Phonetic Detail in the Developing Lexicon*

Daniel Swingley

Max-Planck-Institute for Psycholinguistics

Key words

fixation

language acquisition

lexicon

phoneme

word recognition

Abstract

Although infants show remarkable sensitivity to linguistically relevant phonetic variation in speech, young children sometimes appear not to make use of this sensitivity. Here, children's knowledge of the sound-forms of familiar words was assessed using a visual fixation task. Dutch 19-month-olds were shown pairs of pictures and heard correct pronunciations and mispronunciations of familiar words naming one of the pictures. Mispronunciations were word-initial in Experiment 1 and word-medial in Experiment 2, and in both experiments involved substituting one segment with [d] (a common sound in Dutch) or [g] (a rare sound). In both experiments, word recognition performance was better for correct pronunciations than for mispronunciations involving either substituted consonant. These effects did not depend upon children's knowledge of lexical or nonlexical phonological neighbors of the tested words. The results indicate the encoding of phonetic detail in words at 19 months.

1 Introduction

Over the course of their first year infants learn important aspects of the phonetic structure of their language. For example, infants gradually come to conflate phonetic categories that are not phonologically distinctive in the ambient language, while distinguishing phonetic categories that are phonologically distinctive. This phenomenon was documented by Werker and Tees (1984), who discovered that English-learning infants of 6–8 months discriminated syllables starting with the Hindi sounds [ṭ] and [t̪], whereas English-learning infants of 10–12 months did not; likewise, 6–8-month-olds discriminated Nthlakampx glottalized velar and uvular consonants (neither of which are phonemic in English) whereas 10–12-month-olds did not. The Hindi- or

* *Acknowledgements:* The author would like to thank Manon van Laer and the rest of the MPI babylab staff for their help in this project. We all would like to express our appreciation to the many Nijmegen parents and children who made this work possible through their participation. The author thanks Todd Bailey, Joe Pater, and Sharon Peperkamp for helpful comments and discussion. This research was supported by the NWO Spinoza grant "Native and non-native listening" to A. Cutler.

Address for correspondence: Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia PA, 19104-6196, U.S.A. Phone: 1-215-898-7300; fax 1-215-898-7301; e-mail: <swingley@psych.upenn.edu>.

Nthlakampx-learning infants tested maintained their language's contrast at 11–12 months of age. Similar results demonstrating language-specific tuning have been found for vowels (e.g., Cheour, Ceponiene, Lehtokoski, Luuk, Allik, Alho, & Näätänen, 1998; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Polka & Werker, 1994). Although declines in sensitivity to non-native contrasts are not always found (e.g., Polka, Colantonio, & Sundara, 2001), the positive results reported to date indicate that information in the sound structure of the speech infants hear is sufficient to bias infants' perception in ways that would seem to be advantageous for further progress in learning the language.

Accounts of these language-specific biases describe perception as an active process in which the continuous speech signal is *analyzed* into categories given by the ambient language. Best, in her Perceptual Assimilation Model, argues that non-native speech sounds that are similar to native speech sounds will be "assimilated" to the native categories (Best, 1994), while Kuhl's Native Language Magnet theory holds that a prototypical realization of a vowel "perceptually pulls other members of the category toward it" (Kuhl, 1995, p. 133). This process is automatic, in the sense that listeners generally do not *try* to fit speech sounds into categories; it just happens. In fact, adult second-language learners' best efforts to interpret non-native sounds outside the constraints of the native phonology often fail (e.g., Flege, 1995). This would seem to indicate that from an early age, perception of words in one's native language involves a transformation of the continuous acoustic signal into a sequence of segments whose inventory and phonetic characteristics are language-specific.[1] If so, young children's lexical representations should be detailed enough to perceptually distinguish words of the language whose segmental specifications differ.

However, infants' impressive performance in discrimination experiments does not necessarily imply that lexical representations are stored according to the phonological characteristics of the language. First, most studies demonstrating language-specific phonological tuning have used stimuli that stripped away much of the natural variability of speech. The fact that infants compute phonetic categories based upon exposure to the utterances that constitute their linguistic environment indicates that category formation (and presumably category identification) is not limited to laboratory conditions, but does not imply that this categorization is robust or that it aligns with native adult categorization in all contexts. For example, it is conceivable that the category formation revealed in experiments like Kuhl's is the result of infants'

[1] Here we sidestep the issue of whether it is appropriate to describe adults' speech interpretation in terms of segments, features, or gestures, and refer to segments only, though we take this debate as unresolved (for varying views, see e.g., Fowler, 1986; Hawkins, 1995; Lahiri & Jongman, 1990; Nearey, 1990; Pisoni & Luce, 1987; Stevens, 2002; Warren & Marslen-Wilson, 1987, 1988). For present purposes the relevant question is whether infants perceive and encode the speech signal at a level of detail sufficient for maintaining the language's phonological contrasts, whatever the representational scheme. We also assume that children encode in lexical representations information "above" the level of the segment, such as syllables and feet, but will not discuss whether this encoding faithfully matches the speech children hear.

computations over only a subset of the tokens infants have heard — for example, just those at utterance ends, or in strong syllables, or when infants happened to be paying particular attention. If so, it is not safe to assume that infants analyze all speech into segmental categories.

Second, infants' discrimination performance shows that they remember phonetic categories, but this does not mean that they remember language-specific phonetic detail in words. Translating an acoustic sequence into segments does not entail retaining those sequences in memory as a lexical representation. The present evidence on this question as it pertains to children under 12 months is mixed. Jusczyk and Aslin (1995) found that 7.5-month-olds familiarized with several isolated tokens of a word like *cup* then recognized it in sentences, but infants familiarized to a minimal variant like *tup* did not "false alarm" to *cup* in the same sentences, suggesting that hearing a word several times in isolation suffices for infants to encode the initial segment. However, Hallé and de Boysson-Bardies (1996) tested French 11-month-olds' spontaneous preferences for presumably familiar words like *bonjour* over unfamiliar words like *caduc*, and found preferences not only for *bonjour* but also *ponjour* (voicing change) and *vonjour* (place and manner change) relative to the unfamiliar words. These authors concluded that in contrast to meaningless syllables, actual words are encoded with "global representations" that do not preserve phonetic detail.

The conclusion that infants do not encode words with sufficient detail to distinguish minimal pairs comports with some studies of one-year-olds' word learning. For example, Shvachkin (1973) taught children nonce names for objects (such as *dak* and *gak*) and then tested children's ability to distinguish the words by presenting the objects and requesting one using the taught name. Children ranging from 10 to 24 months of age failed to consistently select the correct objects when the words differed by any of several consonant contrasts that were distinctive in the children's language environment. Replications using similar word-teaching and object choice methods produced similar results, with children succeeding on some contrasts and failing on others (Edwards, 1974; Eilers & Oller, 1976; Garnica, 1973), though two-year-olds generally succeeded when pairs of familiar words were tested (Barton, 1976). Brown and Matthews (1997) used a point-to-picture task with children ranging from 15 to 28 months of age and found considerable variability in performance over a range of minimal-pair contrasts. The words tested were real English words, though it is not clear how many words were known to children before the study began and how many were learned only via a training session that took place the day before the study proper.

In a recent series of studies, Werker and her colleagues have used a habituation technique to determine the degree of detail very young children encode in novel words. In one study (Stager & Werker, 1997), 14-month-olds were shown a film of a moving novel toy which was labeled repeatedly using varied tokens of a single nonce word (e.g., *lif … lif … lif …*). Once infants habituated to this audiovisual stimulus, a new pairing was given (such as the same object with *neem … neem … neem …*). When the words sounded very different, as in *lif* and *neem*, the switch to the new word resulted in dishabituation; when the words sounded similar, as in *bih* and *dih*, 14-month-olds did not dishabituate, suggesting that they had failed to register the sound change. Similar results were obtained when the habituation phase involved the pairing of two words with two objects. In recent experiments using the same method, 17-month-old children dishabituated to a switch in labels like *bih* and *dih*, suggesting a change in infants' abilities between 14 and 17 months (Werker, Fennell, Corcoran, & Stager, 2002).

This body of results has been taken as support for the notion that young children's knowledge of the sound-forms of words lacks either phonetic detail or the segmental structure that is held to be necessary for robust lexical differentiation (e.g., Charles-Luce & Luce, 1990). Discussions along these lines sometimes portray very young children's lexical representations as holistic patterns of sound in which only a few salient phonetic features are present, following child phonologists such as Waterson (1971) and Ferguson and Farwell (1975). For example, Walley (1993) posited that children with small vocabularies might represent a word like *cat* as "[+ abrupt onset]" or *cap* as "[+ labial]" (p. 293).

This suggestion may be placed at the "less specification" endpoint of a continuum spanned by current opinion about children's lexical representations. The "more specification" endpoint appears to be the dominant position in theoretical phonology, particularly among researchers working in Optimality Theory (OT). In OT, it is usually assumed that a segmental tier forms a part of children's lexical input representations (i.e., those representations that govern the recognition of words) and that this segmental specification is usually adultlike even in young children (Gnanadesikan, 1995; Hale & Reiss, 1998). For example, Smolensky (1996) holds that the notion that children's underlying forms "relatively closely approximate the adult forms" is "central to much of [the Optimality Theoretic] literature" (p. 721). The occurrence of misperception is recognized (Macken, 1980), but misperception is generally invoked to explain deviant pronunciations only for isolated cases or for a very short list of particularly confusable phones, such as [f] and [θ] (Vihman, 1996, pp. 161–162; for an exception in OT, see Boersma, 2000). Thus, as Fikkert (1998) wrote, "most approaches assume an input form that is more or less identical to the adult form ... " (p. 173). Children's deviant pronunciations of words are then held to arise from the operation of rules or constraints governing the child's productive phonology (e.g., Ingram, 1992; Smith, 1973).

Although lexical representations are frequently held to be *underspecified* in the theoretical phonology literature, this usually refers to information absent from a lexical representation per se but supplied by the child's grammar, or to information absent from the representation used in *production* (e.g., Levelt, 1994). Typically, neither sort of underspecification is argued to have consequences for word recognition. In fact, some recent work in phonological theory that advocates underspecification in representations takes no explicit position on children's perception in word recognition (e.g., Rice & Avery, 1995; but see Brown & Matthews, 1997).

Some experimental research on older children's lexical representations required children to make explicit similarity judgments about words. These studies revealed substantial false identifications of words (e.g., Gerken, Murphy, & Aslin, 1995; Storkel, 2002; Treiman & Breaux, 1982). For example, Gerken et al. (Experiment 3) asked four-year-olds to listen to a series of monosyllables emanating from a toy dog and indicate (via a button press) when the dog said a given word, such as *lick*. In contrast to adults, who almost never "false alarmed" to minimal-pair foils, children frequently responded positively to words like *nick* (38% false positives) or *gick* (21% false positives).

Thus, some experimental work suggests that children store only incomplete information about how words sound, whereas many theoretical phonologists tend to posit accurate perceptual representations, thereby adhering more closely to conclusions from discrimination experiments in infants. In our view, the most direct tests of the representations children use for word recognition may be provided by experiments in which the recognition of familiar words is the primary task, rather than children's production of words or children's explicit evaluation of whether pairs of words are similar. Children's productions are not ideal guides to children's receptive forms because the relation between perception and production is often opaque (perhaps maximally so when children can recognize words they do not say). Explicit similarity judgments are useful guides to children's metalinguistic interpretation of speech, but not necessarily the nature of the representations used in word recognition (Gierut, 1998). As a result, researchers testing young children's lexical representations have typically used object selection procedures in which children hear a word and then must indicate which of a limited set of pictures or objects is a member of the semantic category denoted by the word (e.g., Barton, 1980; Shvachkin, 1973).

In our work, we have modified the object selection task so that children's visual fixations to named objects served as the dependent measure, rather than children's pointing or touching. This modification was motivated by the desire to minimize task demands. Infants who do not point reliably and children who are reluctant to make an overt response do nevertheless tend to fixate named pictures. This tendency to look at named objects is present in infants hearing sentences like *Where's the ball?* (e.g., Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987; Swingley, Pinto, & Fernald, 1998), and in children and adults when requested to act on objects (e.g., *Touch the blue comb*; Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Trueswell, Sekerina, Hill, & Logrip, 1999). Measurement of fixations has become increasingly popular in the study of language processing across a range of ages, primarily because it provides information about the time-course of on-line sentence understanding without requiring participants to make metalinguistic judgments (Tanenhaus, Magnuson, Dahan, & Chambers, 2000). Given that it is notoriously difficult to induce very young children to make linguistic judgments (or to perform other explicit tasks consistently), a task that takes advantage of children's spontaneous behavior is welcome (see Mills, Coffey-Corina, & Neville, 1993, for a similar argument in favor of electrophysiological measures).

In our implementation of the visual fixation method, children are seated facing a large video monitor. On each trial, a pair of objects is displayed; a few seconds later, one of the objects is named in a sentence. Children's faces are filmed, and these films are then analyzed frame by frame to determine the timing of children's shifts from one picture to the other. These analyses have shown that children generally fixate the named picture: if children happen to be looking at the target (named) picture when its name is spoken, they continue to look at it, whereas if children happen to be looking at the distracter picture when the target word is spoken, they shift their fixation from the distracter to the target. The timing of such shifts from the distracter to the target reflects the timing of children's recognition that the semantic category denoted by the target word does not match the currently fixated picture (see Swingley & Fernald, 2002, for evidence supporting this interpretation). Children's overall tendency to fixate the named picture, expressed as a proportion of their overall looking time, serves as a measure of children's recognition of the word.

One study by Swingley, Pinto, and Fernald (1999, Experiment 2) used this procedure to evaluate 24-month-olds' ability to distinguish similar words. On some trials, children saw two pictures whose names rhymed (*ball* and *doll* or *truck* and *duck*); on others, children saw the same pictures with the pairings rearranged so that the names for the pictures had minimal phonological overlap (viz., *ball* and *duck* or *doll* and *truck*). Children were prompted using sentences of the form *Where's the [target]?*, where *target* was the name of one of the two pictured objects. The chief question of interest was how well children would perform on the rhyme trials relative to the minimal-overlap trials. Analysis of eye movements revealed no effect of overlap. Children were equally likely to fixate the target whether the names of the displayed pictures rhymed or not. Of the 32 children, 30 showed more target than distracter fixation on rhyme trials, and 30 showed more target than distracter fixation on minimal-overlap trials. Gradient measures of performance, including percentage of target fixation and response latency, also did not indicate that children had any difficulty resolving the difference between the similar words. Swingley et al. concluded that by 24 months, children's phonological representations of familiar words are sufficiently detailed to permit rapid and robust recognition without confusion among similar words, though the authors acknowledged that this might be limited to words for which children knew neighboring (phonologically similar) words, a point to which we will return below.

At 24 months, one may ask children to distinguish a ball from a doll (cf. Barton, 1976), but younger children tend not to know minimal-pair words that could be used in this kind of direct comparison. In two studies, Swingley and Aslin (2000, 2002) circumvented this problem by assessing children's responses to correct pronunciations (CPs) of words, and slight mispronunciations (MPs) of words. If children generally retain only a few phonetic features in the words they know, their sensitivity to mispronunciation should be limited; on the other hand, if children encode phonetic detail in words, mispronunciation would be predicted to hinder (though not necessarily prevent) word recognition.[2] A total of 106 children ranging in age from 14;04 (months;days) to 23;02 participated in the studies. All children were tested on six words: *apple*, *baby*, *ball*, *car*, *dog*, and *kitty*. For most children, the MP versions of these words were *opple*, *vaby*, *gall*, *cur*, *tog*, and *pity*; a subset of 14–15-month-olds were tested on the more severe mispronunciations *opal*, *raby*, *shawl*, *kier*, *mog*,

[2] Note that whether children's performance on the mispronounced targets exceeds chance or not is incidental to the primary question of interest; the important question for the hypotheses tested is whether the correct pronunciations elicit recognition to a greater degree than the mispronunciations. Whether responses on MP trials exceed chance addresses a separate question, bearing on how the recognition system handles a mismatch between the speech and the closest real word in the lexicon.

and *yitty*. Over the entire age range studied, children fixated the named picture more, and faster, upon hearing a CP than upon hearing an MP. This effect held for all of the words tested, and was not correlated with children's age, their reported vocabulary size (on a parental checklist measure), or their ability to say the tested words. Swingley and Aslin concluded that for familiar words, children's representations of the onsets of those words were encoded in detail, even among children as young as 14 months. Results consistent with this conclusion have also been presented by Fennell and Werker (using the same procedure as Stager & Werker, 1997, but with familiar words; 2002) and Bailey and Plunkett (2002, using a method similar to Swingley & Aslin's).

In the present experiments, we addressed two limitations of these studies. First, in our previous work only the onsets of words were tested (excepting the vowel in the test item *car*), and onsets may be encoded by young children with greater specificity than sounds later in words. This possibility has a functional motivation: given that young children interpret words incrementally over time (Fernald, Swingley, & Pinto, 2001; Swingley et al., 1999) and given young children's relatively sparse vocabularies (Charles-Luce & Luce, 1990), children might comprehend speech reasonably efficiently even if they only accurately represented, for example, the onset and first vowel of words, with the remaining sounds encoded only partially or even inaccurately. For example, if a child knew the word *baby* and no other words starting with *ba*…, hearing *ba*… would be sufficient for identification of that word. Thus, development might consist partly of the increasing specification of later parts of words. Here, Experiment 1 tested children's responses to MPs at word onsets, while Experiment 2 tested children's responses to MPs word-medially.

Second, the present study permitted a fuller exploration of the possible role played by phonetic *neighbors* in the word learning process. *Neighbors* are words that are phonologically similar enough for one to be converted to another through addition, deletion, or substitution of one segment (Luce & Pisoni, 1998). Several researchers have argued that at first children's representations of words are *holistic*, but as children learn more neighbors of words, those words lose their holistic character and become specified in terms of discrete segments (Charles-Luce & Luce, 1990; Fowler, 1990; Jusczyk, 1993; Logan, 1992; Metsala & Walley, 1998). For some authors, the development of segmental structure in childhood is primarily organizational: clusters of features or gestures become linked into phonological units (e.g., Fowler, 1990). For others, the development of segmental structure is accompanied by significant elaboration of the details of the representation itself. On the latter account, early lexical representations lack some phonetic or featural specifications entirely (Logan, 1992; Metsala, 1997a; Walley, 1993). This position goes beyond stating that children's weighting of subtle phonetic cues develops over a protracted period (e.g., Nittrouer & Miller, 1997; Ohde & Haley, 1997; Parnell & Amerman, 1978); the claim appears to be that children's holistic representations are not specified fully enough to permit the differentiation of minimal pairs.

Fuller specification is then argued to go hand in hand with the learning of neighbors. For example, Metsala (1997a) wrote,

> … representations of lexical items may become increasingly segmented (phonemic) with development from the pressure of an increasing vocabulary size. Young children

> may represent only those distinctions that are necessary for word recognition. As vocabulary increases, discriminating between lexical items will become more demanding. Words that have many similarly sounding neighbors may be forced to become phonemically represented chronologically earlier than words that do not have to be discriminated by many similarly sounding word neighbors (p. 161).

Similarly, Logan (1992) suggested that as children learn more words, "the increased size of the lexicon may serve as an impetus for the child to develop more detailed, adult-like representations of the words in the lexicon in order to keep individual items distinct in memory" (p. 31). Thus, the elaboration of early lexical representations depends upon the shape of the lexicon, such that words in sparse or empty neighborhoods are encoded with less detail than words for which neighbors force sufficient fidelity for word discrimination.

Swingley and Aslin (2002) reasoned that if neighbors are necessary for full phonological specification of words' forms, effects of subtle mispronunciations of words should be shown only for target words for which children knew neighboring words. Children's receptive vocabularies were assessed by a parental report measure (the MacArthur Communicative Development Inventory; Fenson, Dale, Reznik, Bates, Thal, & Pethick, 1994). This permitted division of the experimental items on a child-by-child basis into those for which children were reported to know at least one neighboring word and those for which children knew no neighbors. Swingley and Aslin found that 14–15-month-olds showed robust and significant effects of mispronunciation even for words for which they were reported to know no neighbors. In fact, the sizes of the effects were not different in the neighbor and no-neighbor groups.

This result indicated that children encode sufficient phonetic detail in words to permit superior recognition of correct forms relative to minimally different forms, even in the absence of any "pressure" for refinement from neighbors in the lexicon. However, Swingley and Aslin's assessment of children's vocabularies used parental reports, and might therefore have underestimated the set of potential neighbors. When children's productive vocabularies are small, parents must estimate children's lexical knowledge on the basis of evidence that the child knows what a word means. But children certainly know what some words mean before their parents recognize this. Children are also likely to know many word-forms that have not yet been attached to any meaning at all. Indeed, children probably store some speech sequences that are not meaningful words of the language, but are bits of speech resulting from children's errors in segmentation. It is known that even infants remember some of the speech that they hear; for example, Jusczyk and Hohne (1997) demonstrated that eight-month-olds remembered words they had heard in stories two weeks earlier. It is also clear that infants do not always extract sequences of speech that turn out to be actual words: Because words flow together in speech without clearly demarcated boundaries, segmentation of words from speech is a difficult problem. As a result, if meaningless stored speech sequences enter into neighborhood-based computations, neither parental vocabulary checklists nor lists of words children typically know provide an adequate test of the hypothesis that neighbors are required for the refinement of children's lexical representations. For example, Swingley and Aslin (2002) found that 14-month-olds recognized *dog* better than *tog* as a label for a canine, even if parents reported no words sounding similar to *dog* in their children's vocabularies. However, *dog* may have been encoded in detail only because children had in their lexicon of wordforms a neighbor of *dog*, forming a minimal pair—despite parents' reports to the contrary.

The present experiments considered two hypotheses about the mechanism by which phonological specification might be added to words through processes involving neighbors, where neighbors were considered to be phonetic word-forms potentially stored by children. Both hypotheses stipulate that any word and minimal variants of that word are treated identically in recognition unless the child has learned some word-form that differs from that word minimally, whether or not the new word-form is meaningful to the child or is even an actual word of the language. If, and only if, the child *has* encoded such a "minimal-pair word," then:

> *Hypothesis 1.* The word and the minimal-pair word will be encoded with sufficient detail to differentiate them from each other.

> *Hypothesis 2.* The word and the minimal-pair word will be encoded with sufficient detail to differentiate them from each other and from all other minimal variants differing in the segment not shared by the known pair.

Thus, in contrast to previous research, here we considered a wider range of word-forms that might in principle enter into neighborhood-based computations (i.e., word-forms that children may know but which parents may not detect in their children's vocabulary), and we identified two different hypotheses about the mechanism by which neighbors might influence lexical development.

Both hypotheses depended upon the premise that word-form contrast is instrumental in children's progress from underspecified lexical representations to representations that in principle would be sufficient for differentiating known words from other potential words. To clarify the hypotheses, we consider a few examples in which we assume that articulatory features are the elements by which contrasts are made. Hypothesis 1 suggests that a child knowing the word *pin* but no similar words might not encode the place feature of the initial / p / and would therefore be equally willing to consider [pɪn] and [kɪn] as good realizations of *pin*. Once this child learned *kin*, perhaps as a mis-segmentation of *pumpkin*, the place contrast highlighted by this minimal pair would be specified for both words. As a result, the child would be less willing to consider [kɪn] as a realization of *pin*. Thus, Hypothesis 1 supposes that neighbors cause retraction of words from any phonological space that is jointly occupied by two word-forms.

Here, Hypothesis 1 was tested by examining children's responses to correct pronunciations (CPs) and mispronunciations (MPs). Some of the MPs did not conform to the ambient language's phonology and were therefore extremely unlikely to have been stored by young children as familiar speech sequences. As a result, phonological specification in the segment by which the CPs and MPs differed (as revealed by better performance in recognizing the word upon hearing the CP than upon hearing the MP) could not have been driven by children's having learned the forms of the MPs through previous exposure. Such specification would have to have been driven by some other process than lexical contrast as defined in Hypothesis 1.

Hypothesis 2 suggests that the learning of *kin* would serve not only to fix the onset place feature of *pin*, but *all* of the features of the onset /p/ in *pin*. As a result, the child would be less willing to consider not only [kɪn], but also, for example, [dɪn] and [vɪn], as realizations of *pin*. Hypothesis 2 is broader than Hypothesis 1 in the sense that on Hypothesis 2 full specification in words is easier to achieve: children need not learn a neighbor constraining every contrastive feature, but only a neighbor constraining every segment. Hypothesis 2 was evaluated by examining corpora of speech to children to estimate the likelihood of children's having stored word-forms (possibly as missegmentations) that could have served as neighbors of the words we tested experimentally.

The children tested were Dutch monolinguals. Mispronunciations were formed by replacing one consonant in a word with either [d] (a frequent Dutch sound) or [g] (a very rare Dutch sound). The consonant [g] leads a rather tenuous existence in Dutch, being found only in certain loan words (e.g., /gol/, a goal in sports) and surfacing syllable-finally as a product of regressive assimilation in certain contexts (e.g., *zakdoek*, 'handkerchief', may be pronounced as [zagduk]). The distinction between [g] and the sounds of Dutch is not particularly difficult for Dutch adults to make; the [g] does not appear to be among those non-native sounds that are assimilated to native sounds in identification tasks. Our goal was to select two MPs that would be approximately equally "distant" from the CP, while involving phonemes differing in their input frequency. If MP effects were only found for [d] substitutions (such as pronouncing [beːʁ], *bear*, as [deːʁ]), and not for [g] substitutions (such as pronouncing [beːʁ] as [geːʁ]), Hypothesis 1 could serve as the explanation, on the assumption that children knew a word-form sounding like [deːʁ]. If MP effects were also found for [g] substitutions, Hypothesis 1 could be ruled out. Hypothesis 2 was tested by searching infant-directed speech corpora for word-forms that could in principle have served as neighbors of words for which children showed mispronunciation effects in the experiments. If no such neighbors could be found for at least some of the tested words, it would supply evidence against Hypothesis 2.

In two experiments, children 18–20 months of age participated in a picture fixation task testing their responses to CPs (correctly pronounced words; i.e., pronunciations that may be described by a prototypical sequence of phonemes) and MPs (pronunciations described by an atypical sequence), where MPs were generated by replacing a [b] or [k] with [d] (the MP-d condition) and with [g] (the MP-g condition). In Experiment 1, MPs were created by substituting the first sound in each target word; in Experiment 2, MPs were created by substituting the onset consonant of the second syllable of each target word.

2 Experiment 1

2.1 *Methods*

Children saw pairs of pictures on a large monitor, and then one of the pictures was named in a sentence. Each child heard CP and MP versions of each word on separate trials. Children's faces were videotaped and their eye movements were coded off-line. Analyses of children's proportion of fixation to the target picture focused on whether performance was better on CP trials than on MP trials.

Participants: The age range of the 36 children in the sample was 18;07 (months;days) to 20;03 with a mean of 19;09 or 586 days. Half were girls; the mean ages of girls and boys did not differ (586 days). All children were being raised monolingually, although parents occasionally reported some (< 15%) exposure to various southern Dutch dialects; these dialects do not use [g] more than standard dialects do. Parents were queried about their children's hearing and vision; no parents reported abnormalities. Thirteen children had previously participated in Experiment 2 some days before (range of days, 4 – 30; mean 10.5). An additional nine children were tested but not included in the final sample because the child refused to watch the displays or otherwise did not complete at least nine of the 14 test trials.

Stimuli: Auditory stimuli on test trials consisted of Dutch sentences of the form *Waar is de [target]?* ('Where is the [target]?'). The target words were *bal* ([bal]; 'ball') and *beer* ([beːʁ]); 'bear'). The MP-d items were [dal] and [deːʁ]; the MP-g items were [gal] and [geːʁ]. Of the MP words, only *dal* was a word, meaning *valley*; all parents reported that their child did not know this word. The stimuli were digitally recorded by a female native speaker of Dutch using a moderately infant-directed register. The carrier *Waar is de* averaged 558 ms in duration (range 480 – 630 ms). The durations of the target words (in ms) were as follows: *bal*, 532; *dal*, 490; *gal*, 476; *beer*, 623; *deer*, 615; *geer*, 621. Each target sentence was followed by a pause of 750 ms and then the sentence *Vind je 'm mooi?* (roughly, 'Do you like it?'). Auditory stimuli on filler trials included sentences like *Waar staat de poes?* ('where is the cat?') and *Zie je de koe?* ('Do you see the cow?').

The visual stimuli were digitized photographs of objects on a gray background, presented side by side on a large (127 cm diagonal) video projection screen (Sony KL-X9200M) fed through a VGA connection from a PC. The pictures on test trials were always the same: bear (a teddy bear) and ball (a children's toy with red and white stripes). Pictures slowly moved up and down in synchrony, a feature intended to help maintain children's interest in the displays. On the screen, the pictures were about 21 cm wide and separated horizontally by about 38 cm. On filler trials, pictures were also horizontally aligned and of similar size. Filler pictures included a baby, a book, a car, a cat, and other familiar objects.

Apparatus and procedure: The experiment was conducted in a sound-insulated room containing a three-sided enclosure 2 m tall, 1.3 m wide, and 1.2 m deep. The parent sat on a swivel chair in the open end of the enclosure, and the child sat on her lap, facing the video screen that formed the rear side of the enclosure. The (monaural) speech stimuli were produced from the speakers of the video screen. Children were videotaped onto DVC via a low-light videocamera placed about 15 cm below the monitor. Room lighting was dim.

Before arriving at the laboratory, parents had completed a Dutch version of the MacArthur Communicative Development Inventory: Words and Gestures (Fenson et al., 1994), modified from a Belgian (Flemish) inventory kindly provided by Dr. Maryline Lejaegere. Data were collected on both receptive and productive vocabulary.

When parent and child appeared ready, children were seated about 1 m from the video screen. Parents were instructed to refrain from speaking and to close their eyes and orient their heads downward during the procedure; thus, parents were blind to picture identity and target side (parents' compliance was monitored and enforced by the experimenter). Children first saw an animated display of a moving goldfish or toy duck. Once the experimenter saw that children were oriented to the screen, the first trial was initiated.

The experiment consisted of 14 test trials interspersed with 10 filler trials. Each trial began with the simultaneous presentation of both pictures, which children viewed in silence for 2.5 s. Then the auditory stimulus was played. Trials ended 4 s after the onset of the auditory stimulus. Between trials the screen was black for 0.5 s.

The bear was the target on half of the test trials, the ball on the other half. Target side was approximately counterbalanced by item within children (i.e., the bear was the target on the left three times and the right four times for half of the children, with the imbalance reversed for the other half). Among the 14 test trials were six CP trials, four MP-d trials, and four MP-g trials, with half of the trials for each condition occurring in each half of the experiment, and with the target appearing on the left and right approximately equally in each half. Four stimulus orders were created: the second reversed the trial order of the first, and the third and fourth were left / right reflections of the first and second. An approximately equal number of children were assigned to each order (8 to 10 children per order).

After the experimental procedure was done, parents were asked if their child knew the tested words and if their child knew any words containing the [g] sound; parents were given the words *gol* ('goal') and *taugé* ('bean sprouts') as examples. No parents indicated that their child knew any words with the [g] sound. However, as this experiment and Experiment 2 were nearing completion, we discovered that some children knew the word *buggy* ('baby carriage') which has evidently begun to supplant the native Dutch word *kinderwagen*. The word *buggy* was added to subsequent versions of our vocabulary checklist, and it appears that among children of about 19 months, about 46% (64 of 140 children in our current sample) know the word. As a result we must assume that about half of the children in the present experiments knew the word as well, though we do not know which children they were. This does not affect the first experiment, because *buggy*, however misseg-mented, could not generate a neighbor of *bal* or *beer*. We will return to this issue in Experiment 2.[3] All children were reported to know the word *bal*; two children were reported not to know the word *beer* (removing these 2 children from the sample did not affect the results).

Coding: Videotapes of the children were digitized into MPEG format and stamped with a digital stopwatch identifying each video frame (40 ms intervals). Several highly trained coders used custom software to step through the MPEG film frame by frame, noting for each frame whether the child was looking at the left picture, the right

[3] In the two years since we found *buggy* in the Dutch children's vocabulary, we have attempted to discover other [g] words known by children, without finding any.

picture, or neither, where *neither* contained two categories, *between*, meaning the child was shifting directly from one picture to the other, and *away*, meaning the child was looking away from the pictures entirely. This response-timing information was integrated with trial-timing information generated by automatic detection of tone pulses aligned with stimulus events, yielding an accurate record of the timing of children's responses to the target words. Coder reliability was evaluated by recoding a randomly selected block of six consecutive trials from a random sample of nine infants from Experiments 1 and 2. The mean percent agreement was 96.7% (subjects' range 91.5 to 99.8) and mean Cohen's kappa was .93 (range .83 to .997).

2.2
Results and Discussion

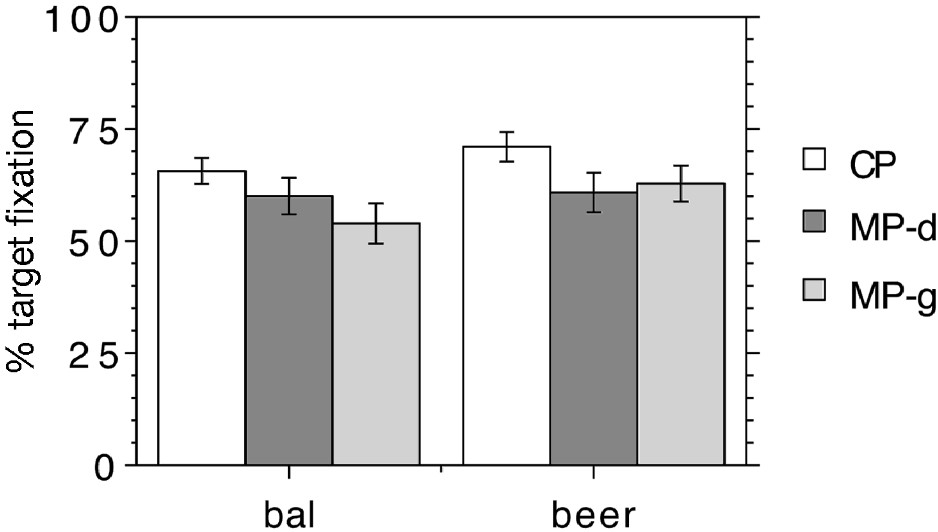
As in previous research (e.g., Swingley & Aslin, 2000, 2002; Swingley & Fernald, 2002) all results were computed over a temporal "window" extending from shortly after the onset of the target word (360 ms), to 2 s after the onset. This is the time period in which young children are generally most likely to exhibit eye movements contingent upon the target words. Within this window, we computed the proportion of time children fixated the target picture. This was calculated by counting the frames on which children fixated the target on a given trial, and dividing this quantity by the number of frames on which children fixated either the target or the distracter.[4] This proportional measure excluded time children spent *away* (off-task), which amounted to 16.8% of the total possible looking time (subjects SD = 11.1). This time did not vary by condition: CP, mean 17.2%; MP, mean 16.4%; $t(35) = 0.27$, ns.

The primary question to be addressed was whether children's performance differed in the CP, MP-d, and MP-g conditions. A repeated-measures analysis of variance (ANOVA) with three levels of condition indicated that fixation behavior differed significantly according to the target's pronunciation, $F(2, 70) = 4.89$, $p = .01$. Children's proportion of fixation to the target when hearing a CP was 67.9%, while this proportion was only 61.8% on MP-d trials and 58.1% on MP-g trials. All three of these proportions were significantly greater than 50% (*min* $t > 2.9$, *max* $p = .006$). Planned comparisons of CP with each variety of MP revealed that children performed significantly better on CP targets than on either kind of MP target, MP-d, $t(35) = 2.2$; MP-g, $t(35) = 3.1$, both p (1-tailed) $< .0167$, while mean target fixation in MP-d and MP-g conditions did not differ significantly, $t(35) = 1.1$, $p = .15$. As shown in Figure 1, these effects held for both of the target words, although the effects were not statistically significant for each word when tested separately (which is to be expected given that each child heard each of the 6 test words only twice).

4 In some previous research we have also computed *response latency*, that is, the time between the onset of the target word and children's initiation of an eye movement from the distracter to the target. In the present study there were too few trials per condition to estimate response latencies accurately, although over all RTs in the experiment (not grouping by subjects), CPs were nonsignificantly faster than MPs, mean CP, 815 ms; mean MP, 896 ms; $t(104) = 1.4$, $p = .15$.

Figure 1

Results of Experiment 1, showing children's proportion of fixation to the target picture upon hearing a correct pronunciation (CP), a mispronunciation involving [d] (MP-d), or a mispronunciation involving [g] (MP-g). Error bars are standard errors



To evaluate children's ability to recognize the tested words, children's target fixation proportions on the first part of the trial (before the onset of the target word) were compared with their target fixation proportions in the test window. Recognition was inferred when children's looking to the target increased significantly over their baseline fixation preferences as revealed in the beginning of each trial, when children viewed the pictures before hearing the target word. On this measure, children as a whole recognized *bal* in all three conditions (CP, mean fixation increase 28%; MP-d, 13%; MP-g, 14%; all $t(35) > 2.5$, all $p < .01$, 1-tailed) and recognized *beer* only in the CP condition (CP, mean fixation increase 9%, $t(35) = 1.9$, $p < .05$, 1-tailed).

Neither *deer* nor *dal* corresponded to words children were reported to know; indeed, only *dal* is a Dutch word. However, it is possible that both could have been stored by children as heard syllables, possibly as mis-segmentations of words like [sty'deːʁ] (from *studeren*, 'to study'). This possibility is very unlikely for *geer* and *gal*, given the rarity of [g] in Dutch. Thus, the MP effect found with MP-g words implies that children need not have learned a word-form corresponding to the MP in order to have encoded the correct form of the target in their lexicon. Stated in terms of phonological features, the results show that children exclude both [velar] and [alveolar] as specifications for the place of articulation of the onsets of *bal* and *beer*, in contradiction to Hypothesis 1.

Swingley and Aslin (2002) showed better recognition of correct pronunciations than of mispronunciations (i.e., MP effects) in children's recognition of words for which children were reported to know no neighbors. To replicate that result in the current Dutch sample, we examined the reported receptive vocabularies of each child, seeking neighbors of *bal* and *beer* that differed from these words at onset. This yielded two words: *peer* ('pear', known by 18 of the children) and *meer* ([meːʁ], 'more', known by 19). No children were reported to know monosyllabic words rhyming with *bal*. Thus, MP effects on variants of *bal* were unlikely to have resulted from children's prior learning of neighboring words. Examination of children's responses to *beer* and its MPs according to whether they knew a neighbor of *beer* yielded no evidence that the MP effects were restricted to children who knew a neighbor: For children knowing no neighbors ($n = 12$), mean target fixation on CP trials was 72.9% ($SD = 16$); on MP-d trials, 61.9% ($SD = 27$); and on MP-g trials, 63.7% ($SD = 22$). For children knowing at least one neighbor ($n = 24$), mean target fixation on CP trials was 66.7% ($SD = 16$); on MP-d trials, 58.6% ($SD = 27$); and on MP-g trials, 60.1% ($SD = 22$). Although the paucity of data available for each item prevents statistical confirmation of these patterns, the results are consistent with our previous findings with American children.

The possibility remains that children knew neighbors such as *peer* simply as word-forms without any meaning. As argued previously, children do learn sound-patterns that occur frequently in their environment, and until children come to know something of what a word means, parents may have no way to detect children's knowledge of these sound-patterns. As a result, Hypothesis 2 remains open. It holds that a word is stored in a phonetically vague form unless the child has also encoded a neighboring word-form (which need not be an actual word with semantic content). This neighbor motivates a retraction in the range of acceptable tokens for the original word: a retraction not only away from the neighbor, but away from other potential neighbors as well. Hypothesis 2 was evaluated by examining a 25,000-word corpus of speech to a Dutch infant (van de Weijer, 1998), searching for words (or parts of words) that might have been known to children in the study and which could have served as neighbors of *bal* or *beer*. In the corpus, some such words were attested, a few with fairly high frequency. For *bal* (6 occurrences), these included *al* ('all', 'already'; $n = 78$); *val* (a form of *vallen*, 'to fall'; $n = 4$); and *zal* (a form of the modal verb *zullen*; $n = 18$). For *beer* (16 occurrences), these included *weer* ('again', $n = 128$); *meer* ('more', $n = 26$); *zeer* ('very', $n = 4$); *keer* ('time', $n = 4$); *peer* ('pear', $n = 3$); and *neer* ('down', $n = 2$). In addition, neighbors of *bal* and *beer* were found within words. For example, the first (stressed) syllable of *heerlijk* ('delightful', $n = 21$), contains the form *heer*, a neighbor of *beer*. Thus, the corpus provided evidence that Dutch children hear syllables (often words) that could in principle serve as neighbors of the words tested in Experiment 1. As a result, Hypothesis 2 was not disconfirmed by this experiment.

A final set of analyses examined children's fixation performance in light of their vocabulary sizes. Reported receptive vocabulary sizes ranged from 58 to 330 words (median 181.5), while productive vocabulary sizes ranged from 3 to 283 words (median 52). There was no relationship between the size of the mispronunciation effect (i.e., target fixation on CP trials minus target fixation on MP trials) and the size of children's receptive vocabulary sizes ($r = .23$, $p > .15$) or productive vocabulary sizes ($r = .01$, *ns*). This replicated previous findings showing that robust mispronunciation effects do not depend upon children's having a large vocabulary.[5] Performance

[5] By contrast, degree of dishabituation in Werker's *Switch* procedure is marginally correlated with receptive vocabulary at 14 and 17 months, and significantly correlated with productive vocabulary at 14 months (Werker et al., 2002).

on CP trials was correlated with receptive vocabulary size ($r = .34, p = .041$) but only weakly with productive vocabulary size ($r = .18, p > .30$); performance on MP trials was not correlated with either vocabulary measure (maximum $r = .15, p > .35$). When the filler trials of the form *Zie je de [target]?* ('Do you see the [target]?') and all of the test trials (CP and MP) were used to compute a fixation score for each subject, correlations between performance and vocabulary size were significant (receptive vocabulary, $r = .52, p < .005$; productive vocabulary, $r = .35, p = .036$). We attribute the stronger correlation to the reduction in variability produced by including more trials in the estimate of each child's performance.

In sum, Dutch 19-month-olds found the words *bal* and *beer* easier to recognize when correctly pronounced than when mispronounced. This was equally true for mispronunciations using the frequent Dutch sound [d] and for those using the rare sound [g]. Mispronunciation effects were not contingent upon children's reported vocabulary size or upon their reported knowledge of words neighboring the targets. However, a corpus analysis suggested that speech to Dutch children contains words and word fragments that could in principle have served as neighbors motivating specification in both *bal* and *beer*. Thus, the results from Experiment 1 provided evidence against the notion that meaningful lexical neighbors, as indexed by parental report, are necessary for phonological refinement in the lexicon. This replicated Swingley and Aslin (2002). Furthermore, children's performance could not be accounted for on Hypothesis 1, the assumption that words are only represented in enough detail to distinguish them from similar-sounding word-forms in the lexicon. Dutch children's lexicons are unlikely to include sound-forms like *gal* or *geer*, whether with or without meanings, yet these pronunciations of *bal* and *beer* elicited less robust recognition responses than correct pronunciations did. However, the results were not sufficient for excluding Hypothesis 2, the assumption that neighboring sound-forms cause a more general refinement of lexical representations, not only away from the neighbor but also away from other potential neighbors (including neighbors containing [g]). Hypothesis 2 remained viable because of corpus-based evidence suggesting that Dutch children are exposed to sound sequences that include forms neighboring the tested target words. The MP effects shown in Experiment 1, including the effects for [g] substitutions, could (on Hypothesis 2) have resulted from children's encoding of these forms.

Experiment 2 provided an additional test of these hypotheses about the potential role of phonological neighbors in lexical development. Experiment 2 also tested children's responses to word-medial mispronunciations, to evaluate the generality of previous results found with word-initial mispronunciations. If children only represent the onsets of words in fine detail, MP effects would not be expected in Experiment 2.

3 Experiment 2

Previous uses of the mispronunciation method to test one-year-olds' knowledge of lexical form have focused on word-initial sounds. Here, children's responses to word-medial MPs were tested. The targets employed were *baby* ([beːbi], 'baby') and *beker* ([beːkəχ], 'cup'); the mispronunciations tested substituted the second consonant of each target, replacing the [b] or [k] with [d] (MP-d condition) and [g] (MP-g condition). These words were selected because (a) they are known to children in this age range; (b) they are picturable; and (c) they overlap phonetically at onset. The third constraint was necessary because of children's incremental interpretation of speech. If children viewing a car and a baby heard *Where's the bady?*, they might fail to show an MP effect simply because hearing [be] was sufficient for rejecting the car as a possible referent (Fernald et al., 2001; Swingley et al., 1999). By pairing *baby* and *beker* the informational relevance of the manipulated consonant was assured, just as the initial consonants of *bal* and *beer* had been.

3.1 Methods

Participants: The 24 children in the sample had a mean age of 19;05 (583 days) with a range of 17;30 to 20;17. Eleven were boys; the mean ages of boys and girls were nearly identical (boys 583 days, girls 584). Eleven children had participated in Experiment 1 some days before (range of days, 7–14; mean 11.5). An additional eight children were tested but not included in the final sample because of failure to complete at least 9 of the 14 test trials. Four children in the final sample were reported not to know the word *baby*, and one of these was the only child reported not to know the word *beker*. Analyses rerun excluding the latter participant and including only the *beker* data for the other three participants did not affect the pattern of results or the pattern of significance.

Stimuli and Procedure: As in Experiment 1, auditory stimuli on test trials were Dutch sentences of the form *Waar is de [target]?* ('Where is the [target]?'). Stimuli were recorded by the same speaker using an infant-directed register. The carrier *Waar is de* averaged 517 ms in duration (range 480–573 ms). The durations of the target words (in ms) were as follows: *baby*, 709; *bady*, 743; *bagy*, 720; *beker*, 713; *beder*, 708; and *beger*, 728. Each target sentence was followed by a pause of 750 ms and a second (uninformative) sentence such as *Vind je'm mooi?* ('Do you like it?') or *Kun je 'm zien?* ('Can you see it?'). Auditory stimuli on filler trials included sentences like *Kijk naar de grote bal* ('Look at the big ball.') or *Zie je de auto?* ('Do you see the car?').[6] The visual stimuli were displayed as in Experiment 1. Pictures on test trials were of a baby (seated, facing forward) and a child's sipping cup.

The procedure was as in Experiment 1, with the same counterbalancing restrictions. Experiment 2 contained 14 test trials and 12 filler trials. An approximately equal number of children were assigned to each order (5 to 7 children per order). Coding was done as in Experiment 1.

3.2 Results and Discussion

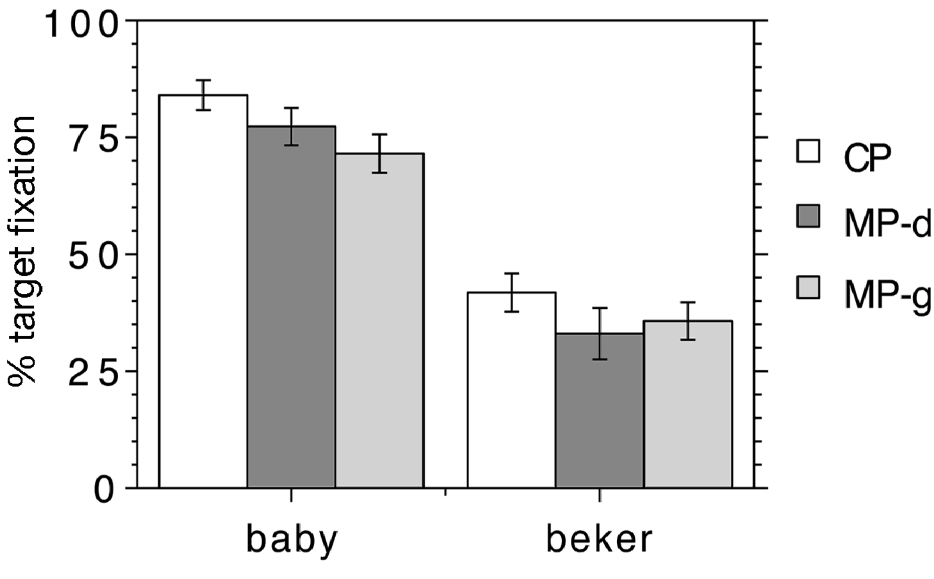
Once again, the chief question to be addressed was whether children's recognition performance, as indexed by percentage of target fixation, was greater for CP words

6 Filler trials with adjectives used pictures contrasting on that adjective, such as a large ball and a small ball. Children's performance on these trials was not different from chance and not correlated with measures of vocabulary size.

than for MP-d or MP-g words. A repeated-measures analysis of variance (ANOVA) with three levels of condition indicated that fixation behavior differed significantly according to the target's pronunciation, $F(2, 46) = 3.79, p = .03$. Children's proportion of fixation to the target when hearing a CP was 64.2%, while this proportion was only 56.2% on MP-d trials and 54.4% on MP-g trials. Children's *away* (off-task) fixations averaged 12.8% of the total possible looking time (subjects $SD = 12.7$), and did not vary by condition (CP, mean 14%; MP, mean 12%; $t(23) = 1.2$, *ns*). Planned comparisons of CP with each variety of MP revealed that children performed significantly better on CP targets than on either kind of MP target (MP-d, $t(23) = 2.1$, $p = .023$; MP-g, $t(23) = 3.1, p < .005$), while mean target fixation in MP-d and MP-g conditions did not differ significantly ($t(23) = 1.1, p > .3$; all p 1-tailed). As shown in Figure 2, these effects were present for both of the target words.

Figure 2

Results of Experiment 2, showing children's proportion of fixation to the target picture upon hearing a correct pronunciation (CP), a mispronunciation involving [d] (MP-d), or a mispronunciation involving [g] (MP-g). Error bars are standard errors



The figure also suggests that children had a general preference for fixating the baby picture rather than the cup. This preference was expected given prior piloting work (and was the reason for using different words in Experiment 1), but does not affect our conclusions here. Children's target fixation proportions in the test window were compared with childrens' target fixation proportions in the first part of the trial, before the target word's onset (as in Experiment 1). On this measure, children recognized *baby* and *beker* in the CP condition, baby, $t(23) = 4.2, p < .0005$; beker, $t(23) = 2.1$, $p = .022$. Children's baseline-corrected proportions of target fixation did not exceed chance on MP-d or MP-g trials for either word.

The MP effects found here replicated the primary finding of Experiment 1: even when MPs were words children were unlikely ever to have heard (even as fragments), children showed reliable decrements in recognition performance. Thus, a word like *baby* does not occupy the nearby phonetic space of *bagy*, as Hypothesis 1 would predict. Furthermore, analyses of children's reported receptive vocabularies revealed no neighbors of *baby* or *beker*.[7] It is possible that the MP-g effect for *bagy* was in part a consequence of competition from the word *buggy*, which some children probably knew. We consider it unlikely that this competition contributed much to the effect, however, given the similarity of the results in the MP-d and MP-g conditions and the absence of neighborhood competition effects in the two experiments of Swingley and Aslin (2001, 2002).

To evaluate the possibility that children might have extracted word-forms neighboring *baby* or *beker*, a search for such neighbors was conducted on the van de Weijer (1998) corpus, as in Experiment 1. In the corpus, *beker* occurred once and the diminutive variant *bekertje* occurred five times. Two bisyllables in the corpus corresponded to neighbors of *beker*: *zeker* ('certain'; 5 occurrences) and *beter* ('better'; 17 occurrences). The latter is a neighbor that could in principle help refine the second consonant of *beker*, as the two words form a minimal pair differing in that consonant. However, no syllable pairs in the corpus were neighbors of *baby* (3 occurrences, plus 4 diminutive variants). Note that this search was conducted not only over *words* in the corpus but over *syllable pairs* regardless of their lexical status; thus, for example, the analysis showed that the corpus contained no instances of two adjacent monosyllabic words together forming a neighbor of *baby*. If the corpus may be taken as representative of the speech that young Dutch children hear, then it is unlikely that the participants of Experiment 2 showed MP effects on the word *baby* as a consequence of having previous experience differentiating *baby* from any neighboring speech sequence.

A similar analysis was completed for the bisyllabic English words tested by Swingley & Aslin (2000, 2002).[8] Those words included *apple*, *baby*, and *kitty*. The English corpus was constructed by concatenating and regularizing several corpora from the CHILDES database (MacWhinney, 1995), as described in Swingley (1999). The corpus consisted of approximately 50,000 word tokens spoken to a total of 15 children under 18 months of age. The word *apple* occurred 19 times in the corpus. No neighbors were present. The word *baby* (n = 208) had three neighbors: *babies* (n = 20), *maybe* (n = 18), and *bubby* (n = 1). The word *kitty* (n = 108) had one neighbor (*city*; n = 2). Thus, the English analysis suggested that children might know word-forms neighboring *baby* and *kitty*, but not *apple*.[9]

7 The nearest words were *lekker* ('yummy') and *kikker* ('frog').

8 Neighbors of the monosyllabic test items were attested in corpora of English infant-directed speech, so this analysis included only the bisyllables.

9 One might argue that the phrase *an apple* might be misanalyzed as *a napple*, thereby yielding a neighbor of *apple* that could trigger refinement of the onset of *apple*. While this kind of segmentation error undoubtedly occurs, in our view it is unlikely to be the source of fine phonetic specification in the lexicon.

Although these analyses of corpora do not establish the full range of possible bisyllables that might have been stored by the participants in our mispronunciation studies, they do suggest that neighbors of the bisyllabic words tested in these studies cannot be assumed to have been learned by children. This reinforces the conclusion that children learn the phonological specifications of words without the potential benefit of "pressure" to resolve ambiguities in minimal pairs.

A final set of analyses examined children's responses relative to their reported receptive vocabulary sizes (range 58 to 315 words, median 177) and productive vocabulary sizes (3 to 283 words, median 59). There was no relationship between the size of the mispronunciation effect and the size of children's receptive vocabularies ($r = .06$) or productive vocabularies ($r = .07$; both *ns*). No significant correlation was found between either vocabulary measure and any measure of fixation performance, in contrast to Experiment 1. We attribute the difference to the smaller number of participants in Experiment 2, and the inclusion of fewer simple fillers of the *Do you see the [target]?* variety, leading to a smaller number of trials upon which to establish each child's performance level. (Recall that Experiment 2 included a number of difficult filler trials that yielded only chance performance.)

Children's reduced performance on MP trials implied that children's representations of the tested words were sufficiently precise to favor the CP realizations in our stimuli. This result extended to word-medial consonants earlier data indicating precision in children's lexical representations. As in previous research, MP effects were not contingent upon the presence of neighbors nor upon vocabulary size (Bailey & Plunkett, 2002; Swingley & Aslin, 2000, 2002). The finding of impaired performance on MP-g trials provided evidence against Hypothesis 1, under which MP effects are only expected when children are tested using mispronunciations they have stored in the lexicon as neighbors of the CP words. A corpus analysis revealed that one of the two tested words was unlikely to have been specified in detail as a result of children's extraction of neighboring sound-forms, thereby providing evidence against Hypothesis 2.

4 General discussion

The premise of these studies was that young children with vague or phonologically underspecified representations of spoken words should be indifferent to mispronunciations involving substitutions of one sound with a similar sound, and that better recognition of correctly pronounced words is therefore evidence for well-specified lexical representations. The results show that children's representations of familiar words are better specified than they need to be to permit differentiation of the words in the child's vocabulary. By the age of 19 months, even mispronunciations containing a rare sound in place of a frequent one impaired recognition, as did word-medial mispronunciations.

These results contrast with some studies in which children appeared not to distinguish the forms of newly-learned words from either variant pronunciations, or from other just-taught words. As reviewed in the introduction, research testing children's phonological knowledge through word teaching or explicit word comparison has not supported theories in which infants' finely-tuned perceptual abilities are relevant for word learning. The mispronunciation effects shown here with familiar words do not implicate language-specific phonological tuning, but they do help rehabilitate the possibility that young children store words the same way adults do — not as vague or "holistic" soundscapes containing only a few phonetic features, but as precise, well-specified phonetic forms.

This suggests that the natural learning procedure that Dutch children undergo in learning words like *bal* and *beker* (or American children *ball* and *dog*) and the laboratory learning procedures used in some word teaching experiments have different consequences for children's knowledge of words' forms. However, other explanations of the data available to date remain possible.

Negative results in mispronunciation studies (such as those of Hallé & de Boysson-Bardies, 1996) can be explained either by appeals to underspecified representations, or to activation processes. As shown here and in Swingley and Aslin (2000, 2002), mispronunciations of words can be recognized as instances of their targets by young children; semantic activation occurs, albeit more slowly or less reliably than when children hear correct pronunciations. If, for example, the procedure of Stager and Werker (1997, Experiment 2) were not sensitive to subtle differences in the degree to which words are activated, MP effects would not be found. This alternative would be excluded if children in the same procedure did show MP effects on known words; studies of this sort are in progress (see Fennell & Werker, 2003, for a step in this direction).

It is also not entirely clear why 14-month-olds fail to learn the phonetic details of pairs of similar words taught in the laboratory (e.g., Shvachkin, 1948/1973; Stager & Werker, 1997, Experiment 1; Werker et al., 2002). One possibility is that children encode the sounds of the words accurately, but cannot remember which sound-form corresponds to which word. When an object previously referred to as a *bih* is then referred to as a *dih* (the name given to a second object and therefore a familiar sound-form), children's apparent failure to detect the change could signal uncertainty regarding the appropriateness of the referent. On this account, however, one must treat as coincidental the developmental shift in children's behavior between 14 and 17 months: in both the "mispronunciation" version of Werker's Switch procedure (one word taught) and the "other word" version (two words taught), 14-month-olds generally fail and 17-month-olds generally succeed.

For the purpose of discussion, then, let us assume that under some conditions, when one-year-olds first learn a word, the phonetic form of that word is underspecified such that one-feature deviations from the taught form do not impair recognition. The question that immediately arises is what leads to refinement in the representation of partially specified words. The simplest answer is probably *repetition*: if children heard the taught words more, they would encode the phonetic detail better. It is worth bearing in mind, though, that the 14-month-olds in Werker's studies heard the taught words many, many times — usually between 50 and 100 repetitions, immediately before the children's encoding was tested. If repetition alone is the key to fine-grained encoding, either an enormous amount of exposure is required, or massed repetition in a single session is not an effective teaching method.

Another answer that has been offered is that the acquisition of neighbors in the vocabulary leads to refinement of lexical representations (Charles-Luce & Luce, 1990; Jusczyk, 1993; Metsala & Walley, 1998). We have provided evidence casting doubt upon some potential mechanisms of this sort.

First, the representation of a word might occupy all nearby phonetic space that is not assigned to another meaningful word (thus, learning *bear* and *pear* might refine the b/p distinction in both words). This view was ruled out by Swingley and Aslin (2000, 2002) and by the present results: subtle mispronunciations that did not correspond to correctly-pronounced words in the participants' lexicons nevertheless resulted in impaired recognition. Second, knowledge of two neighbors might cause the phonetic space assigned to each to recede, not only away from each other but from other potential words as well (thus, learning *baby* and *maybe* might result in MP effects upon hearing a baby referred to as *vaby*). This view was ruled out by Swingley and Aslin (2002) and by the present results: subtle mispronunciations of words for which children knew no neighbors resulted in impaired recognition. Thus, if one stipulates that the set of neighbors is limited to words parents report as known by their children (thereby including actual words for which children seem to know some meaning), neighbors do not appear to be necessary for the refinement of lexical representations, at least not in the sense that vaguely encoded words gain full phonological specifications only when neighbors of those words are learned.

The present experiments provide parallel conclusions for a much broader set of neighbors: not only the words children were reported to know, but all word-forms that children might *potentially* know, constrained only by the phonology of the language and by the set of syllables that constitute the child's linguistic environment (as estimated by corpora). Our Hypothesis 1 was that knowledge of a word-form, though not necessarily a word and perhaps without semantic content, might retract the space assigned to a word. Thus, Dutch children knowing *beer* and having extracted the word-form *geer* might therefore show an MP effect upon hearing a bear referred to as *geer*. This account was rendered unlikely by the present results, given that [g]-initial syllables rarely occur in Dutch.

Our Hypothesis 2 was that knowledge of a neighboring word-form might lead to a more general refinement in the sound differentiating the neighbors. Ruling out this account definitively would require knowledge of which word-forms the participants in our experiments had stored, a modeling problem with many unknowns (Brent & Cartwright, 1996; Christiansen, Allen, & Seidenberg, 1998; Swingley, under review). However, examination of corpora of infant-directed English and Dutch speech suggested that for at least two of the words for which children have revealed mispronunciation effects, prior extraction of neighboring forms was unlikely.

Of course, these hypotheses do not exhaust the set of effects that neighbors might have on phonological specification. For example, suppose that learning *bear* and *pear* helped children to acquire the skill of correctly identifying the sounds [b] and [p]. If so, the effects of learning this minimal pair could, in principle, percolate through the lexicon.[10] A child who previously failed to store a [voicing] value for

[10] The author thanks Sharon Peperkamp and Joe Pater for suggesting proposals related to this one.

pony might do so as a result of learning to make the [voicing] distinction in *bear/pear*. To account for children's reduced performance on *beger* relative to *beker*, however, one must assume that children can identify feature values (such as [±voicing]) independently of the segments in which they appear. That is, because [k] and [g] do not contrast in Dutch infant-directed speech, it cannot be the case that children learned to treat [k] and [g] as distinct in *beker* and *beger* because they had previously learned some other k/g minimal pair. If neighbors are to be implicated in the setting of [–voice] for [k] in *beker*, it is necessary to assume that children first learned at least one pair of words contrasting in voicing (such as *beer* and *peer*), abstracted [voice] as a feature critical for lexical contrast, and then applied this to the word *beker*.

An analysis of Dutch children's reported vocabularies revealed surprisingly few minimal pairs differing only in voicing—the CDI included only the examples *beer* and *peer*. These two words were also found in the van de Weijer infant-directed speech corpus, which otherwise yielded no voicing minimal pairs that were also object labels. Apart from a few nonsense words, the only minimal pairs contrasting in voicing at onset (other than *beer* and *peer*) were *de* and *te* (both function words), and *doe* and *toe* (a light verb and a preposition). Nevertheless, the existence of such minimal pairs precludes definitive exclusion of this neighbor-driven feature-generalization account.

The assumption that children can abstract knowledge of feature values from one segment (such as [–voice] in the [p] of *peer*) to other segments in other words (such as [–voice] in the [k] of *beker*), calls for a level of phonological proficiency that appears to be incompatible with the positions of Walley and Metsala and other psycholinguists who have suggested that young children have vague lexical representations. For example, Metsala (1997b) argued that phonological detail in the lexicon emerges "on an item-by-item basis" and not in a "system-wide fashion" (p. 48). That is why it is clusters of neighbors that are held to be encoded in detail, rather than all words containing segments (or features) whose distinctiveness is discovered via comparison of minimal pairs elsewhere in the lexicon.

In our view, the notion that children *can* perform the analyses required for the feature-generalization proposal, but *do not* perform such analyses until motivated to do so by a minimal pair, is unlikely to be true. The chief problem with this hypothesis is one shared by all theories that assume that children ignore phonetic or phonological detail until learning minimal pairs forces children to take this detail seriously, and that is the problem of learning the minimal pairs in the first place. If Dutch children initially fail to represent [voicing] in *beer* or *peer*, why do they not continue to treat the words as homophones?

The mechanism requires that when children know a word that seems to have a bimodal set of meanings (or, possibly, differing syntactic distributions), children then scrutinize the phonetics of that word more closely, seeking the phonological feature that distinguishes them. If Dutch children notice that [?eːʁ] sometimes refers to 'bears' and sometimes refers to 'pears', this could in principle lead children to attend more closely to the realizations of the (initially conflated) word-form, and eventually discover that [–voice] tends to be used with the fruit. But this implies an enormous computational endeavor. Even if children always correctly identified adults' intended referents for a given word, this set of referents would often appear "bimodal," triggering phonetic scrutiny. Grizzlies and Winnie-the-Pooh can both be called *bear*; whole Bartletts and pureed Boscs can be referred to as *pear*. Similarly, *can* as a verb and *can* as a noun have very different syntactic distributions, which could trigger formation of two lexical categories. But any phonological conclusions drawn by separating these *pears* and *cans* would be false. It is possible, of course, that children await substantial supporting evidence before permitting apparent minimal pairs or sets thereof to reconfigure their system of phonological contrasts; if so, in the long run any accidental, misleading phonetic differences between realizations of pear-as-fruit and pear-as-puree should average to nothing. Nevertheless, the use of semantic or syntactic categories as a means of structuring the phonological system does not strike us as a robust procedure. In our view it is more likely that the phonological system is built upon the child's discovery of language-specific phonetic categories in infancy—categories that are the necessary ones for the task.

We are still left with the question of why young children sometimes appear not to encode phonological detail in new words, and what subsequently leads to refinement in these words. Theories of speech perception that wish to account for infants' and adults' learning of phonetic categories hold that non-native sounds are assimilated to native categories (e.g., Best, 1994; Kuhl, 1995; see also Flege, 1995). If so, surely native sounds are assimilated to native categories as well. Furthermore, perceptual assimilation to native categories is not assumed to require attention. Why, then, should infants *ever* make encoding errors when presented with the clear speech used in word-teaching experiments? We suspect that in many cases the answer is simple: children, like adults, perceive speech in terms of the kinds of distinctions that they learned to make as infants. But under some circumstances, this phonetic information that was accurately perceived is *forgotten*; that is, at the moment of hearing a word, its sounds may be accurately categorized, but nevertheless not stored in memory in a format that can then help guide the process of word recognition. Young children's problem is not perception per se; it is encoding in memory. The notion that infants can forget is not controversial; adults and older children forget the sound-forms of new words too (Aitchison & Chiat, 1981; Smith, Macaluso, & Brown-Sweeney, 1991). Although 14-month-olds' forgetting of the initial sound of *bih* after just having heard it 100 times remains mysterious, the more general case of children only gradually learning the words in their natural environment is consistent with both the automatic nature of perceptual categorization and children's apparent failure to store words accurately under some circumstances.

Of course, it remains unclear exactly when children will encode a word-form in detail, and what is left in the representation when some forgetting has taken place. At present it is possible to set a few limits. Children's "best" words, such as *ball* and *baby*, appear to be represented with fine phonetic detail at least as early as 14 months (Fennell & Werker, 2003; Swingley & Aslin, 2002). Novel words taught in Werker's "Switch" procedure appear to be vaguely encoded at 14 months, and well-encoded at 17 months (e.g., Werker et al., 2002). Studies in progress are testing for MP effects in words taught under varying conditions at different ages (e.g., Swingley, 2002).

Positive results in mispronunciation studies admit various interpretations. It seems clear that the best cue to children's recognition of a familiar word is a realization that may be described as a CP. This means that children have encoded whatever detail is necessary for making the distinction between the CP and MP, whether this is best considered as a phonemic representation or not. However, it is not yet known what the functional consequences of this ability are. Mispronunciation effects do not imply that children are aware of deviations from the ideal form; indeed, young children's apparent failures in word-learning and similarity-judgment tasks suggest that they are not. Awareness aside, mispronunciation effects also do not prove that variation in a word's realization leads children (or children's implicit statistical processors) to keep track of associated linguistic or extralinguistic probabilities. For example, German children hearing *Mutter* ([mʊtɐ], 'mother') and *Mütter* ([mʏtɐ], 'mothers') might detect a difference in vowels without this detection triggering an analysis of differential linguistic co-occurrences (e.g., *Mutter* occurs with singular verbs, *Mütter* with plurals) or situational co-occurences (e.g., one mother vs. more than one mother). We suspect that statistical analyses of this sort begin in infancy, but little is known about this process.

The experiments presented evaluated whether neighbors are *necessary* for the specification of words, where *specification* refers to the presence of sufficient phonetic detail in the representation to cause correctly-produced tokens of the word to be recognized better than incorrectly-produced tokens. The present work does not indicate whether neighbors might help refine representations that are already specified to this standard. For example, the notion that children entirely fail to specify the [place] feature for the onset of *bal* or the medial consonant of *baby* conflicts with our results. But the learning of *dal* or other neighbors of *bal* could nevertheless contribute to a better phonetic representation of the word, beyond what might be present in the phonological specification. Children who know that *bal* begins with /b/ nevertheless probably do not have adultlike knowledge of the phonetic cues (and the weightings of these cues) that signal the /b/–/d/ distinction, and it is possible that gaining this knowledge depends upon, or is facilitated by, the presence of minimal pairs or near-pairs in the lexicon (Walley, 1993; Walley & Flege, 1999). Our results showing no effects of the presence or absence of single neighbors failed to support this prediction, but the present experiments cannot be considered a strong test of it either. Such neighborhood effects might refine children's representations of words at a fine-grained level that affects recognition of less clear tokens of the words, or has a stronger impact in other tasks. In addition, fine-grained neighborhood-driven development might take place only under the influence of large numbers of neighbors, in which case this mechanism would have relatively little influence in children as young as those tested here, given their small vocabularies (Charles-Luce & Luce, 1990).

Lexical neighborhoods might also indirectly relate to phonetic encoding in word learning via *phonotactic probability*. Words with many phonological neighbors tend to be made up of frequent sounds in frequent combinations (Vitevitch & Luce, 1998). Recent research testing 3- to 6-year-olds' learning of phonotactically probable or improbable words (i.e., nonce words constructed from frequent sequences of segments or infrequent sequences of segments) showed better learning of words made of phonotactically probable sequences (Storkel, 2001; see also Storkel, 2002). Given that the phonotactically probable words almost certainly had more phonological neighbors than the phonotactically improbable words, this effect might have been a consequence of neighborhood structure rather than phonotactic probability per se, though these somewhat different descriptions are difficult to disentangle empirically (Bailey & Hahn, 2001; Vitevitch & Luce, 1998). In either case, we interpret Storkel's results as suggesting that although young children can perceive the contrasting sounds of their language, the formation of a robust and accurate lexical representation depends to some degree on how the word's component sounds are used in the language: well-trodden paths are more readily traveled upon again.

The present results, together with those of Swingley and Aslin (2000, 2002) and those of Fennell and Werker (2003) and Bailey and Plunkett (2002), help to establish how much phonological detail is present in the words young children know, and set some constraints on the processes by which this detail is learned. Across a range of words and a variety of phoneme substitutions, both Dutch- and English-learning children find correct pronunciations of words easier to recognize than mispronunciations. Correct pronunciations can only provide a better match to children's stored representations if those representations specify the phonetic characteristics that distinguish the correct pronunciations from the mispronunciations. By 19 months, this specification is present for word-initial and word-medial consonants, even in words for which children do not know neighbors. We conclude that this specification is not a consequence of children's knowledge of neighbors. Although lexical encoding is not error-free in childhood, we believe that it is guided by learning that begins in infancy, including the learning of phonetic categories that are a part of the language's phonology.

References

AITCHISON, J., & CHIAT, S. (1981). Natural phonology or natural memory? The interaction between phonological processes and recall mechanisms. *Language and Speech*, **24**, 311–326.

BAILEY, T. M., & HAHN, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, **44**, 568–591.

BAILEY, T. M., & PLUNKETT, K. (2002). Phonological specificity in early words. *Cognitive Development*, **17**, 1265–1282.

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

BARTON, D. (1980). Phonemic perception in children. In G. H. Yeni-Komshian, J. F. Kavanagh, & C. A. Ferguson (Eds.), *Child phonology* (Vol. 2: *Perception*, pp. 97–116). New York: Academic Press.

BEST, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception* (pp. 167–224). Cambridge, MA: MIT Press.

BOERSMA, P. (2000). *Phonetically-driven acquisition of phonology.* (unpublished manuscript, University of Amsterdam.)

BRENT, M. R., & CARTWRIGHT, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition*, **61**, 93–125.

BROWN, C., & MATTHEWS, J. (1997). The role of feature geometry in the development of phonemic contrasts. In S. J. Hannahs M. & Young-Scholten (Eds.), *Focus on phonological acquisition* (pp. 67–112). Amsterdam: Benjamins.

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

CHEOUR, M., CEPONIENE, R., LEHTOKOSKI, A., LUUK, A., ALLIK, J., ALHO, K., & NÄÄTÄNEN, R. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience*, **1**, 351–353.

CHRISTIANSEN, M. H., ALLEN, J., & SEIDENBERG, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes*, **13**, 221–268.

EDWARDS, M. L. (1974). Perception and production in child phonology: The testing of four hypotheses. *Journal of Child Language*, **1**, 205–219.

EILERS, R. E., & OLLER, M. K. (1976). The role of speech discrimination in developmental sound substitutions. *Journal of Child Language*, **3**, 319–329.

FENNELL, C., & WERKER, J. (2003). (this issue). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, **46**, 245–264.

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 number 242.

FERGUSON, C. A., & FARWELL, C. B. (1975). Words and sounds in early language acquisition. *Language*, **51**, 419–439.

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. (2001). When half a word is enough: Infants can recognize spoken words using partial acoustic-phonetic information. *Child Development*, **72**, 1003–1015.

FIKKERT, P. (1998). The acquisition of Dutch phonology. In S. Gillis & A. De Houwer (Eds.), *The acquisition of Dutch* (pp. 163–222). Amsterdam: Benjamins.

FLEGE, J. E. (1995). Second language speech learning. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Baltimore, MD: York Press.

FOWLER, A. E. (1990). How early phonological development might set the stage for phoneme awareness. In S. A. Brady & D. P. Shankweiler (Eds.), *Phonological processes in literacy* (pp. 97–117). Hillsdale, NJ: Lawrence Erlbaum.

FOWLER, C. A. (1986). An event approach to the study of speech perception from a direct-realist perspective. *Journal of Phonetics*, **14**, 3–28.

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.

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.

GIERUT, J. A. (1998). Production, conceptualization, and change in distinctive featural categories. *Journal of Child Language*, **25**, 321–341.

GNANADESIKAN, A. E. (1995). *Markedness and faithfulness constraints in child phonology.* (ms, Rutgers Optimality Archive 67)

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.

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

HALLÉ, P. A., & BOYSSON-BARDIES, B. de (1996). The format of representation of recognized words in the infants' early receptive lexicon. *Infant Behavior and Development*, **19**, 463–481.

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.

INGRAM, D. (1992). Early phonological acquisition: A crosslinguistic perspective. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 423–435). Timmonium, MD: York Press.

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., & HOHNE, E. A. (1997). Infants' memory for spoken words. *Science*, **277**, 1984–1986.

KUHL, P. K. (1995). Mechanisms of developmental change in speech and language. In K. Elenius & P. Branderud (Eds.), *Proceedings of the XIIIth international congress of phonetic sciences* (Vol. 2, pp. 132–139). Stockholm: Stockholm University.

KUHL, P. K., WILLIAMS, K. A., LACERDA, F., STEVENS, K. N., & LINDBLOM, B. (1992). Linguistic experience alters phonetic perception in infants by six months of age. *Science*, **255**, 606–608.

LAHIRI, A., & JONGMAN, A. (1990). Intermediate level of analysis: Features or segments? *Journal of Phonetics*, **18**, 435–443.

LEVELT, C. C. (1994). *On the acquisition of place*. Unpublished doctoral dissertation, Holland Institute of Generative Linguistics.

LOGAN, J. S. (1992). *A computational analysis of young children's lexicons*. Unpublished doctoral dissertation, Indiana University, Bloomington, Indiana.

LUCE, P. A., & PISONI, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, **19**, 1–36.

MACKEN, M. A. (1980). The child's lexical representation: The "puzzle–puddle–pickle" evidence. *Journal of Linguistics*, **16**, 1–17.

MACWHINNEY, B. (1995). *The CHILDES project: Tools for analyzing talk* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

METSALA, J. L. (1997a). Spoken word recognition in reading disabled children. *Journal of Educational Psychology*, **89**, 159–169.

METSALA, J. L. (1997b). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory and Cognition*, **25**, 47–56.

METSALA, J. L., & WALLEY, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, NJ: LEA.

MILLS, D., COFFEY-CORINA, S., & NEVILLE, H. (1993). Language acquisition and cerebral specialization in 20-month-old infants. *Journal of Cognitive Neuroscience*, **5**, 317–334.

NEAREY, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*, **18**, 347–373.

NITTROUER, S., & MILLER, M. E. (1997). Predicting developmental shifts in perceptual weighting schemes. *Journal of the Acoustical Society of America*, **101**, 2253–2266.

OHDE, R. N., & HALEY, K. L. (1997). Stop-consonant and vowel perception in three- and four-year-old children. *Journal of the Acoustical Society of America*, **102**, 3711–3722.

PARNELL, M. M., & AMERMAN, J. D. (1978). Maturational influences on perception of coarticulatory effects. *Journal of Speech and Hearing Research*, **21**, 682–701.

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

POLKA, L., COLANTONIO, C., & SUNDARA, M. (2001). A cross-language comparison of /d/–/ð/ perception: Evidence for a new developmental pattern. *Journal of the Acoustical Society of America*, **109**, 2190–2201.

POLKA, L., & WERKER, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 421–435.

RICE, K., & AVERY, P. (1995). Variability in a deterministic model of language acquisition: A theory of segmental elaboration. In J. Archibald (Ed.), *Phonological acquisition and phonological theory* (pp. 23–42). Hillsdale, NJ: Lawrence Erlbaum.

SEDIVY, J. C., TANENHAUS, M. K., CHAMBERS, C. G., & CARLSON, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, **71**, 109–147.

SHVACHKIN, N. K. (1948/1973). 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.

SMITH, B. L., MACALUSO, C., & BROWN-SWEENEY, S. (1991). Phonological effects shown by normal adult speakers learning new words: Implications for phonological development. *Applied Psycholinguistics*, **12**, 281–298.

SMITH, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge, U.K.: Cambridge University Press.

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

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.

STEVENS, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, **111**, 1872–1891.

STORKEL, H. L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research*, **44**, 1321–1337.

STORKEL, H. L. (2002). Restructuring of similarity neighborhoods in the developing mental lexicon. *Journal of Child Language*, **29**, 251–274.

SWINGLEY, D. (1999). Conditional probability and word discovery: A corpus analysis of speech to infants. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st annual conference of the cognitive science society* (pp. 724–729). Mahwah, NJ: LEA.

SWINGLEY, D. (under review). Statistical clustering and the contents of the infant vocabulary.

SWINGLEY, D. (2002b). On the phonological encoding of novel words by one-year-olds. Paper presented at the 27th annual Boston University conference on language development. Boston.

SWINGLEY, D., & ASLIN, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, **76**, 147–166.

SWINGLEY, D., & ASLIN, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, **13**, 480–484.

SWINGLEY, D., & FERNALD, A. (2002). Recognition of words referring to present and absent objects by 24-month-olds. *Journal of Memory and Language*, **46**, 39–56.

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.

TANENHAUS, M. K., MAGNUSON, J. S., DAHAN, D., & CHAMBERS, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, **29**, 557–580.

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.

TRUESWELL, J. C., SEKERINA, I., HILL, N. M., & LOGRIP, M. L. (1999). The kinder-garten-path effect: Studying on-line sentence processing in young children. *Cognition*, **73**, 89–134.

VIHMAN, M. M. (1996). *Phonological development*. Oxford: Blackwell.

VITEVITCH, M. S., & LUCE, P. A. (1998). When words compete: Lexical processing in perception of spoken words. *Psychological Science*, **9**, 325–329.

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., & FLEGE, J. E. (1999). Effect of lexical status on children's and adults' perception of native and non-native vowels. *Journal of Phonetics*, **27**, 307–332.

WARREN, P., & MARSLEN-WILSON, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. *Perception and Psychophysics*, **41**, 262–275.

WARREN, P., & MARSLEN-WILSON, W. (1988). Cues to lexical choice: Discriminating place and voice. *Perception and Psychophysics*, **43**, 21–30.

WATERSON, N. (1971). Child phonology: A prosodic view. *Journal of Linguistics*, **7**, 179–211.

WEIJER, J., Van De (1998). *Language input for word discovery*. Unpublished doctoral dissertation, MPI Series in Psycholinguistics 9.

WERKER, J. F., FENNELL, C. T., CORCORAN, K. M., & STAGER, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, **3**, 1–30.

WERKER, J. F., & TEES, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, **7**, 49–63.