this does not imply that children would reject the same mispronunciations that adults would. Third, children know words they do not say at all. Thus, while children's speech is an invaluable resource for revealing early linguistic knowledge, it cannot decisively establish whether children have well-specified representations for recognition, particularly when children have small vocabularies or do not articulate clearly. To understand the complex relationship between receptive and productive lexical representations, they must be assessed independently, a point made by many investigators (e.g. Jusczyk, 1993; Shattuck-Hufnagel, 1986).

Although many studies of infant speech perception suggest that infants represent and categorize speech sounds in a manner appropriate for language learning, some of these results may not be generalizable to real-world language use. For example, although young infants dishabituate when a series of CV syllables (e.g. ba…ba…ba…) changes to another CV series (pa…), this does not necessarily imply that infants would also categorize a maternal utterance of, say, 'That's a bat' differently from an utterance of 'That's a pat', or would recognize the identity of 'bat' in sentences spoken in different contexts by different talkers. Discrimination studies generally do not require categorization over sets of syllables varying in talker, rate, and sentence context, although proper categorization in spite of this variability is essential for word recognition. This distinction between simplified laboratory contexts and speech in natural contexts has often been invoked to reconcile infants' ability to discriminate minimal pairs, and young children's apparent failure to differentiate minimally different words (e.g. Locke, 1988).

However, more recent studies using relatively naturalistic stimuli have shown that infants recognize at least some subtle distinctions between words. For example, Jusczyk and Aslin (1995) familiarized 8-month-olds with spoken lists of tokens of a word such as 'dog', and then assessed infants' preferences for six-sentence passages containing 'dog' or not containing 'dog'. Infants preferred the passages containing the familiarized word. However, this preference was *not* shown among infants familiarized to a similar word, such as 'bog'. These results suggest that some phonetic detail is preserved, at least for a short time, by 8-month-olds, even in a situation incorporating some of the natural variability of speech.

If even 8-month-olds maintain well-specified representations of words in sentential contexts, as suggested by Jusczyk and Aslin (1995), why do young children in Shvachkin-type tasks so frequently fail to distinguish minimal pairs? One possibility is that the tasks used with infants reflect the detectability of changes (or of similarity) in the acoustic properties of speech sounds, whereas the tasks used with young children reflect the nature of children's word recognition systems, which may operate according to different criteria. Word recognition should be robust over (i.e. ignore) certain types of variation, such as speaker identity, while maintaining fine distinctions in other types of variation, such as voice onset time. If 1-year-olds have not discovered which types of variation are relevant in word recognition, they might fail to encode some important phonetic details in their lexicons. This hypothesis highlights the need to test children's lexical representations using referential tasks, rather than tasks that only assess the familiarity or discriminability of the sound patterns of words.

A second possibility is that young children are in principle capable of distinguishing relevant and irrelevant variation in speech, but in practice cannot make use of this ability in word learning because the attentional demands of word learning are too great. Thus, when discriminating speech sounds alone, children perform well, but given concurrent presentation of other interesting stimuli (such as one or several objects), children are sensitive only to relatively large changes in the speech signal. For example, Stager and Werker (1997), using a habituation method, found that 14-month-olds could differentiate minimal pairs (such as 'bih' and 'dih') while looking at a still checkerboard pattern, but not while looking at a moving novel object display. Infants' failure to discriminate the minimal pairs may be attributed either to their attention-demanding fascination with the objects (and consequent disregard for subtle aspects of the speech signal) or, on the account favored by Stager and Werker, to their attempt to learn a mapping between the syllable and the object. Presumably this mapping demands processing resources that are therefore unavailable for precise phonetic analysis.

If the demands of matching sounds to objects during word learning preclude attention to detailed aspects of the sound patterns of words, young children will start with vague lexical representations. These representations might be refined through auditory or vocal experience, or they might be relatively immune to revision until children undergo a general change in the manner with which they represent speech sounds, such as the representational reorganization proposed by Werker (e.g. Werker & Tees, 1999). If these refinements do not occur until early childhood, this delay may account for the poor performance of 1.5-year-olds in Shvachkin-type tasks.

A third possibility is that the manual picture-selection task, using newly-taught words, underestimates children's representational capacities, and that in general children's representations of more familiar words are well-specified. As suggested by the studies by Barton (1976, 1978), children perform better when tested with words they say spontaneously than with words they have just been taught. In addition, even for familiar words, the manual object selection task may not be the most sensitive measure of lexical knowledge. Thus, children may have well-specified representations of how words sound, but assessments may underestimate the fidelity of those representations. This position is explicitly assumed in several theoretical discussions of phonological development (e.g. Bernhardt & Stemberger, 1998; Hale & Reiss, 1998; Smolensky, 1996). On this account, there would be substantial *continuity* in the forms by which infants encode familiar sound patterns and children encode words.

Our goal in the present study was to evaluate the phonetic detail present in the lexical representations that very young children use in word recognition. We employed a visual fixation procedure to assess word recognition (Swingley, & Fernald, 1998). Children's eye movements were monitored as they viewed pairs of pictures and heard sentences naming one of the pictures. Previous research has found that children tend to fixate named pictures promptly upon hearing their labels (e.g. Fernald, Pinto, Swingley, Weinberg & McRoberts, 1998).

We argue that fixation responses provide an index of the time-course of word recognition. As children hear the spoken target word on each trial (e.g. 'ball'), that sound pattern is continuously compared with the sound patterns of words in the child's lexicon. The words in the lexicon that match are activated, resulting in the retrieval of semantic information in the lexical entry (e.g. [round toy]).[1] If this semantic information is consistent with the currently fixated picture (Am I looking at a round toy?), the child continues to gaze at that picture. If it is inconsistent (No, I'm looking at a car) the child rapidly shifts her gaze.

This interpretation of the task is supported by a recent series of experiments by Swingley and Fernald (under review). In these studies, 24-month-olds saw pictures of two familiar objects and heard sentences containing (a) a word matching one of the objects (baseline trials), (b) a familiar word not matching either object (mismatch trials), or (c) a nonce word (nonce trials). Children fixating the distractor on baseline trials rapidly shifted to the target while hearing the target word. Children fixating either picture on mismatch trials produced exactly the same initial response – rapidly and reliably shifting away from the current (mismatching) picture, even though the alternative picture was just as inconsistent with the spoken word. (After this initial rapid response, children gazed randomly.) On nonce trials, children did not reliably shift – rather, they showed a diffuse array of responses, sometimes shifting and sometimes not. Swingley and Fernald interpreted these findings as showing that children's initial responses to the spoken words were based primarily on the picture the child was fixating when the target was heard, and that the rapid and robust shifts in gaze reflected the child's detection of a mismatch between the retrieved semantic category and the initially fixated picture.

In the present research, 18–23-month-old children participated in the visual fixation task. As discussed above, previous research leaves open the question of whether children this age use vaguely-specified representations in word recognition. To address this question, children were presented with correct pronunciations of familiar target words, as well as subtle mispronunciations of those words. If children's representations of words are vague or underspecified, they would not be expected to respond differently to small changes in pronunciation. Alternatively, if children's

---

[1] Our terminology here is taken from research on spoken word recognition in adults. The major focus of this research is to specify what counts as a 'match' and to characterize the relevant mental operations. For a fuller treatment than we can offer here, see, for example, Frauenfelder and Floccia (1998), or Marslen-Wilson (1993).

lexical representations are accurately specified, they should respond differently to correct and incorrect pronunciations. In particular, children hearing mispronounced words might be (a) less likely to fixate the target, and (b) slower to initiate a shift from the distractor to the target. Finally, possible relations between children's speech productions and their sensitivity to mispronunciations were evaluated by comparing children who varied in the extent of their spoken vocabularies and in their ability to say the tested words.

2. Methods

2.1. Overview

Participants were 56 18–23-month-olds. On each trial, two pictures of familiar objects were displayed on a large computer monitor. A few seconds later, a prerecorded pair of sentences was played. The first sentence named one of the displayed pictures. The realization of the target word was either 'correct' (i.e. described by a prototypical sequence of phonemes) or 'mispronounced' (described by an atypical sequence of phonemes). The target words tested were apple, baby, ball, car, dog, and kitty; the mispronounced versions of these words were opple (with the vowel of 'hop'), vaby, gall, cur, tog, and pity. Thus, two mispronunciations involved vowels, and four involved initial consonants. Consonant mispronunciations substituted sounds that are relatively likely to be confused by adults (e.g. Miller & Nicely, 1955). Children's visual fixations were recorded and coded off-line by coders who noted the timing of stimulus onsets and changes in children's fixations.

2.2. Subjects

The children ranged in age from 18,02 (months, days) to 23,02, with a mean of 20,08 and median of 19,16 (the 23,02 boy was the only 23-month-old in the sample, which was weighted toward 18–19-month-olds to help diversify the range of vocabulary sizes). Half were girls. Boys and girls were roughly age-matched (mean male age 20,11; mean female age 20,06). All children were full-term well-baby births, with American English as the primary language of the caregivers. An additional 37 children participated but were not included in the final data set. Of these, 27 began the study but did not complete at least 20 of the 24 test trials. Other subjects were excluded because the child refused to sit on the parent's lap (5), no Communicative Development Inventory (Fenson, Dale, Reznick, Bates, Thal & Pethick, 1994) was obtained (3), the parent peeked at the displays during test trials (1), or experimenter error (1).

2.3. Visual stimuli

The visual stimuli were digitized photographs of objects on a gray background, presented side by side on an 80 cm Sony PVM-3230 color video monitor. The pictures shown on test trials included an apple, a baby, a ball, a car, a cat, and a dog. Pictures were of similar sizes, averaging about 13 cm in length, and were separated by about 18 cm. Children were seated on their parent's lap about 80 cm from the screen, with the pictures at eye level. Four filler trials with new pictures were included to help maintain children's interest in the procedure. On filler trials, four pictures were displayed, each about the same distance from the center of the screen. The four filler pictures were of a toy duck, a robin, a toy dump truck, and a baby sneaker. On filler trials, each picture occupied each of four screen positions (left, right, top center, bottom center) once.

2.4. Auditory stimuli

The speech stimuli were digitally recorded by a female native speaker of American English in a soundproofed room using Sound Designer software, sampling at 22 050 Hz. Her speaking rate was slow and in a moderately 'infant-directed' register. The first sentence spoken on each test trial was 'Where's the [target]?'. The 'where's the' carrier phrase was about 765 ms long. The lengths of the target words in ms (including the stop gap of the initial stop consonants) were as follows: apple, 837; opple, 824; baby, 813; vaby, 821; ball, 640; gall, 688; car, 660; cur, 568; dog, 782; tog, 746; kitty, 791; and pity, 794. The second sentence, which began 750 ms after the offset of the first, was 'Can you find it?' for the targets apple/opple and baby/vaby, 'Do you like it?' for ball/gall and car/cur, and 'Do you see it?' for dog/tog and kitty/pity.

2.5. Apparatus and procedure

The experiment was conducted in a room containing a three-sided cloth-walled booth measuring 1 × 1.2 × 2 m tall. The parent sat on a chair in the open end of the booth, holding her child on her lap facing the monitor, which formed part of the back wall of the booth. Speech stimuli were delivered through a concealed central speaker beneath the monitor. The child's eye movements were monitored using a videocamera placed about 10 cm below the monitor.

Before coming to the laboratory, parents completed a consent form, and the MacArthur Communicative Development Inventory (Words and Sentences CDI; Fenson et al., 1994). (Parents were asked to add to the CDI a list of words said by their child but not included on the checklist; typically parents added about 10 words, mostly proper names.) When parent and child arrived at the lab, they were brought into the dimly-lit room containing the booth; the child was invited to play with some toys while the procedure was explained to the parent. Parents were instructed to try to keep their child on their lap and facing the monitor throughout the procedure. Parents were also asked to refrain from speaking, and to close their eyes and bend their head downward throughout the trials. (Parents not closing their eyes when the first trial began were instructed to do so; thus, parents were blind to target side.)

Two audiovisual displays preceded the experiment proper. First, about 1 min of the animated film 'Snow White' was shown in a 10-cm window at the center of the monitor. Then, an experimenter-controlled animation was initiated, in which a 3-cm picture of a beagle, goldfish, or duck moved around the screen, periodically stopping and making a brief noise at each of nine locations. These displays were used to gather calibration data for a semi-automated eye tracking system (sparse data from this system precluded their use in the present study). On occasion, children began fussing during the animations; when this happened the sequence was cut short and the experiment proceeded.

The experiment proper consisted of 28 trials, including 24 test trials and four filler trials. Each test trial began with the simultaneous presentation of two horizontally-aligned pictures. Three seconds later, the first of the two sentences began. The trial ended 6 s after the onset of the first sentence. Trials were separated by a 1-s pause, during which the monitor was black. On filler trials, four pictures were presented, but trial timing was the same. The entire procedure lasted about 5 min.

Four stimulus orders were created. The second order was a left/right reflection of the first; the third and fourth reversed the trial order of the first and second. Each picture served as the target four times (twice on the left and twice on the right) and the distractor four times (twice left, twice right). Pictures were yoked in pairs: the apple with the ball, the baby with the dog, and the car with the kitty. No picture appeared twice on consecutive trials. Each of the 12 target words was presented once in each half of the experiment, and each picture appeared an equal number of times on the left and right in each half of the experiment. Finally, excepting one mis-assignment, each order was presented to an equal number of boys and girls.

Following the experimental procedure, parents were asked whether their child understood or had ever said each of the target words. If a child was reported to have said a word, the experimenter asked the parent to imitate or otherwise characterize their child's pronunciation. The child was also encouraged to try to say the words, typically by showing each picture to the child and asking 'What's that?'.

2.6. 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. (Because the children were close to the screen, fixations to the left and right pictures were readily discernible.) Coding in each study was done by several highly trained coders who were unaware of the auditory stimulus 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. Filler trials were not coded. Reliability over the data set was established by having second coders re-code a randomly selected block of six test trials for each of 16 randomly selected subjects. Reliability was high (mean percent agreement 95.4; mean Cohen's kappa 93.2).

3. Results

Previous research using this visual fixation method within a comparable age range has shown that children tend to fixate the target picture shortly after the onset of the spoken target word (Fernald, Swingley & Pinto, under review; Swingley, Pinto & Fernald, 1999). As in previous research, we established a 'window' of time during which fixation responses were examined. This window began 367 ms after the onset of the target word. Because responses to the spoken word require the mobilization of an eye movement, they cannot be instantaneous; it is often assumed that the *minimum* latency to initiate an eye movement in infants is on the order of 233 ms, with mean latencies considerably longer (e.g. Canfield & Haith, 1991; Dougherty & Haith, 1997; Haith, Hazan & Goodman, 1988; see also Canfield, Smith, Brezsnyak & Snow, 1997; Hood & Atkinson, 1993). At present, research specifying mean eye movement latencies in 17–23-month-olds has not been done, particularly with regard to eye movements between two continuously shown images (most infant work cited above used a 'gap' task in which an initial fixation point is removed as the target fixation point is presented; 233 ms is probably an underestimate of infants' minimum latency under the more complex visual conditions used in our study). The lower bound of 367 ms is an 'educated guess' based on studies such as those cited above, and our data (here and in other studies) showing that target and distractor fixations tend to diverge at around 400 ms. In any event, the results reported here are the same when other minima (such as 200 and 400 ms) are used. Thus, removal of responses faster than 367 ms excluded some eye movements too quick to be plausibly considered responses to the spoken target word. Similar criteria are used in research using fixations to study word recognition in adults (e.g. Dahan, Swingley, Tanenhaus & Magnuson, 2000). The window of analysis ended 2 s after the onset of the target word. Previous research has suggested that the few eye movements occurring after this time are usually spontaneous re-fixations unrelated to the spoken stimulus.

Within this analysis window, we report two measures: children's proportion of fixation to the target (*accuracy*), and children's *response latency* to initiate a shift from the distractor to the target (on those trials where children were looking at the distractor at the onset of the spoken target word). Proportional measures provided a means of evaluating the extent to which children linked the spoken target word and the target picture (Reznick, 1990).[2] Target fixation proportions were calculated by dividing the time children fixated the target by the sum of time they spent fixating the target and the distractor. The second measure, response latency, has been used in other research on infants to assess efficiency in word recognition (e.g. Fernald et al., 1998) and, of course, in numerous studies of linguistic processing in adults (e.g. Cattell, 1886).

The first question to be addressed was whether children behaved differently when hearing correct and incorrect pronunciations. If children's representations of these words were vaguely specified, they should have recognized words equally well in the two conditions; if children's representations were well-specified, their accuracy

---

[2] One alternative is to compare raw target and distractor fixation times (e.g. Golinkoff, Hirsh-Pasek, Cauley & Gordon, 1987). If children are consistent in the amount of time spent 'off task' on each trial, raw-duration and proportional measures provide essentially the same results.

and speed should have been inferior on mispronunciation (hereafter 'MP') trials. The latter result was obtained. Children's mean proportion of fixation to the target picture in the correct-pronunciation (CP) condition was 73.0%, whereas their target fixation proportion was only 61.3% in the MP condition ($t(55) = 7.48$, $P < 0.0005$; all reported t-tests are two-tailed unless noted otherwise). Accuracy exceeded 50% in both conditions (CP, $t(55) = 16.5$, $P < 0.0005$; MP, $t(55) = 9.0$, $P < 0.0005$), indicating that children recognized the targets in both conditions.

A two-way ANOVA with sex and condition as factors yielded a significant main effect of condition ($F(1, 54) = 54.9$, $P < 0.0005$) but no effect of sex and no interaction. The mispronunciation effect for looking proportions was significant by items ($F(1, 5) = 12.3$, $P = 0.017$) and held for all six pairs.

Analyses of response latencies yielded similar results. At the onset of the spoken target word, children were fixating the distractor picture on about 47% of the trials (across conditions). Of these trials, children shifted to the target within the window of analysis 64% of the time. (Incorrect shifts, from the target to the distractor, occurred only 40% of the time.) Thus, a response latency to initiate a shift from the distractor to the target was measured on about 30% of all test trials (i.e. 0.47×0.64). Two children (one 18-month-old, one 22-month-old) did not produce an RT in both conditions and were excluded from these analyses. For the remaining 54 children, the mean CP latency was 718 ms, while the mean MP latency was 850 ms ($t(53) = 3.48$, $P = 0.001$). An ANOVA revealed no significant effects or interactions involving sex. The mispronunciation effect for RT was significant by items ($F(1, 5) = 15.7$, $P = 0.011$) and held for all six pairs.

Further analyses evaluated in more detail the possible role of vocal production in refining children's mental representations of words. Vocabulary sizes (as estimated by parental report on the MacArthur CDI; Fenson et al., 1994) ranged from 0 to 584 words (median 191.5). Vocabulary size was strongly correlated with age ($r = 0.590$, $P < 0.0005$), and age was correlated with accuracy on both MP trials ($r = 0.574$, $P < 0.0005$) and on CP trials ($r = 0.325$, $P < 0.02$). However, vocabulary size was only correlated with accuracy on MP trials ($r = 0.421$, $P < 0.005$), and not on CP trials ($r = 0.105$, *ns*). Thus, older children performed better than younger children in both conditions, but vocabulary size (a correlate of age) was only predictive of better performance on MP trials. Response latencies on CP and MP trials were negatively correlated with age and vocabulary size, but none of these correlations was significant (age: r for CP condition, -0.209; MP condition, -0.127; vocabulary: CP, -0.162; MP, -0.060; all $P > 0.10$).

Children with larger vocabularies did not show greater effects of mispronunciation than children with smaller vocabularies. The size of the mispronunciation effect ($\mathrm{CP} - \mathrm{MP}$) on accuracy was *negatively* correlated with vocabulary size, though this correlation was marginal ($r = -0.239$, $P = 0.076$). The mispronunciation effect on accuracy was not significantly correlated with age ($r = -0.165$, $P > 0.20$). The mispronunciation effect on RT was not significantly correlated with age or vocabulary size (both $r < 0.10$, $P > 0.40$). Regression analyses on condition differences for accuracy and for latency failed to find any variance significantly accounted for by age or vocabulary size.

Most of the children in the sample were more accurate, and faster, when responding to CP targets. Of the 56 children, 47 had higher %-to-target (accuracy) scores on CP trials than on MP trials. If all children are separated into four equal-sized groups of 14 subjects according to age, the number of children showing the effect is nearly equal across groups (12, 12, 12, and 11). If all children are separated into four equal groups according to vocabulary size, a similar result is found (0–39 words, 11 children; 49–178, 14; 205–291, 11; 324–584, 11). In each case the result is significant by sign test (11 or more of 14, $P = 0.029$). Fig. 1 displays the accuracy scores for each age quartile, split by condition. Overall performance improved with age, but the size of the mispronunciation effect was as great in the younger children as in the older children.

Turning to the response latency analyses, 35 children (of the 54 contributing RTs in both conditions) showed faster responses on CP trials. Grouping by age quartile revealed no differences in the numbers of children showing the effect (8, 7, 10, 10; $\chi^2 = 0.77$, *ns*), nor did grouping by vocabulary-size quartile (8, 8, 10, 9; $\chi^2 = 0.31$, *ns*).

Even if vocabulary size did not provide special leverage in predicting performance, aspects of children's pronunciations of the tested words might be expected to be related to their differentiation of good and bad pronunciations in perception, operationalized again by difference scores (CP accuracy − MP accuracy). Three factors were considered: whether children *said* the target or not, whether children saying the target said it *correctly* or not, and finally, whether children saying the target said the *onset* correctly or not (for the item car/cur, the last analysis concerned the vowel rather than the onset). In these three within-subjects analyses, a given child could only contribute to an analysis if he or she had words in both of the analyzed categories; for example, 25 children said either zero or six of the target words, and were therefore excluded from the say/no-say analysis. Results showed no significant effects of any of the three factors (saying versus not saying targets, $t(30) = 0.095$; good versus bad child pronunciations, $t(38) = 0.826$; good versus bad onsets, $t(17) = -0.089$; all $P > 0.40$). Inspection of the mean accuracy scores

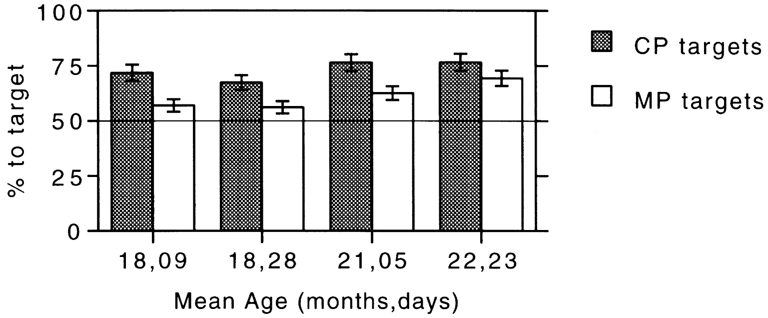

Fig. 1. Children's proportions of fixation to the target picture for correctly pronounced targets (CP) and mispronounced targets (MP). Subjects are grouped by age quartile, with the mean age in each group shown. Error bars are standard errors.

for each ‘child pronunciation’ type yielded no indication that children’s ability to say the words was related to their tendency to fixate the target picture. Thus, the results provided no support for the notions that spoken vocabulary size, or practice in saying the tested words, were related to the size of the mispronunciation effect.

The difference in children’s responses to the CP and MP trials is evident in the time-course of their eye movements (Fig. 2). Time is represented on the *x*-axis, with zero corresponding to the onset of the target word. Trials have been separated according to condition (CP, MP) and according to where children happened to be looking when the target word began (target, distractor), yielding four groups of trials corresponding to the four lines on the graph (the graph excludes the 13% of trials on which children were initially fixating neither the target nor the distractor). CP trials are shown with filled symbols, MP trials with empty symbols. Trials on which children initially fixated the distractor are represented by squares, and trials on which children started at the target are represented by circles. The *y*-axis indicates, for each 33 ms, the proportion of trials on which children at that time were no longer fixating the picture that they had been fixating at the onset of the target word. Thus,

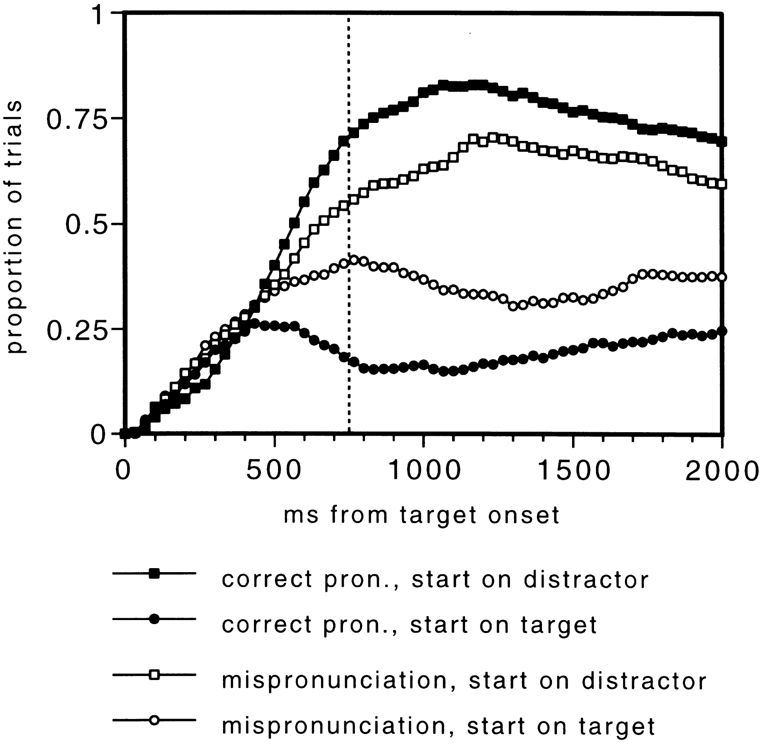

Fig. 2. Children’s responses over time, separated by condition and by whether children fixated the target or the distractor picture at the onset of the spoken target word. The ordinate indicates the proportion of trials on which children were (at that moment) fixating a different picture than the one they fixated at the onset of the target word. The dashed vertical line indicates the average offset of the target word.

perfect performance would be shown by circles staying at zero, and squares rising to one. For example, the filled squares concern CP trials on which children were initially fixating the distractor. After 600 ms, children on over half of these trials were then looking at the target. The filled circles concern CP trials on which children initially fixated the target. After 600 ms, children were looking at the distractor on only about 25% of the trials. The vertical dashed line shows the mean time at which the target word ended.

The time-course plot shows that the effects of the mispronunciation did not depend on where children were looking at the onset of the target word. Furthermore, the difference between responses on CP and MP trials emerged quite early; it was not the case that children heard the MP words and ‘retroactively’ reacted to the difference; rather, it seems that the MP words were less effective at driving responses than CP words were, right from the start. This interpretation may be justified statistically by finding the first 33 ms time-slice in which CP accuracy (calculated over subjects) exceeded MP accuracy. For trials on which children initially fixated the distractor, this time was 533 ms (mean difference 9.8%, $t(55) = 2.0$, $P < 0.025$, one-tailed). For trials on which children initially fixated the target, this time was 567 ms (mean difference 8.0%, $t(55) = 1.8$, $P < 0.05$, one-tailed). Considering all trials regardless of children’s initial fixations, CP accuracy first significantly exceeded MP accuracy at 467 ms (mean difference 6.7%, $t(55) = 2.1$, $P < 0.02$, one-tailed).[3] Considering that it generally takes at least 200 ms to program an eye movement, these results suggest that the effects of mispronunciation on word recognition are nearly immediate, as would be predicted by continuous-activation models of word recognition.

4. Discussion

The nature of early lexical representations has been a matter of some dispute. Early studies suggested that 1-year-olds cannot reliably distinguish words differing by a single sound (e.g. Shvachkin, 1948/1973), whereas later studies showed that infants under 12 months of age distinguish familiar and unfamiliar minimal-pair words (e.g. Jusczyk & Aslin, 1995). The present research used a task with modest behavioral demands, but which was nevertheless referential, and found that in fact young children (from 18 to 23 months) represent familiar words with sufficient detail to be hindered in recognition when words are slightly mispronounced. Furthermore, we found no evidence that children’s differentiation of correct and deviant pronunciations was related to their vocabulary size or to their ability to say the tested words.

The present results are consistent with research evaluating the lexical representations of older children. For example, Barton (1976, 1978) found that 2-year-olds had little difficulty differentiating minimal pairs in familiar words. The present results can also be accommodated with previous findings within the 18–23 month age range

---

[3] For each of these times, all subsequent CP versus MP comparisons were significant at alpha levels of 0.01 or less; thus, this analysis did not capitalize on multiple t-tests to produce Type I errors.

by considering differences in the tasks used. Previous work using object selection tasks and newly-taught words found inconsistent performance on many phonemic contrasts, even for contrasts not especially confuseable by adults (e.g. Garnica, 1973). This may have been due in part to the task, which required that the child manually indicate one of two objects. Though this is not a strenuous task, it does require the performance of an overt movement indicating a choice. By contrast, eye movements are relatively automatic, and under appropriate conditions may reflect cognitive processes that are masked when children must make an overt choice (for examples in other domains see Clements & Perner, 1994; Spelke, 1991).

Also, as previously stated, children's representations of familiar words may well be better-specified than their representations of newly-learned words. It is possible that the performance seen in experiments like Shvachkin's (or like Stager and Werker's) reflects the nature of words to which children's exposure has been relatively limited, but does not reflect children's representations of more familiar words (Barton, 1978). We do not have age-of-acquisition data on the words tested here, but it seems likely that most of the words had been known by most of the children for several months. We suspect that newly-learned words will not always be represented in such detail.

The present results may be described in terms of 'activation', which has become a generally accepted metaphor in the psycholinguistic literature on word recognition. As discussed in Section 1, the activation of a word is, in part, a function of the degree to which that word's mental representation matches the heard speech, according to a similarity metric that is not well understood. Activation is taken to be continuous; good tokens of a word will result in more, or swifter, activation of that word than poor tokens. Sufficient activation results in access to the word's meaning, as evidenced by, for example, facilitation in semantic priming tasks. At present it is not clear whether the activation of meaning is a probabilistic all-or-none process (in which poor tokens are less likely to activate word meaning, or are slower to do so), or if the strength of semantic activation is a continuous function of the match between the token and the lexical representation (poor tokens activate word meaning to a lesser degree). The available research on adults' word recognition does not differentiate these accounts; however, it is known that mispronounced words may result in less semantic facilitation on average, across trials, than correctly pronounced words (Milberg, Blumstein & Dworetsky, 1988), and that sufficiently close competitors of heard words are semantically activated to some degree (e.g. Allopenna, Magnuson & Tanenhaus, 1998; Connine, Blasko & Titone, 1993).

We argue that similar processes are responsible for the mispronunciation effects seen in the current study. The fact that children performed above chance on MP trials, yet significantly worse than they performed on CP trials, suggests that mispronunciations activate the semantics of words less frequently or less strongly than correct pronunciations, just as in adults. Thus, children hearing 'vaby' were less likely to activate the notion of [baby], or activated it less strongly, than children hearing 'baby', resulting in suboptimal but non-chance performance on MP trials.

No claim is made here about children's conscious decisions. Mispronunciations were relatively ineffective in driving fixation responses, but this does not imply that children were aware of the mispronunciations, or that children wondered whether the mispronounced words were, in fact, novel words. Children's responses are not best viewed as the outcome of a lengthy deliberation in which children compared the fit of the MP word against both pictures and eventually chose the best match; recall that MP and CP responses began to diverge about 500 ms from the *onset* of the targets – a figure that includes *at least* 233 ms for the programming of the eye movement itself. The task measures word recognition, not mispronunciation detection per se. Consequently, the results presented here do not permit us to specify the conditions under which children would consider a word like 'vaby' to be a mispronunciation of 'baby', or a new word that is a candidate for learning and entry into the lexicon.

However, the results do suggest that young children encode familiar words in detail. This finding contrasts with two of the hypotheses described in Section 1: first, that children's word recognition systems are insensitive to much of the lexically relevant phonetic information in words; and second, that the attentional demands of word learning preclude sensitivity to phonetic detail. Both hypotheses account for young children's failure to discriminate newly-taught minimal pairs by appealing to a lack of specificity in children's representations of the words' phonetic forms. Given our finding that children were sensitive to subtle mispronunciations in a referential task, we suggest that previous experiments with young children, such as Shvachkin's, underestimated children's knowledge. Whether this was due to the novelty of the words typically tested in object-selection procedures, or the difficulty of the task itself, remains an open question.

It has been suggested that early in language development, young children have 'holistic' representations of words, which are adequate for efficient word recognition only until the lexicon becomes crowded with similar-sounding words, or 'neighbors'; then, segmental representations are required (Charles-Luce & Luce, 1990, 1995; Jusczyk, 1986; Walley, 1993). For example, Charles-Luce and Luce (1990, p. 207) write, "We suggest that, as the size of the lexicon increases, the child must begin to organize the acoustic-phonetic information into segment-size units in order to identify a word uniquely."

While we agree that increases in vocabulary size would render *inaccurate* or *vague* phonetic specifications inadequate, it is not clear that non-phonemic representations need be inaccurate or vague. It may be that even in adults word recognition is not a process that obligatorily involves the translation of the speech signal into phonemes. This is a controversial issue, and we will not rehearse the debate here (see, for example, Hawkins, 1995; Klatt, 1989; Pisoni & Luce, 1987). However, without more conclusive evidence that phonemic specification is necessary to account for word recognition, it would be premature to consider the present results as evidence for (or against) phonemic specification in 1-year-olds' lexical representations. We may say only that our results contradict any view holding both that (a) segmental representations are criterial for the differentiation of minimal pairs, and that (b) children of 18–23 months do not represent speech in terms of segments.

Previous research found that 24-month-olds did not have more trouble differentiating 'ball' and 'doll' from each other, than from the more phonetically distinct words ‘duck’ and ‘truck’ (Swingley et al., 1999). Swingley et al. noted that this could have been a consequence of children having learned the two similar words, which may have drawn their attention to the subtle differences between them. This mechanism cannot account for the present results, because the MP words were either invented words or words that are surely unknown to the vast majority of children. Furthermore, children are unlikely to know words whose onsets match those of all of the mispronunciations. For example, words starting with ‘vay’ are extremely rare even in the speech of first-graders (Moe, Hopkins & Rush, 1982). The Toddler CDI, which contains many of the words 1-year-olds are likely to know, includes no entries starting with ‘vay’, ‘cur’, or ‘gaw’ (with the possible exception of ‘all-*gone*’). Although the other three MP words overlapped at onset with words known by many children (e.g. ‘tog’ and ‘talk’; ‘pity’ and ‘pig’ or ‘picture’; ‘opple’ and perhaps ‘on’), the results for these words were not distinguishable from the results for the other words. Thus, while children’s representations of ‘ball’ and ‘doll’ may be refined by the need to tell those words apart, the children tested here probably did not know any ‘vay…’ words to play this role in reinforcing the ‘vaby/baby’ contrast. Nevertheless, children treated ‘baby’ as a better token of the word than ‘vaby’. This suggests that the refinement of lexical representations does not require a *direct* competitor of the ball/doll sort.

However, the effects of lexical neighbors may not be limited to delineating specific contrasts in this way. For example, it is possible that knowledge of a word like ‘maybe’ serves to refine the representation of ‘baby’, not only by highlighting the [m]/[b] contrast, but also by focusing children’s attention on the initial [b] of ‘baby’, thereby encouraging elaboration of that sound. Such a process could result in a well-specified representation of [b], leading to reduced activation given mispronunciations like ‘vaby’.

This hypothesis would be supported by evidence that expansion in the tested children’s vocabularies was correlated with the degree to which children differentiated CP and MP words. Our failure to detect any effects of vocabulary size, despite a range extending from zero to several hundred words, suggests that expansion of the *spoken* vocabulary is not the driving force behind the accurate phonetic specification of words. To the extent that there were any effects of vocabulary size, they were in the wrong direction: children with small vocabularies showed slightly (and non-significantly) greater differentiation of the correct and deviant pronunciations than children with large vocabularies.

Of course, the proposal that ‘crowding’ in phonetic space forces detailed specification of neighboring word-forms hinges on perception, not production; it is the child’s need to differentiate the words she hears that leads to closer attention to contrasting phonetic information. Because we have data only on the productive lexicon and not the receptive lexicon, we cannot rule out the possibility that even the children saying very few words still knew neighbors of the target words, and that these neighbors forced the detailed specification of the targets. In sum, although the current study offered the potential for neighborhood effects on lexical specification to be demonstrated, our findings of accurate specification in children with tiny spoken vocabularies are not decisive evidence against the operation of such neighborhood effects. It is clear, however, that practice in articulating a word correctly is not a prerequisite for accurate specification as operationalized here; no differences in recognition performance emerged on the basis of whether or not children said the targets correctly, or said them at all.

In the present experiment, four of the six mispronunciations involved onset consonants and two involved vowels; no coda consonants were tested. This was partly a matter of practicality. Given that young children interpret speech incrementally, as adults do (Fernald et al., under review; Swingley et al., 1999), a non-initial mispronunciation might be detected only after the target word had been activated. For example, children hearing 'Where's the dawp?' might begin to fixate a picture of a dog even before the mismatching /p/ was heard. For this reason, null effects would be difficult to interpret.

However, apart from methodological considerations, this question bears further investigation. Syllable onsets may be more precisely specified than syllable offsets in children, simply because onsets are often more perceptually distinctive. For example, adults are more accurate in identifying initial Cs of CVC syllables than in identifying final Cs (e.g. Redford & Diehl, 1999; see also Malécot, 1958, for data bearing on this point). Speakers of English frequently fail to release prepausal stops, resulting in reduced identifiability (Lisker, 1999), and assimilation tends to obscure the first, rather than the second, C of VCCV sequences (e.g. Ohala, 1990).[4] Although none of this work has been done with children or with child-directed speech, there is reason to suppose that children might have less well-specified representations of coda consonants than of onset consonants – not because of any particular developmental incapacity, but as a consequence of differential identifiability in the speech signal. It is also possible that children's representations of the later parts of words are less well-specified for reasons particular to children. For example, very young children may attend to word or syllable onsets to a greater degree than older children. In a recent set of experiments, Jusczyk, Goodman and Baumann (1999) found that 9-month-olds detected repetition in the onset C or CV of lists of CVC sequences, but not the rime (VC), suggesting that infants may pay special attention to syllable onsets. Whether this tendency holds with respect to children's early words remains to be sorted out experimentally.

In the present study, children's recognition of familiar words was impaired, but not prevented, when those words were mispronounced. This general pattern of results has also been found in adults (e.g. Connine et al., 1993). Our results are compatible with models of word recognition in which lexical activation is a continuous function of the degree to which heard words match stored lexical representations. By 18 months, the lexical representations used for recognition appear to be specified in fine detail, even among children with small spoken vocabularies.

---

[4] Of course, other phonetic processes result in variability in onsets, but in English, coarticulation of adjacent phonemes is usually anticipatory (e.g. Gimson & Cruttenden, 1994, p. 255).

Acknowledgements

This work was supported by grants from the National Institutes of Child Health and Human Development (NIH HD-37082) and the National Science Foundation (SBR-9421064) to R.N.A., and the National Institutes of Health (F32-HD08307) to D.S. Portions of the research reported here were presented at the 1999 Biennial Meeting of the SRCD. The authors would like to extend their thanks to Elizabeth Gramzow and to the students at the Infant Perception Lab at Rochester for their invaluable assistance. We would also like to thank the many parents and children who made this work possible through their participation.

References

Allopenna, 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 D. Kuhn, & R. Siegler (Eds.), *Handbook of child psychology: cognition, perception, and language.* (Vol. 2, pp. 147–254). New York: Wiley.

Barton, D. (1976). Phonemic discrimination and the knowledge of words. *Papers and Reports on Child Language Development*, *11*, 61–68.

Barton, D. (1978). The discrimination of minimally-different pairs of real words by children aged 2;3 to 2;11. In N. Waterson, & C. Snow (Eds.), *The development of communication* (pp. 255–261). Chichester: Wiley.

Bernhardt, B. H., & Stemberger, J. P. (1998). *Handbook of phonological development: from the perspective of constraint-based nonlinear phonology*. San Diego, CA: Academic Press.

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., Brezsnyak, M. P., & Snow, K. L. (1997). Information processing through the first year of life. *Monographs of the Society for Research in Child Development*, 62(2, Serial No. 250).

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *11*, 63–65.

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, *17*, 205–215.

Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, *22*, 727–735.

Clements, W. A., & Perner, J. (1994). Implicit understanding of false belief. *Cognitive Development*, *9*, 377–395.

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.

Dahan, D., Swingley, D., Tanenhaus, M. K., & Magnuson, J. S. (2000). Linguistic gender and spoken word recognition in French. *Journal of Memory and Language*, *42*, 465–480.

Dodd, B. (1975). Children's understanding of their own phonological forms. *Quarterly Journal of Experimental Psychology*, *27*, 165–172.

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(5, Serial No. 242).

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.

Fernald, A., Swingley, D., & Pinto, J. P. (under review). When half a word is enough: infants can recognize spoken words using partial phonetic information.

Frauenfelder, U. H., & Floccia, C. (1998). The recognition of spoken word. In A. D. Friederici (Ed.), *Language comprehension* (pp. 1–40). Berlin: Springer.

Garnica, O. K. (1973). The development of phonemic speech perception. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 215–222). New York: Academic Press.

Gimson, A. C., & Cruttenden, A. (1994). *Gimson's pronunciation of English* (5th ed.). London: Arnold.

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.

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.

Hale, M., & Reiss, C. (1998). Formal and empirical arguments concerning phonological acquisition. *Linguistic Inquiry*, *29*, 656–683.

Hawkins, S. (1995). Arguments for a nonsegmental view of speech perception. In K. Elenius, & P. Branderud (Eds.), *Proceedings of the XIIIth International Congress of Phonetic Sciences* (Vol. 3, pp. 18–25). Stockholm: Stockholm University.

Hood, B. M., & Atkinson, J. (1993). Disengaging visual attention in the infant and adult. *Infant Behavior and Development*, *16*, 405–422.

Jusczyk, P. W. (1986). Towards a model for the development of speech perception. In J. Perkell, & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 1–19). Hillsdale, NJ: Erlbaum.

Jusczyk, P. W. (1993). Some reflections on developmental changes in speech perception and production. *Journal of Phonetics*, *21*, 109–116.

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., Goodman, M. B., & Baumann, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language*, *40*, 62–82.

Kiparsky, P., & Menn, L. (1977). On the acquisition of phonology. In J. Macnamara (Ed.), *Language learning and thought* (pp. 47–78). New York: Academic Press.

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

Lisker, L. (1999). Perceiving final voiceless stops without release: effects of preceding monophthongs versus nonmonophthongs. *Phonetica*, *56*, 44–55.

Locke, J. L. (1988). The sound shape of early lexical representations. In M. D. Smith, & J. L. Locke (Eds.), *The emergent lexicon* (pp. 3–22). San Diego, CA: Academic Press.

Malécot, A. (1958). The role of releases in the identification of released final stops. *Language*, *34*, 370–380.

Marslen-Wilson, W. (1993). Issues of process and representation in lexical access. In G. T. M. Altmann, & R. Shillcock (Eds.), *Cognitive models of speech processing: the second Sperlonga meeting* (pp. 187–211). Hillsdale, NJ: Erlbaum.

Milberg, W., Blumstein, S., & Dworetzky, B. (1988). Phonological factors in lexical access: evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society*, *26*, 305–308.

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.

Moe, A. J., Hopkins, C. J., & Rush, R. T. (1982). *The vocabulary of first-grade children*. Springfield, IL: Thomas.

Ohala, J. J. (1990). The phonetics and phonology of aspects of assimilation. In J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology I: between the grammar and physics of speech* (pp. 258–275). Cambridge: Cambridge University Press.

Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition*, *25*, 21–52.

Redford, M. A., & Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *Journal of the Acoustical Society of America*, *106*, 1555–1565.

Reznick, J. S. (1990). Visual preference as a test of infant word comprehension. *Applied Psycholinguistics*, *11*, 145–166.

Shattuck-Hufnagel, S. (1986). Comment: why we need more data. In J. Perkell, & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 77–84). Hillsdale, NJ: Erlbaum.

Shvachkin, N. K. (1973). The development of phonemic speech perception in early childhood. In C. A. Ferguson, & D. I. Slobin (Eds.), *Studies of child language development* (pp. 91–127). New York: Holt, Rinehart, and Winston. (Original work published 1948).

Smith, N. V. (1973). *The acquisition of phonology: a case study*. Cambridge: Cambridge University Press.

Smith, N. V. (1978). Lexical representation and the acquisition of phonology. *Studies in the Linguistic Sciences*, *8*, 259–273.

Smolensky, P. (1996). On the comprehension/production dilemma in child language. *Linguistic Inquiry*, *27*, 720–731.

Spelke, E. S. (1991). Physical knowledge in infancy: reflections on Piaget's theory. In S. Carey, & R. Gelman (Eds.), *The epigenesis of mind: essays in biology and cognition* (pp. 133–169). Hillsdale, NJ: Erlbaum.

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. (under review). Visual context and word recognition at 24 months.

Swingley, D., Pinto, J. P., & Fernald, A. (1998). Assessing the speed and accuracy of word recognition in infants. In C. Rovee-Collier, L. P. Lipsitt, & H. Hayne (Eds.), *Advances in infancy research* (Vol. 12, pp. 257–277). Stamford, CT: Ablex.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*, 73–108.

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review*, *13*, 286–350.

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: toward a new synthesis. *Annual Review of Psychology*, *50*, 509–535.