

Sub-segmental detail in early lexical representations

Katherine S. White*, James L. Morgan

Department of Cognitive and Linguistic Sciences, Brown University, 190 Thayer Street, Providence, RI 02912, USA

Received 21 October 2007; revision received 22 February 2008

Available online 9 May 2008

Abstract

In previous studies of phonological sensitivity, toddlers have failed to differentiate mispronunciations of varying severity. We provide evidence of more sophisticated phonological knowledge. Nineteen-month-olds were presented with displays consisting of one familiar and one unfamiliar object. In Experiment 1, names of familiar objects were pronounced correctly or had onset mispronunciations of one, two, or three phonetic features. Under these referential conditions, subjects demonstrated linearly graded sensitivity to the degree of mismatch. In Experiment 2, mispronunciations involved one-feature place, voice, or manner changes; in Experiment 3, subjects heard three types of two-feature onset mispronunciations. Within each of these two experiments, responses were similar to the three types of mispronunciations. Moreover, the pattern across Experiments 2 and 3 confirmed the graded sensitivity observed in Experiment 1. These results converge to suggest that 19-month-olds' representations of familiar words are quite mature and that lexical processing in toddlers (as in adults) is affected by sub-segmental detail.

© 2008 Published by Elsevier Inc.

Keywords: Phonological sensitivity; Word learning; Phonetic features; Graded sensitivity; Lexical access

Introduction

Humans possess prodigious abilities for recognizing spoken words in familiar languages. Across enormous ranges of different words, different talkers, and different contexts, we recognize words with very high accuracy, seemingly instantaneously and effortlessly. This simplicity is, however, illusory, as attempting to find words in an unfamiliar language makes obvious; psycholinguistic research has explicated some of the complexities involved. Chief among these are that there are no universal cues marking boundaries of words in continuous

speech, that no two word tokens are ever identical, and that speech is fleeting, so that there is a great premium on efficient processing. Another difficulty is that we often encounter new words; for young language learners, this must occur many times each day. If exemplars of known words vary, how are we to distinguish new tokens of old words from novel words?

Instances of words differ acoustically along many dimensions. Ultimately, however, only phonemic contrasts (e.g., /p/ vs. /b/) are important for establishing lexical identity (e.g., the difference between *pat* and *bat*). Experience with the phonology of their native language allows adults to process the words they hear efficiently, enabling them to focus on the linguistic dimensions critical for distinguishing among potential candidates in the lexicon. Moreover, phonological knowledge alerts them to the possibility that a new lexical entry should be cre-

* Corresponding author. Address: Department of Brain & Cognitive Sciences, University of Rochester, Meliora Hall, Rochester, NY 14627, USA.

E-mail address: kwhite@bcs.rochester.edu (K.S. White).

ated, if a token differs sufficiently along phonological dimensions and fails to be recognized as an existing word. Thus, both word processing and word learning are intimately dependent on phonological knowledge.

Phonological knowledge is of particular importance to infant learners who are faced with the daunting task of building a lexicon from the ground up. This complex task requires, most fundamentally, that infants create stable representations of words in memory, so that they can learn the mappings between words and their referents, grammatical categories, and pragmatic properties. Early knowledge of language-specific phonological structure has the potential to make learning more efficient, by allowing listeners to background dimensions of speech that vary orthogonally with lexical identity (e.g., talker gender or affect, and, for some languages, vowel duration or tone) and focus on critical dimensions of contrast—dimensions that must be encoded to differentiate lexical representations.

In this article, we demonstrate that, contrary to previous suggestions, toddlers as young as 19 months have highly detailed and apparently adult-like lexical representations. Under appropriate conditions, 19-month-olds display a sensitivity to the degree of phonological mismatch that shows striking parallels with the graded sensitivity observed in adults' lexical processing. These patterns suggest that the architecture underlying lexical representation and processing is constant over development. In addition, these patterns can provide insight into the role of sub-segmental detail in the organization of the early lexicon.

Languages differ considerably, both in the sound distinctions employed to convey meaning and in the organization of these sounds within and across words. Phonological knowledge is clearly experience-dependent. It is likely that infants are born with perceptual sensitivities and/or biases that can be used in the service of early speech perception; however, infants cannot begin life with knowledge about the phonology of particular languages. Previous research has demonstrated that very young infants are sensitive to the major dimensions of phonological variation that define phonetic categories (what we will refer to as “features”, in keeping with the psycholinguistic literature), including voicing, manner and place (Eimas, 1974; Eimas & Miller, 1980; Eimas, Siqueland, Jusczyk, & Vigorito, 1971; Miller & Eimas, 1983). However, they are also highly sensitive to non-linguistic variation—including changes in talker gender and affect, stress, and pitch—during non-referential word recognition tasks (Bortfeld & Morgan, submitted for publication; Houston & Jusczyk, 2000; Singh, Morgan, & White, 2004; Singh, White, & Morgan, 2008). That is, young infants appear to assign equivalent weights to linguistic and non-linguistic sources of variation. By the end of the first year, however, infants' sensitivities have been refined to reflect the phonological structure of the ambient language: they

are less sensitive to non-native phonetic contrasts (Anderson, Morgan, & White, 2003; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992; Werker & Tees, 1984) and weigh non-linguistic acoustic variation less heavily in word recognition tasks (Houston & Jusczyk, 2000; Singh et al., 2004).

Despite the strides infants make during the first year, there has been a great deal of debate concerning whether young learners are sensitive to phonological detail when tasks of word processing and word learning require attention to meaning. Several early studies reported phoneme discrimination failures by children as old as eight years across a variety of tasks, including picture selection and phonological similarity judgment (Barton, 1976, 1980; Eilers & Oller, 1976; Garnica, 1973; Kay-Raining Bird & Chapman, 1998; Schwachkin, 1973). However, the processing demands on children in many of these studies were high—discrimination failures could have been due to any number of factors. For example, a probe–target similarity judgment task requires the child to perceive the phonological detail in the target, to encode and rehearse this information over time, to remember the task instructions, to compare the probe representation to a representation of the target held in memory, to determine on what basis to make a decision, and to make a response while inhibiting other possible responses.

Yet even when processing demands are reduced, learners sometimes still fail to demonstrate sensitivity to phonological detail (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). When tested in a habituation task on a novel object–novel label pairing, 14-month-olds fail to notice a minimal phonetic change in the label. However, in the same task, when habituated to an already familiar pairing, 14-month-olds successfully detect the change (Fennell & Werker, 2003). These results indicate that familiarity with particular lexical items or with object–label pairings influences learners' phonological sensitivity (though see Ballem & Plunkett (2005) for evidence that toddlers show some sensitivity to phonological changes in recently learned words).

When tested in tasks with minimal processing demands—like habituation and preferential looking—on highly familiar words, infants and young children demonstrate a higher degree of competence: they represent familiar words in considerable phonological detail and use this detail during word recognition. For example, if toddlers are presented with pictures of two familiar objects and asked to look at one of them (e.g., “where's the ball?”), they look less to the target object when its name is mispronounced (“where's the gall?”) than when it is pronounced correctly (Swingley & Aslin, 2000). In these tasks, toddlers in the second year of life have demonstrated sensitivity to even one-feature onset mispronunciations of familiar words (Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Swingley & Aslin, 2000, 2002).

Despite such fine-grained sensitivity, in some respects toddlers' performance still appears immature in these tasks. In particular, in several studies, they have responded similarly to multiple-feature and single-feature onset mispronunciations (Bailey & Plunkett, 2002; Swingle & Aslin, 2002). For example, subjects as young as 14 months and as old as 23 months do not look any less at a ball when the label is pronounced as *shawl* (a three-feature deviation) than when it is pronounced as *gall* (a one-feature deviation), although in both cases they look less than when the label is pronounced correctly. These findings suggest that multiple-feature deviations involving a single segment are no more disruptive to young learners' lexical access than are one-feature mispronunciations. This contrasts strikingly with reports that adults are affected by the degree of acoustic-phonetic mismatch during word recognition and in semantic priming tasks (Connine, Titone, Deelman, & Blasko, 1997; Milberg, Blumstein, & Dworetzky, 1988). For example, adults show significantly less priming for the word *dog* when preceded by the non-word prime *gat* than by the prime *cat* (a one-feature mispronunciation), and still less for the non-word prime *wat* (a three-feature mispronunciation).

One interpretation of these findings with toddlers is that the lack of graded sensitivity is the result of a ceiling effect. As Swingle and Aslin (2002) state: "we suspect that our null result on this factor reflects the fact that the effects of phonetic distance on word recognition are not linearly proportional to the number of phonetic features (or any other a priori phonetic-distance metric) incorrectly realized in a pronunciation. That is, under clear listening conditions, any mispronunciation may impair recognizability. Phonetic distance should have an impact, but variation due to distance may not be as influential as the mere fact of being a mispronunciation." (p. 482) On this view, the absence of distance effects in Bailey and Plunkett (2002) and Swingle and Aslin (2002) is uninformative about the nature of phonetic detail in toddlers' lexical representations.

However, the failure to distinguish single-segment mispronunciations of varying severity in these studies is also consistent with views that lexical representations become increasingly specified over development. It is possible that toddlers do not have fully specified representations and in these studies detected mispronunciations using a metric that was not based on sub-segmental detail. To understand this point, consider the fact that toddlers could have responded in the same way in these studies if they were simply counting the number of mismatching segments (e.g., both *shawl* and *gall* differ from *ball* by one segment). If toddlers encode words by including only rudimentary information about segments, then any difference in a segment (whether it involves a single feature, or multiple features) might have the same effect on their responses. Thus, previous

results are consistent with the possibility that early representations do not contain enough detail about segments to enable finer differentiations (e.g., in the degree of mismatch on a single segment).

Holistic theories posit that early lexical representations include only the information necessary to distinguish one word from another in the lexicon (Charles-Luce & Luce, 1990, 1995; Storkel, 2002; Walley, 1993, 2005). As a result, in a small lexicon, little detail needs to be represented. A strong version of holistic view therefore predicts that children should only be sensitive to very large mispronunciations, perhaps involving the deletion or alteration of multiple segments (see Hallé & de Boysson-Bardies (1996) for data that are consistent with this prediction). Another formulation of this view is that early representations, although organized around segments, are underspecified (Brown & Matthews, 1997), and that phonological features are added to representations only after exposure to minimal pairs that are contrastive along those dimensions. Thus, toddlers may fail to show sensitivity to certain changes because these contrasts have not yet been acquired. On these views, acquisition of new words creates pressure to specify and organize lexical items, in order to avoid confusion. Representation of phonetic features thus emerges over development as a consequence of increasing experience with words. Hence, children with larger vocabularies are expected to exhibit greater sensitivity to the degree of featural mismatch on a single segment, as indeed they do by three years of age (Gerken, Murphy, & Aslin, 1995).

The sensitivity to single-segment changes in previous studies (Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Swingle & Aslin, 2000, 2002) appears difficult to accommodate on the strong version of the holistic view above. However, the failure to observe graded effects in these studies is consistent with a weaker version of the holistic view, in which there is still a developmental shift in the specificity of lexical representations. In other words, this work is consistent with the possibility that toddlers encode words by including only rudimentary information about segments and can therefore not make fine differentiations based on the degree of mismatch on a single segment. Also consistent with these findings is the possibility that toddlers *do* encode specific detail about segments in their lexical representations, but use an immature metric for computing similarity during processing which is based on the number of matching segments between the representation and the heard token. In the present work, we explore an alternative hypothesis, namely that the typical method of testing—in which toddlers are presented with only familiar objects with known labels—has contributed to an underestimation of early phonological sophistication and that representations contain detailed phonological information from early in development. This typical method of

testing may depart from learners' everyday experiences in a fashion that works against finding fine-grained phonological sensitivities.

In real-world situations, infants and toddlers likely know labels for some, but not all, of the objects in their environment. Therefore, when a learner hears a token that is phonologically similar to a known word, this could be interpreted either as a novel word that should be mapped onto a novel referent or as a mispronunciation of the known word. Seventeen-month-olds, and possibly younger infants, can map novel labels that are phonologically dissimilar to known labels onto novel objects (Halberda, 2003; Markman, Wasow, & Hansen, 2003; Pruden, Hirsh-Pasek, Golinkoff, & Hennon, 2006). For example, when shown a ball and a novel object, 17-month-olds can map the label *dax* (which is phonologically dissimilar to *ball*) onto the novel object. Thus, a referential context that includes both familiar and novel objects might be more natural for assessing the interpretation of new phonological forms. In this situation, the degree of phonological similarity to the known word might influence learners' processing and interpretation of novel labels. Research with older toddlers and pre-schoolers has demonstrated that in a context containing familiar and novel objects, the tendency to choose a familiar object as the referent of a new label is affected by that label's phonological similarity to the known label (Jarvis, Merriman, Barnett, Hanba, & Van Haitsma, 2004).

As noted, the standard paradigm employed in mispronunciation studies with young toddlers has been to present visual displays containing two *familiar* objects (with known labels) and to name (either correctly or incorrectly) one of the objects. Given the dissimilarity of the competitor object's known label, it is most natural for subjects in this referential context to interpret an incorrect label as a mispronunciation of the target, because there is no viable alternative (unknown) object available. For example, consider a subject who is presented with a visual display containing a ball and a car. The novel tokens *gall*, *zall*, and *shawl* are all unsuitable labels for the car. Thus, in addition to being attracted (perhaps differentially) to the lexical entry for *ball*, all of these pronunciations may be equally repelled from the lexical entry for *car*. Therefore, in this case we might expect the labels to be interpreted as mispronunciations of *ball*, regardless of subtle differences in phonological similarity. In other words, this type of referential context may render it difficult to observe graded sensitivity to mispronunciations of varying severity because the high dissimilarity between the referent labels may overwhelm the smaller differences among mispronunciations.

In the present studies, we assessed 19-month-olds' sensitivity to mispronunciations in a referential context in which the competitor object was unfamiliar. We predicted that the availability of a viable referent for the

mispronounced label would make this a more sensitive referential context in which to address the question of graded sensitivity. Furthermore, this type of referential context allows us to determine how tolerant toddlers' lexical processing system is to phonological mismatch. Because learners successfully map novel words onto unfamiliar objects by 17 months, we hypothesized that more severe mispronunciations might be interpreted as labels for the unfamiliar object in the display.

Across three experiments we sought evidence for the representation of sub-segmental detail by young learners. In all three studies, 19-month-olds were presented with pairs of familiar and unfamiliar objects using the Intermodal Preferential Looking Procedure; labels were either correctly pronounced or had onset mispronunciations. In Experiment 1, we varied the degree of the mispronunciation. We predicted that if they are sensitive to the degree of phonological mismatch, 19-month-olds should display graded sensitivity, directing their attention to the familiar object less as the severity of the mispronunciation increased. In Experiments 2 and 3, we explored this effect of phonological distance further: in Experiment 2, the name of the familiar object was either pronounced correctly or contained one of three different types of *one-feature* onset mispronunciations (place, voice, or manner); in Experiment 3, three different types of *two-feature* onset mispronunciations were used. These manipulations allowed us to determine whether patterns observed in Experiment 1 generalize beyond the particular set of changes employed in that experiment. Further, if graded similarity plays a role in toddlers' lexical access, then comparison across Experiments 2 and 3 should reveal greater effects of the more deviant mispronunciations of Experiment 3.

Experiment 1

This experiment was designed to establish whether 19-month-olds exhibit sensitivity to varying degrees of phonological mismatch when the visual display contains a familiar object paired with an unfamiliar object. The experimental session consisted of 18 trials, each of which involved a unique familiar object–novel object pair. In this study we included one-feature, two-feature, and three-feature mispronunciations, a greater range than has been systematically studied previously. Bailey and Plunkett (2002) studied one- and two-feature changes, as did Pater, Stager, and Werker (2004), using a different methodology. Neither found differences in toddlers' performance as a function of distance. Mani and Plunkett (2007) did not manipulate distance explicitly, but included vowel mispronunciations of different sizes. Other studies (Swingley & Aslin, 2000, 2002) have included a range of mispronunciations, from single-feature changes to full segment deletions or insertions,

involving both vowels and consonants. If toddlers exhibit graded sensitivity to the degree of phonological mismatch, we expect them to show decreases in looking to the familiar object as the distance between the mispronunciation and the correct pronunciation increases.

Subjects

Forty-one 19-month-olds were tested using the Intermodal Preferential Looking Procedure (described below). Ten subjects did not complete enough trials for analysis due to fussiness or disinterest in the stimuli (6), experimental error (2), and failure to look at both objects during the salience phase (2). The data from three additional subjects were removed because of large negative (< -1.5) difference scores in the correct condition. The data from these subjects were removed because the procedure was predicated on the assumption that children had familiarity with the referents of the words included as familiar stimuli; toddlers with low scores in the correct condition did not fit this assumption, either because they did not know (some of) the words or because they were not engaged in the task. Although this may result in some inflation of the correct condition mean, the comparisons of the mispronunciation conditions to chance, to each other, and to the novel condition are all unaffected. Because we were primarily interested in the effects of different types of mispronunciations (one, two, three features), we adopted this approach.¹ This left 28 subjects distributed across four counterbalancing groups (17 females and 11 males, mean age = 575 days). Seven subjects were assigned to each of the four counterbalancing groups.

Stimuli

The familiar stimuli comprised a set of words that are comprehended by the majority (>50%) of infants by 14 months, according to parental report norms (Dale & Fenson, 1996). Thus, on average, the majority of our 19-month-olds would have been familiar with the familiar words for at least five months prior to the experimental session. To ensure that the visual stimuli selected to depict these words were recognizable to 19-month-olds, a pilot study was conducted with a separate group of twenty 19-month-olds. During the pilot, also conducted using the Intermodal Preferential Looking Procedure, subjects were presented with visual displays consisting of two familiar objects and asked to look at one. Each trial involved a unique pair of objects, presented first in silence to establish baseline looking prefer-

ences and followed by the labeling of one of the objects. All 20 subjects saw the same 12 object pairs. Ten subjects were tested on their recognition of 12 items; the other 10 subjects were tested on their recognition of the paired 12 items. Thus, 24 familiar words were piloted. Eighteen items were retained for use in the current experiment; for all of these items, pilot subjects increased their looking to the target object following mention of the object's label. The unfamiliar items selected for the study were real objects, similar in visual complexity and category status (e.g., artifacts, living things) to the known-label objects. In some cases, colors were altered to maximize their novel appearance. With the exception of *pickle*, the names for the unfamiliar objects are not included on lists of familiar words on either the infant or toddler versions of the MacArthur CDI (Dale & Fenson, 1996). An example stimulus pair is depicted in Fig. 1; a list of displayed objects is given in the Appendix.

Auditory stimulus conditions

Of 18 total trials, nine trials involved correct labeling of one object and nine involved a mispronunciation of the familiar object's label. Each subject received five trials in which the familiar object's label was pronounced correctly, three trials in which the familiar object's label was pronounced with a one-feature change in the onset consonant, three trials in which the familiar object's label was pronounced with a two-feature change in the onset consonant, three trials in which the familiar object's label was pronounced with a three-feature change in the onset consonant, and four novel trials in which the unfamiliar object was named. Novel labels were phonologically distinct from familiar labels. Novel trials were included so that labels were not always phonologically similar to the name of the familiar object. This was done to reduce the possibility that subjects would adopt a strategy of always looking at the familiar object. Our inclusion of an unfamiliar object was motivated by the hypothesis that this would be a more sensitive context in which to evaluate toddlers' knowledge; it was, therefore, important to indicate to them that the unfamiliar object was a possible referent. Additionally, it provided a baseline from which to measure looking behavior in the context of a completely unfamiliar word.



Fig. 1. Sample visual stimulus pair. Actual stimuli were in color.

¹ We note, however, that the pattern of results remained the same when the data from these subjects were included in the analyses.

The decision to present subjects with 18 trials involving unique items differs from other studies of toddlers' sensitivity to mispronunciations: in other studies, each subject typically hears the same small set of words presented in multiple pronunciation conditions. However, it is possible that hearing a mispronunciation of a word after having just heard it pronounced correctly in the same session alters subjects' responses to the mispronunciation. For example, toddlers might be more sensitive to a deviant pronunciation directly following a correct pronunciation, because the differences between the two might be highlighted. Conversely, prior exposure to a correct pronunciation may make it easier for subjects to activate the same lexical representation from a mispronunciation—making subjects appear less sensitive to the mispronunciation. Therefore, in order to eliminate possible effects of repetition, each item was presented only once during the session.

One-feature mispronunciations involved changes in place of articulation, two-feature mispronunciations involved changes in place and voicing, and three-feature mispronunciations involved changes in place, voicing and manner. Examples of mispronunciations are given in Table 1. All mispronunciations resulted in non-words or in words judged unlikely to be familiar to toddlers at this age. All stimuli were naturally produced by a trained female speaker of American English who produced the utterances with positive infant-directed affect. The mean length of target items across all conditions was 734 ms ($SD = 95$ ms); lengths of target items did not differ significantly across conditions, $F(4, 53) = 1.45$, *ns*. Pairings of familiar and unfamiliar objects remained constant across subjects. However, the assignment of these stimulus pairs to pronunciation condition was counterbalanced across subjects, with the exception of six trials (two correct filler and four novel) that were constant across subjects. The order of presentation was pseudo-random with the constraint that the first two trials always contained one correct and one novel trial. A complete list of stimuli and conditions is provided in Appendix A.

Procedure

Testing was conducted in a sound-treated laboratory room. The parent sat with the child on his/her

lap, while listening to instrumental music over noise-cancelation headphones to mask the audio stimuli. Approximately 90 cm in front of the child were two 51 cm television monitors mounted side-by-side, together subtending approximately 55 degrees of visual angle. A speaker was located centrally between the two television monitors behind a pegboard panel. At the subjects' eye level, a blue light was mounted on the panel between the two television monitors. The subjects were monitored over a closed-circuit video system and recorded on a digital camcorder at 30 fps for later off-line coding. Speech stimuli were played at conversation level (70 dB).

Each trial began with the blue light flashing until the subject fixated at midline. At that point, the experimenter turned off the center light and initiated the salience phase. During the salience phase, the two objects were presented simultaneously in the absence of an audio track to establish baseline looking preferences. After 4 s, the two monitors went dark. Piloting revealed an optimal length of 4 s for the salience phase, which gave subjects time to look at both objects. Following a pause of at least 1 s, the experimenter started the center light flashing. Once the toddler had again fixated centrally, the experimenter initiated the test phase. The use of a re-centering period differs from many other studies, in which a continuous (salience + test) trial is used. The motivation for re-centering the subjects was to prevent them from staring fixedly at either the left or right side for the duration of the trial. This modification to the design appears to have been effective: in contrast to the typical procedure, in our data there were no significant contingencies between side of fixation at the end of the salience period and at the beginning of the test period.

During the test phase, the same two visual stimuli were presented simultaneously. The first audio stimulus ("Where's the X?") was synchronized with the onset of presentation of the visual stimuli. Three seconds after the offset of the target word a second audio stimulus was presented ("Find the X!") to maintain subjects' interest. The test phase lasted 9 s in total. Following an interval of at least 1 s, the next trial commenced. Side of presentation of the familiar object was randomized between trials by the customized experimental software, but was consistent across salience and test phases within each trial. The dependent measure was the change in subjects' looking to the familiar object between the (silent) salience phase and the test phase. Of interest was whether looking behavior would differ as a function of mispronunciation condition.

Following the session, the parent completed a questionnaire on his/her toddler's comprehension and production of the stimulus items (familiar and unfamiliar).

Table 1
Sample audio conditions, Experiment 1

Condition	Example target
CORRECT	"SHOE"
1 FEATURE (place)	"FOO"
2 FEATURE (place + voice)	"VOO"
3 FEATURE (place + voice + manner)	"GOO"

Results and discussion

Results from the parental questionnaire indicated that the items we selected were appropriate for this sample of 19-month-olds. On a scale of 1 (not visually familiar, label unknown), 2 (visually familiar, label unknown), 3 (visually familiar, label familiar), and 4 (visually familiar, label highly familiar), items used as familiar words received an average score of 3.8 ($SD = .15$), indicating that they were very familiar to this group of toddlers. In addition, parents indicated that the children were producing a number of the words themselves. Items used as unfamiliar words received an average score of 1.24 ($SD = .23$). On the present scale, this indicated that some of the objects may have been visually familiar to some of the subjects, but that the names for these objects were not. Parents reported that their children were not producing the names of any of the unfamiliar objects.

Looking behavior was coded off-line frame-by-frame (1 frame = 33 ms) using the SuperCoder program (Hollich, 2003). For the salience phase, looking behavior was coded for the 4-s duration of the phase. For the test phase, looking behavior was coded only for the 3 s following the onset of the first occurrence of the target word. This was done in order to include only subjects' initial response to the target word, before the onset of the second audio stimulus. For each phase, the proportion of looking towards each of the objects was computed over the total time the subject spent looking at both objects for that phase. For each subject, a difference score was computed for each condition. This difference score measured the change in looking toward the familiar object after the target was named, and was com-

puted using the following formula, where averages are computed over trials in a single condition:

$$\% \text{Looking(Familiar)}_{\text{Test}} - \% \text{Looking(Familiar)}_{\text{Salience}}$$

Comparison across test and salience phases allowed us to use each stimulus pair as its own control, thereby controlling for any inherent preference for a particular stimulus in each pairing.

As expected, subjects exhibited a general preference for the familiar object in the salience phase. Average looking proportion for familiar objects across all conditions was .57 (95% CI: .54–.59). A single-sample *t*-test comparing mean salience proportions (across conditions) to .5 was significant ($t(27) = 5.51, p < .001$). This replicates a finding by Schafer, Plunkett, and Harris (1999), who showed that 17-month-olds prefer to look at familiar objects with known names, even in the absence of audio input. In that study, subjects preferred to look at objects with known names over other familiar objects; in the present study, the preference held when the competitor was unfamiliar. Because of the importance of establishing a baseline preference, trials in which subjects did not look at both objects during the salience phase were not included in the analysis. Across all subjects, 47 trials were discarded for this reason (approximately 9% of trials). For each subject, conditions that did not contain two usable trials were not analyzed (8 data points total across subjects—approximately 6% of the data).

Test-minus-salience difference scores are depicted in Fig. 2. Three primary analyses were conducted on these difference scores. First, to establish that the inclusion of novel trials was effective, we compared responses on cor-

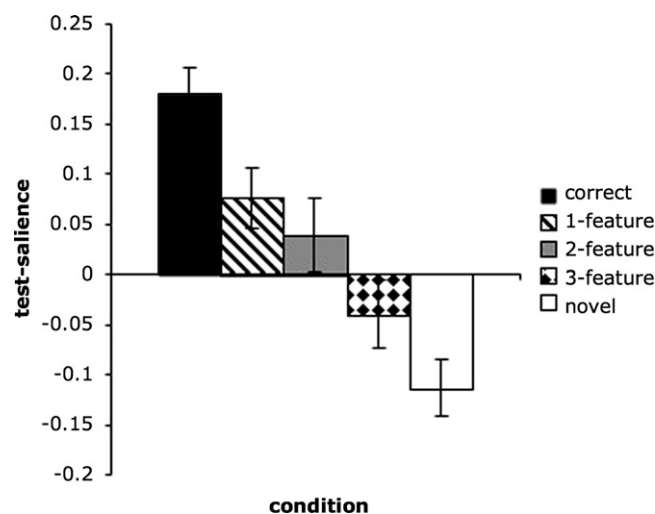


Fig. 2. Proportional looking times and standard errors, Experiment 1. Condition is represented on the *x*-axis. The *y*-axis represents the difference between proportion looking at the familiar object in the test phase and proportion looking at the familiar object in the salience phase.

rect and novel trials. In correct trials, attention to the familiar object increased significantly between salience and test, mean .18 (CI: .13–.23), $t(27) = 7.02$, $p < .001$ (all Confidence Intervals reported are 95%; all t -tests are two-tailed). In novel trials, on the other hand, attention to the familiar object *decreased* significantly, $t(27) = -4.05$, $p < .001$, mean $-.11$ (CI: $-.17$ – $-.06$). A significant condition \times phase interaction (with the two pronunciation conditions and two experimental phases) confirmed that subjects behaved differently in these two conditions, $F(1, 27) = 47.6$, $p < .001$.

Second, to establish whether there was an effect of mispronunciation on responses, subjects and items ANOVAs on conditions 1–4 (excluding novel trials) were conducted. These analyses revealed different patterns of looking across the conditions, significant both by subjects, $F_1(3, 63) = 5.77$, $p < .001$, and by items, $F_2(3, 33) = 5.61$, $p < .005$ (min $F(3, 86) = 2.84$, $p < .05$). Moreover, planned comparisons between the correct condition and each of the three mispronunciation conditions were significant, $t(23) = 2.64$, $p < .05$, mean difference .1 (CI: .02–.17); $t(26) = 3.12$, $p < .005$, mean difference .14 (CI: .05–.24); $t(24) = 4.27$, $p < .001$, mean difference .22 (CI: .11–.32) for one-feature, two-feature and three-feature changes, respectively. Thus, the present group of toddlers, like toddlers in previous studies, was significantly affected by single-segment mispronunciations. Because preferences in the salience phase were similar across conditions, $F(3, 63) = .45$, *ns*, differences in looking behavior can be attributed to effects of the audio stimulus at test.

Third, to explore the effects of mispronunciation degree, a trend analysis on conditions 1–4 (excluding novel trials) was performed. There was a striking linear trend in difference scores, $F_1(1, 21) = 14.13$, $p < .001$, $F_2(1, 11) = 15$, $p < .005$ (min $F(1, 29) = 7.27$, $p < .02$), whereas the quadratic and cubic trends were negligible. The linear trend captured 98% of the between-condition variance in looking behavior. In keeping with this trend, analyses of simple main effects revealed significant increases in looking towards the familiar object in the correct and one-feature conditions, respectively, $t(27) = 7.02$, $p < .001$, mean .18 (CI: .13–.23); $t(23) = 2.52$, $p < .05$, mean .08 (CI: .01–.14). In contrast, there was no change between salience and test in the two-feature condition, $t(26) = 1.06$, $p < .3$, mean .04 (CI: $-.04$ –.11). Finally, there was also a non-significant change between salience and test in the three-feature condition, although the mean difference score in this case was negative, $t(24) = -1.17$, $p < .25$, mean $-.04$ (CI: $-.11$ –.03). The finding that subjects interpreted novel labels as labels for the unfamiliar object (and to some extent, did the same for three-feature mispronunciations) is particularly striking given that each pair of objects was only presented once. In other examples of precocious learning by mutual exclusivity, learners have

been repeatedly exposed to the same object pairs (Halberda, 2003).

This pattern of results reveals that 19-month-olds are in fact sensitive to varying degrees of mispronunciation. Moreover, this sensitivity is not limited to a few lexical items; the linear effect of phonological mismatch on looking behavior was significant by items as well as by subjects. Why did toddlers demonstrate such fine-grained sensitivities in this experiment? Subjects in this study were, on average, five months older than subjects in Swingley and Aslin (2002). Perhaps this type of sensitivity develops over the second year. However, our subjects were, on average, younger than the toddlers in Bailey and Plunkett (2002). Thus, age alone cannot explain the discrepant findings. An alternative explanation is that the graded sensitivity observed here can be attributed to the referential context: a referential context in which both familiar and unfamiliar objects are present is not only more representative of the real world, but also appears to constitute a more sensitive context for assessing learners' phonological knowledge. These results also reveal the limits of 19-month-olds' tolerance for mispronunciations: only one-feature mispronunciations were interpreted as labels for the familiar object.

To our knowledge, this is the first demonstration of graded sensitivity during lexical access in young toddlers. However, it is possible that the linear pattern observed is unique to the particular sequence of changes made. In this experiment, one-feature changes always involved a place of articulation change, two-feature changes always involved the addition of a change in voicing, and three-feature changes always involved the addition of a manner change. Experiments 2 and 3 were designed to determine whether this pattern of sensitivity generalizes to other combinations of features.

Experiment 2

The results of Experiment 1 suggest that lexical access in 19-month-olds is affected by the phonological distance between the heard token and the correct pronunciation. Further, they suggest that one-feature mispronunciations are mapped in this context onto the familiar object, whereas more deviant mispronunciations are not. However, one-feature changes in Experiment 1 always consisted of place changes. The dimensions of place, voicing, and manner involve different sets of acoustic changes (Lieberman & Blumstein, 1988). The acoustic correlates for place include the target formant values and pattern of formant transitions into adjacent vowels, which differ as a function of where the closure occurs in the supralaryngeal vocal tract. The acoustic cues to voicing include amplitude of the release burst and aspiration, duration of the closure, nature of the first formant transition, and vowel length (for conso-

nants in final position). Finally, cues to manner of articulation include the presence and strength of the closure and release burst (for stops and affricates), the presence of noise (for fricatives), the presence of a nasal murmur (for nasals), and the rate of formant transitions (e.g., the formants in approximants change at a slower rate). In addition to these acoustic differences, there is some perceptual evidence from adults to suggest that place contrasts are less perceptible than other types of one-feature contrasts (Miller & Nicely, 1955). Perhaps voice or manner mispronunciations would have had different effects on toddlers' looking behavior. On the other hand, the linear trend in Experiment 1 suggests that the addition of each type of featural change had an equivalent effect on subjects' looking behavior; thus, it is possible that changes along each of the three dimensions are perceptually equivalent. Experiment 2 was designed to address these possibilities by directly comparing the effects of three types of one-feature mispronunciations—place, voice, and manner.

Subjects

Forty-eight 19-month-olds were tested using the Intermodal Preferential Looking Procedure. As in Experiment 1, data from subjects who were missing more than one trial per experimental condition in more than two conditions were discarded. Nine subjects did not complete enough trials for analysis due to fussiness (6), experimenter error (1), and failure to look at both objects during the salience phase (2). The data from three additional subjects were removed because they had large negative (< -0.15) difference scores in the correct condition. This left 36 subjects distributed across three counterbalancing groups (14 females and 22 males, mean age = 581 days), 12 subjects per group.

Stimuli, apparatus, and procedure

The familiar and unfamiliar objects from Experiment 1 were used in Experiment 2. However, in some cases the assignment of items to pronunciation conditions was changed so that all three one-feature mispronunciations would result in non-words or in words unfamiliar to 19-month-olds. For example, in Experiment 1, the item *foot* occurred in both correct and mispronunciation trials; this was not true in Experiment 2 because a manner mispronunciation of *foot* would have resulted in the familiar word *put*. All stimuli for Experiment 2 (including those that remained the same across experiments) were recorded anew by the same trained speaker.

Auditory stimulus conditions

Subjects again completed 18 trials. As in Experiment 1, in nine trials subjects heard a correct labeling of one

object and in nine they heard a mispronunciation of the familiar object's label. Subjects in Experiment 2 received four trials in which the familiar object's label was pronounced correctly, five trials in which the unfamiliar object was named, and nine trials in which the familiar object's label was pronounced with a one-feature change in the onset consonant. Three trials involved a place mispronunciation, three involved a voice mispronunciation, and three involved a manner mispronunciation. Manner mispronunciations always involved changes from stops to fricatives or fricatives to stops.

All mispronunciations resulted in non-words or in words judged unlikely to be familiar to this age group. The mean length of target items across all conditions was 701 ms ($SD = 124$ ms); there were no systematic differences in the lengths of target items across conditions. Pairings of familiar and unfamiliar objects remained constant across subjects. In addition, all subjects received the same correct and novel trials. However, the assignment of stimulus pairs to mispronunciation condition was completely counterbalanced across subjects. A complete list of counterbalancing conditions is provided in Appendix B. The visual displays were presented in the same order for each counterbalancing group; consequently the order in which different types of pronunciations were presented differed across groups. The first two trials always consisted of a correct and a novel trial.

Procedure

The procedure was identical to Experiment 1, with the exception that there was a single audio stimulus per test event ("Where's the X?"), reducing the length of the test phase to 5 s. The dependent measure was the amount of time subjects looked at each object in the absence and presence of the audio stimulus. To maintain consistency with Experiment 1, looking behavior was again analyzed over the duration of the 4-s salience (no audio) phase and for the 3 s following target onset in the test (audio) phase. As in Experiment 1, parents rated their toddlers' familiarity with the stimulus items.

Results and discussion

On a scale of 1 (unfamiliar) to 4 (highly familiar), familiar words received an average score of 3.78 ($SD = .16$) on the parental questionnaire. In addition, parents indicated that the children were producing a number of the words themselves. Unfamiliar words received an average score of 1.22 ($SD = .21$). Again, this indicates that although some of the unfamiliar objects were visually familiar to some of the subjects, the names for these objects were not. These ratings were virtually identical to those obtained in Experiment 1.

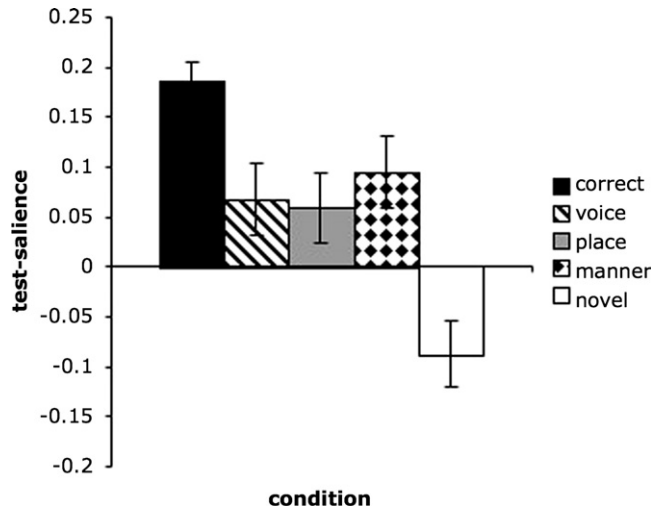


Fig. 3. Proportional looking times and standard errors, Experiment 2. Condition is represented on the x-axis. The y-axis represents the difference between proportion looking at the familiar object in the test phase and proportion looking at the familiar object in the salience phase.

Also, as in Experiment 1, there was a significant bias to look at the familiar object in the absence of an audio stimulus. Average looking proportion in the salience phase for familiar objects across all conditions was 0.56 (CI: .54–.58). A single-sample *t*-test comparing mean salience proportions (across conditions) to chance (0.5) was significant ($t(35) = 5.79, p < .001$).

The analyses of interest were conducted on average test-minus-salience difference scores for each condition (see Fig. 3). As in Experiment 1, trials in which subjects did not establish a baseline preference by looking at both objects during the salience phase were not included in the analysis. Across all 36 subjects, 72 trials were discarded for this reason (approximately 11% of trials). For each subject, conditions that did not contain two usable trials were not analyzed (8 data points total across subjects—approximately 4% of the data).

These difference scores were first used to compare responses on correct and novel trials. In correct trials, attention to the familiar object increased significantly between salience and test, $t(35) = 9.5, p < .001$, mean .18 (CI: .14–.22). In novel trials, on the other hand, attention to the familiar object decreased significantly, $t(35) = -2.65, p < .05$, mean $-.09$ (CI: $(-.15)$ – $(-.02)$). A significant condition \times phase interaction (with both pronunciation conditions and two experimental phases) confirmed that subjects behaved differently in these two conditions, $F(1,35) = 40.6, p < .001$. Thus, 19-month-olds again showed evidence of applying novel labels to novel objects. Also, as in Experiment 1, analyses of simple main effects revealed that subjects showed significant or marginal increases in looking to the familiar object in the correct and in all three one-feature conditions (cor-

rect: $t(35) = 9.5, p < .001$, mean .18 (CI: .14–.22); place: $t(32) = 1.73, p < .094$, mean .06 (CI: $(-.01)$ –.13); voice: $t(31) = 1.88, p < .07$, mean .07 (CI: 0–.14); manner: $t(33) = 2.7, p < .05$, mean .09 (CI: .02–.17)). These analyses reveal that the pattern of increased looks to the familiar object in the one-feature trials of Experiment 1 was not specific to the use of place changes in that experiment, but rather holds for place, voice and manner mispronunciations.

An additional question of interest was whether the three types of one-feature changes would be treated equivalently. A planned contrast between the correct condition and the three one-feature conditions was significant ($F(1,27) = 15.74, p < .001$),² replicating the penalty found for onset mispronunciations in Experiment 1 and elsewhere in the literature. Planned pair-wise comparisons of the three feature conditions yielded non-significant results (voice vs. place, $t(29) = .34, ns$; voice vs. manner, $t(29) = -.73, ns$; place vs. manner, $t(30) = -.56, ns$). This suggests that all three types of mispronunciations had similar effects on subjects' ultimate interpretation of the labels. Of interest, voice and

² Items analyses were not conducted in Experiments 2 and 3. Unlike in Experiment 1, the designs of Experiments 2 and 3 were such that no items occurred in the correct condition as well as in the mispronunciation conditions. Thus, it was impossible to examine the effect of different types of mispronunciations (vs. correct pronunciations) by items. In addition, pairwise comparisons of the mispronunciation conditions by items would have involved only nine items, limiting our power to reject the null hypothesis that all of the mispronunciation conditions had equivalent effects.

place mispronunciations also showed no difference as a function of whether the initial place of articulation was labial or dorsal (place: $t(31) = .2$, *ns*; voice: $t(31) = .6$, *ns*). Therefore, although acoustic correlates for phonetic features differ at different places of articulation, subjects' ultimate treatment of the changes was qualitatively very similar regardless of the places of articulation involved.

However, manner changes appear to have had the least disruptive effect on 19-month-olds' interpretation of the mispronounced labels. In fact, although the correct condition differed significantly from the voice and place conditions (for voice, $t(31) = 3.43$, $p < .005$, mean difference .12 (CI: .05–.19); for place, $t(32) = 2.93$, $p < .01$, mean difference .12 (CI: .04–.2)), it was only marginally different from the manner condition ($t(33) = 1.92$, $p < .06$, mean difference .08 (CI: 0–.17)). The manner condition was also the only type of mispronunciation in which looks to the familiar object increased significantly from salience to test. It is possible that the manner changes we used were difficult to perceive; most of them involved changes from /k/→/h/ and /b/→/v/ (there was no difference for trials involving dorsal (/k/) or labial (/b/) initial words, $t(32) = .55$, *ns*). Miller and Nicely's (1955) analysis of adults' perceptual confusions of onset phonemes revealed that these segments were more likely to be confused with their manner counterparts than their voice or place counterparts (e.g., /b/ and /v/ were highly confusable). Perhaps toddlers and adults are susceptible to the same sorts of confusions. It is possible, and indeed likely, that the effect of a particular change is influenced by the salience of its acoustic cues. An alternative possibility is that the presence vs. absence of phonological neighbors played a role. In early lexicons there are few /v/ initial words (Dale & Fenson, 1996). Perhaps toddlers were more tolerant of mispronunciations with these initial segments as a result.³

The primary finding to emerge from Experiment 2 is that the patterns involving the three mispronunciation conditions were similar to the pattern observed with the one-feature condition of Experiment 1: increased looking to the familiar object, but less so than in the correct condition. In addition, there were no significant differences among the three types of mispronunciations. In Experiment 3, we sought converging evidence for graded sensitivity by presenting subjects with two-feature mispronunciations. If early representations contain detail that allows graded access, then two-feature mispronunciations should have a more disruptive effect on looking behavior; in Experiment 1, subjects did not increase their looking to either object following the two-feature place + voice mispronunciations. If

the pattern observed in Experiment 1 was not due to the particular choice of place + voice changes, then we should observe similar responses for the different types of two-feature mispronunciations presented in Experiment 3.

Experiment 3

In Experiment 3 we presented subjects with three different types of two-feature changes: place + voice, voice + manner, and manner + place. Two predictions can be made on the basis of the previous experiments. First, in accordance with graded sensitivity, the effects of two-feature changes should be greater than the effects of one-feature changes observed in Experiment 2. Second, if the behavior observed for the two-feature changes in Experiment 1 was not specific to the particular combination used (place + voice), different combinations of two-feature changes should all have similar effects on looking behavior.

Subjects

Forty 19-month-olds were tested using the Intermodal Preferential Looking Procedure. As in the two previous experiments, data from subjects who were missing more than one trial per experimental condition in more than two conditions were discarded. Three subjects did not complete enough trials for analysis due to fussiness/crying (2) and failure to look at both objects during the salience phase (1). One infant's data were lost due to experimenter error. In addition, the data from six subjects who had large negative difference scores ($< -.15$) in the correct condition were discarded. This left 30 subjects distributed across three counterbalancing groups (17 females and 13 males, mean age = 584 days), ten subjects per group.

Stimuli, apparatus and procedure

Subjects completed 15 trials. Many of the stimulus items from Experiments 1 and 2 were used in this experiment, including all of the novel items and nine of the familiar items. Six new familiar items were added so that the three types of mispronunciations could be produced, and to ensure that mispronunciations would result in non-words or words unfamiliar to this age group. Four of the added familiar items had been used in the original pilot (see Experiment 1 methods). Although the two remaining familiar items were not piloted, all of the familiar items are reported to be familiar to a majority of infants by 14 months (with the exception of fish, which is reported as familiar by 15 months; Dale & Fenson, 1996). In addition, there were no significant differ-

³ We thank an anonymous reviewer for this suggestion.

ences in performance on correct trials across the three experiments (see below). Each subject received three trials in which the familiar object's label was pronounced correctly, three trials in which the unfamiliar object was named, and nine trials in which the familiar object's label was mispronounced. Three trials involved a place + voice onset mispronunciation, three involved a voice + manner onset mispronunciation, and three involved a manner + place onset mispronunciation.

All mispronunciations resulted in non-words or in words judged unlikely to be familiar to this age group. All auditory stimuli were recorded anew by the same speaker as in Experiments 1 and 2. The mean length of target items across all conditions was 779 ms ($SD = 73$ ms); there were no systematic differences in the lengths of target items across conditions. Pairings of familiar and unfamiliar objects remained constant across subjects. In addition, as in Experiment 2, all subjects received the same correct and novel trials. However, the assignment of stimulus pairs to mispronunciation condition was completely counterbalanced across subjects. A complete list of counterbalancing conditions is provided in Appendix C. The visual displays were presented in the same order for each counterbalancing group; consequently the order in which different types of pronunciations were presented differed across groups. The first two trials always consisted of a correct and a novel trial.

Procedure

The procedure was identical to Experiment 2. Again, the dependent measure was the amount of time subjects looked at each object in the absence and presence of the audio stimulus (for 3 s post target onset). As in Experiments 1 and 2, parents rated their toddler's familiarity with the stimulus items.

Results and discussion

On a scale of 1 (unfamiliar) to 4 (highly familiar), familiar words received an average score of 3.82 ($SD = .23$) on the parental questionnaire. In addition, parents indicated that the children were producing a number of the familiar words themselves. Unfamiliar words received an average score of 1.2 ($SD = .22$). Again, this indicates that although some of the unfamiliar objects might have been visually familiar to some of the subjects, the names for these objects were not. Parents reported that their children were not producing the names of any of the unfamiliar items. Due to experimenter error, data were missing for seven of the 30 experimental items. However, ratings for the remaining items were virtually identical to those obtained in Experiments 1 and 2.

Also, as in Experiments 1 and 2, there was a significant bias to look at the familiar object in the absence of an audio stimulus. Average looking proportion in the salience phase for familiar objects across all conditions was .56 (CI: .54–.58). A single-sample *t*-test comparing mean salience proportions (across conditions) to chance (0.5) was significant ($t(29) = 4.51, p < .001$).

The analyses of interest were conducted on average test-minus-salience difference scores for each condition (see Fig. 4). As before, trials in which subjects did not establish a baseline preference by looking at both objects during the salience phase were not included in the analysis. Across all 30 subjects, 48 trials were discarded for this reason (approximately 11% of trials). For each subject, conditions that did not contain two usable trials were not analyzed (9 data points total across subjects—approximately 6% of the data).

These difference scores were first used to compare responses on correct and novel trials. In correct trials, attention to the familiar object increased significantly between salience and test, $t(29) = 4.04, p < .001$, mean .12 (CI: .06–.18). As before, in novel trials, attention to the familiar object decreased significantly, $t(27) = -2.64, p < .05$, mean $-.07$ (CI: $(-.12)$ – $(-.01)$). A significant condition \times phase interaction (with both pronunciation conditions and two experimental phases) confirmed that subjects behaved differently in these two conditions, $F(1,27) = 18.35, p < .001$.

In contrast to Experiment 2, analyses of simple main effects revealed that 19-month-olds did not alter their looking to the familiar object in the three mispronunciation conditions (place + voice: $t(26) = .31, ns$, mean .01 (CI: $(-.06)$ – $(.08)$); voice + manner: $t(26) = -.34, ns$, mean $-.01$ (CI: $(-.09)$ – $(.06)$); manner + place: $t(28) = -.35, ns$, mean $-.01$ (CI: $(-.08)$ – $(.05)$). This is consistent with the results from the two-feature condition of Experiment 1, in which looking did not change between the salience and test phases.

A planned contrast between the correct condition and the three mispronunciation conditions was conducted. This analysis was significant ($F(1,22) = 12.11, p < .005$) replicating the penalty found for onset mispronunciations in Experiments 1 and 2. To evaluate whether subjects treated the three types of two-feature changes equivalently, planned contrasts between the correct condition and each mispronunciation condition were performed. The correct condition was significantly different from all three types of mispronunciations (place + manner: $t(28) = 3.9, p < .001$, mean difference .13 (CI: .06–.2); place + voice: $t(26) = 2.83, p < .01$, mean difference .11 (CI: .03–.19); voice + manner: $t(26) = 2.62, p < .01$, mean difference .13 (CI: .03–.24)). Pair-wise comparisons of the three two-feature conditions yielded non-significant results (place + voice vs. place + manner: $t(25) = .65, ns$; place + voice vs. voice + manner: $t(23) = .41, ns$; place + manner vs.

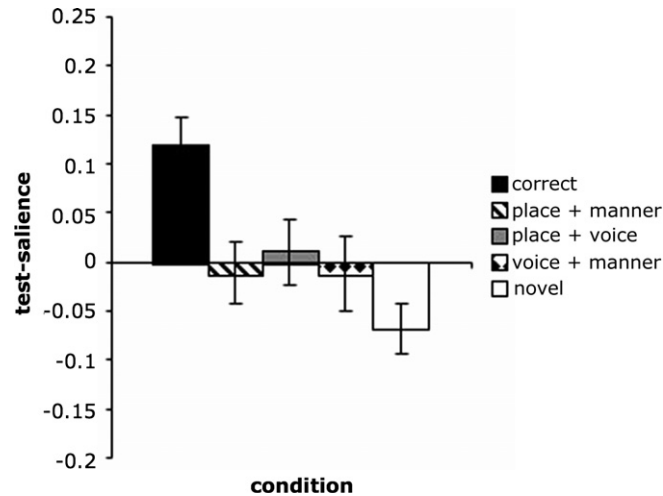


Fig. 4. Proportional looking times and standard errors, Experiment 3. Condition is represented on the x-axis. The y-axis represents the difference between proportion looking at the familiar object in the test phase and proportion looking at the familiar object in the salience phase.

voice + manner: $t(25) = .05$, *ns*). This suggests that all three types of mispronunciations had similar effects on toddlers' ultimate interpretation of the labels.

To evaluate whether the pattern across Experiments 2 and 3 would replicate the graded pattern observed in Experiment 1, a mixed ANOVA was performed with experiment (2 or 3) as a between-subjects factor and pronunciation as a within-subjects factor, where pronunciation included three levels: correct, average mispronunciation, and novel. In this analysis, there were significant effects of pronunciation, $F(1.73, 107)$ Greenhouse–Geisser corrected values) = 34.68, $p < .001$ and experiment, $F(1, 62) = 8.6$, $p < .005$. There was also a marginal interaction of pronunciation and experiment, $F(1.73, 107) = 3.01$, $p < .053$. Given this interaction, one-way ANOVAs were conducted to explore the effect of experiment at each level of pronunciation. Whereas the effect of experiment was marginal or non-significant for correct and novel pronunciations, $F(1, 64) = 3.79$, $p < .06$ and $F(1, 62) = .19$, *ns*, respectively, it was highly significant for mispronunciations, $F(1, 64) = 9.27$, $p < .005$. Average difference scores for the mispronunciation trials of Experiment 2 (.07) were significantly higher than average difference scores on the mispronunciation trials of Experiment 3 (–.01). Recall that the mispronunciations of Experiment 3 involved two-feature changes, whereas the mispronunciations of Experiment 2 were one-feature changes. Thus, the pattern across Experiments 2 and 3 confirms the pattern observed within-subjects in Experiment 1, providing evidence that performance is determined by the degree of mismatch (as we manipulated it, in the number of phonetic features), and is not specific to particular combinations of

features. Thus, this comparison provides additional support for graded lexical access in 19-month-olds.

General discussion

Across three studies, we assessed 19-month-olds' sensitivity to phonological mismatch and their ability to apply this sensitivity during the semantic processing of familiar words. In all three studies, the Intermodal Preferential Looking Procedure was used to present toddlers with pairs of familiar and unfamiliar objects and ask them to look at one of the objects. In Experiment 1, subjects heard the name of the familiar object either correctly pronounced, or mispronounced in onset position by one, two, or three phonetic features. As in previous studies (Bailey & Plunkett, 2002; Swingley & Aslin, 2000), toddlers were sensitive to single-segment mispronunciations: this was manifest as a decrease in looking to a familiar object when its name was mispronounced relative to when its name was pronounced correctly.

However, the present results go further, by demonstrating that when both familiar and unfamiliar objects are presented, 19-month-olds are sensitive to the *degree* of phonological mismatch on a single segment. This was revealed by a highly significant linear trend in Experiment 1: linear increases in the severity of the mispronunciation (in features) mapped almost perfectly onto linear decreases in looking. In Experiments 2 and 3, we explored this sub-segmental sensitivity further, by directly comparing responses to different types of one-feature and two-feature onset mispronunciations,

respectively. While subjects were less likely overall to interpret mispronunciations as labels for the familiar object, there were no significant differences among the three mispronunciation conditions within each experiment. Moreover, the disruption caused by the mispronunciations was greater in Experiment 3 than Experiment 2, in keeping with the linear pattern observed in Experiment 1. The results of these studies converge to demonstrate high sensitivity to the degree of phonological mismatch.

Several findings of interest emerge from this set of studies. We briefly list them here and address each of them in turn. First, despite previous failures to exhibit sensitivity to varying severity of mispronunciations in this type of task (Bailey & Plunkett, 2002; Swingley & Aslin, 2002), toddlers tested in the present, more natural referential situation demonstrated just such sensitivity. In fact, their performance shows strong parallels with adults' behavior during lexical processing. Second, subjects were relatively intolerant of phonological mismatch in this referential situation, mapping some one-feature mispronunciations onto familiar referents, but failing to map two- and three-feature mispronunciations onto familiar referents. Third, the pattern of results across experiments raises interesting questions about the role of phonetic features in the organization of the early lexicon.

The finding that 19-month-olds exhibited graded sensitivity to the degree of phonological deviation contrasts with previous reports that toddlers fail to differentiate between minor and more distant mispronunciations of a single segment (Bailey & Plunkett, 2002; Swingley & Aslin, 2002). These previous results were consistent with the use of a segmental (or larger) metric for representation and processing; for example, in those studies toddlers could have represented only rudimentary information about segments and detected mispronunciations by attending to the number of matching segments. However, toddlers in our study showed distance effects even when distance was manipulated on a single segment, suggesting that 19-month-olds have access to sub-segmental phonological detail. Moreover, during lexical processing they appear to compare heard tokens to stored representations using a similarity metric that is based on this level of detail.

We hypothesize that the discrepancy between our findings and previous findings can be attributed to the referential status of the visual competitors. In previous studies, subjects saw two familiar objects on each trial. The presence of two familiar objects meant that effects of the mispronunciations were necessarily due to computations of similarity between the heard mispronunciations and the lexical representations of the target words alone. This is because the presence of two familiar objects made it unlikely that the competitor object would enter consideration as a potential referent for the mispronounced label. In the present case, when there is a novel object present, mispronunciation effects likely stem from another source as well—the process of deciding whether the mispronounced label refers to the novel referent. We hypothesize that the competitor object's status as a potential referent for the mispronunciation in the current study may have allowed graded effects of phonological mismatch to emerge. Moreover, this type of referential context is more analogous to real-world situations, in which learners do not know the name of every object in their environment and many objects have highly similar names.

One interesting observation that can be made on the basis of this larger set of studies is that learners appear to deploy their phonological sensitivities flexibly as a function of the referential context: in the presence of a novel referent, a phonologically deviant label may be interpreted as a new word; in contrast, when there is no new referent available, toddlers may interpret the same label as a mispronunciation. Swingley and Aslin (2002) found that 14- to 15-month-olds looked at the target object significantly more than chance with both "close" and "far" mispronunciations; in Bailey and Plunkett (2002), 18-month-olds still looked to the target object for two-feature mispronunciations of some words. Note that, in our study, toddlers did not have a general bias to interpret mispronunciations as novel labels in the presence of novel referents. Rather, the interpretation of a mispronunciation in this context was dependent on the degree of mismatch. This is evidenced by the fact that one-feature mispronunciations were mapped onto familiar objects, but the two- and three-feature mispronunciations were not. This type of adaptive sensitivity to referential context (e.g., the presence or absence of novel objects) is ideal for word learning.

These results also have implications for developmental accounts of lexical representation. In the Introduction, we discussed views hypothesizing that young learners' representations become increasingly specified over development as a result of exposure to phonologically similar words. According to holistic theories, young word learners store and organize words using large units, such as the syllable (Metsala & Walley, 1998; Walley, 1993) and it is exposure to phonological neighbors that creates pressure to represent phonological detail. On a strict interpretation, the present results are inconsistent with this view, since the mispronunciations were not familiar neighbors, but rather non-words or words unfamiliar to this age group. More generally, this type of holistic view predicts that the representation of phonological detail should differ for words residing in sparser or denser lexical neighborhoods. The fact that the results of Experiment 1 were generalizable across items as well as subjects argues tentatively against this

view as well, for no attempt was made to control neighborhood size. However, post hoc examination revealed that many of the items used in that experiment may have multiple neighbors even for this age group, so holistic views cannot be categorically ruled out on the basis of these findings.

However, on both the holistic and developmental underspecification (Brown & Matthews, 1997) views, lexical representations continue to undergo significant structural change until well into childhood. Yet even the 19-month-olds in our experiments exhibited behavior resembling adults' in semantic tasks: sensitivity to phonological distance. Under appropriate experimental conditions, toddlers show that they not only encode highly specific detail about each segment, but also that, like adults, they use this detail when computing similarity between heard tokens and representations. Our results, therefore, suggest that the architecture underlying lexical representation and processing is adult-like by 19 months. In addition, they are consistent with the view that phonological structure acquired during the first year (Anderson et al., 2003; Kuhl et al., 1992; Werker & Tees, 1984) is incorporated in lexical representations from the earliest stages of referential word learning, at least for familiar word-forms with known referents. Of course, we cannot rule out the possibility that the representation of this phonological detail emerges with age in the absence of data from earlier points of development.

With respect to the second finding, toddlers' intolerance of phonological mismatch, the results of Experiment 1 indicate that under conditions that approximate the natural world, the lexical processing system of 19-month-olds tolerates only very limited degrees of phonological deviation. Learners are typically exposed to new words in the context of a variety of objects, some of which have known names and some of which do not. In situations where novel referents are sufficiently salient, phonological sensitivity and the presence of learning biases, such as mutual exclusivity (Halberda, 2003; Markman et al., 2003) or novel-name-nameless-category (Golinkoff, Mervis, & Hirsh-Pasek, 1994), may conspire to make learners unwilling to accept mispronunciations as labels of known objects. In Experiment 1, toddlers required a high degree of phonological match (i.e., equal to or less than a one-feature deviation) for novel tokens to be interpreted as existing lexical items. Although subjects increased their attention to the familiar object in the one-feature mispronunciation condition, this was not the case in the two-feature or three-feature conditions. These patterns were replicated in Experiments 2 and 3: subjects again looked more toward the familiar objects with one-feature mispronunciations in Experiment 2, but failed to map the two-fea-

ture mispronunciations onto familiar objects in Experiment 3. And in Experiment 1, 19-month-olds increased their attention to the unknown object with three-feature mispronunciations. Of course, the reality is that all of these types of mispronunciations—even one-feature—*should* be interpreted as novel words. Yet, despite their sensitivity to the one-feature mispronunciations, 19-month-olds still looked significantly more at familiar objects after hearing these mispronunciations. Whether this pattern of performance indicates that their interpretation of phonological changes differs from adults' is unclear; the issue has important implications for lexical acquisition because, even in a small lexicon, many words are phonological neighbors (e.g., *hat*, *cat*, *rat*, *bat*, *pat*; Coady & Aslin, 2003; Dollaghan, 1994). The present results suggest that in a typical referential context (where some objects have familiar labels and others do not), words that differ by a single phonetic feature from known words may be difficult to learn. Swingley and Aslin (2007) provide support for this hypothesis (but see Nazzi (2005) for evidence that toddlers can learn novel words that differ by a single phonetic feature).

With respect to the third finding, toddlers' responses in Experiment 1 varied linearly as a function of our manipulation of phonological distance, in phonetic features. This correspondence suggests that semantic activation in toddlers is a function of phonological distance, and moreover, that this distance is represented in a metric that may be commensurate with the broad features of place, voice, and manner. This is because, although responses were graded in a linear fashion according to the distance in features, the acoustic changes between conditions were not equivalent. One-feature mispronunciations involved a change in place of articulation, two-feature mispronunciations added a change in voicing, and three-feature mispronunciations added a change in manner. The three types of features have very different acoustic correlates, involving different patterns of temporal and spectral change (Lieberman & Blumstein, 1988). Nevertheless, the addition of each type of change caused a similar decrement in subjects' looking towards the familiar object. In addition, the results of the items analysis demonstrated that the pattern was not specific to particular items or onset segments. The results of Experiments 2 and 3, in which looking behavior for the three types of one-feature and two-feature changes, respectively, was similar, are consistent with this pattern. Furthermore, in Experiment 2, the nature of the initial segment (labial or dorsal) did not appear to affect subjects' performance. That is, the effect of a place (or voice) mispronunciation was similar, regardless of the place of articulation of the original segment. However, none of our experiments were explicitly designed to test the effect of a particular change on

different segments and we were limited in our choice of stimuli by the words familiar to this age group. Thus, although these results are consistent with the notion that the early lexicon is organized along dimensions broadly commensurate with phonetic features, they are not conclusive.

In addition, although we characterized our changes in phonological terms, involving featural units, it is unclear from our results whether the relevant metric is phonological or acoustic–phonetic distance. In these studies, the two were confounded. In other words, in Experiment 1, the number of feature changes was confounded with the acoustic–phonetic distance (e.g., two-feature changes involve a more complex set of acoustic changes than one-feature changes). Likewise, the two-feature mispronunciations in Experiment 3 involved more complex differences than the one-feature mispronunciations in Experiment 2. Like the linear trend, the fact that different one-feature mispronunciations in Experiment 2 (and two-feature mispronunciations in Experiment 3) had equivalent effects is consistent with the notion that dissimilarity is determined by the number of phonological dimensions with deviant values. Still, it is possible that these feature changes, despite their different acoustic correlates, have similar perceptual consequences (at least, in the resulting perceived distance from the correct pronunciation). It is also possible that the acoustic–phonetic differences among the one-feature mispronunciations were registered by toddlers, but were not great enough to produce significant differences in our looking proportion measures.

There are two aspects of our data that are consistent with the notion that the acoustic–phonetic differences among the features had different perceptual consequences: first, manner changes (and not place or voice changes) were significantly above baseline and not significantly different from the correct condition in Experiment 2. Second, for trials in which toddlers were not initially looking at the familiar object, they were slower to shift their gaze towards the familiar object for place and voice mispronunciations than for manner changes (which disrupted looks to the familiar object later in the trial). Although the perceptual hierarchy of different feature changes likely depends on the particular segments and changes used (e.g., manner changes here included /b-v/ and /k-h/), these patterns suggest that 19-month-olds are sensitive to the acoustic–phonetic match between the perceived exemplar and stored word types. Thus, the question of whether dissimilarity in toddlers is driven by phonological or acoustic–phonetic distance remains one for future research. One possible avenue would be to use, for example, phonologically equivalent contrasts that are more or less acoustically complex.

For adults, changes along featural dimensions are functionally equivalent—that is, they lead to a change in lexical identity or status. Moreover, meta-linguistic judgments of similarity treat phonetic features as equivalent units. For example, Greenberg and Jenkins (1964) reported on subjects' judgments of the similarity of six stop consonants. Subjects consistently judged pairs differing with respect to a single phonetic feature as more similar than pairs differing with respect to two phonetic features. Bailey and Hahn (2005) recently compared several measures as predictors of phoneme similarity judgments and concluded, "our current best measure of phoneme similarity ... is based on simple counts of the number of major articulatory features—place of articulation, voicing, and manner of articulation ...—in which the two phonemes fail to match" (p. 357). At the same time, acoustic differences among feature classes have perceptual consequences (the confusability of phonemes appears to differ across the classes of place, voice, and manner, though particular rankings appear to vary across testing contexts; Benki, 2003; Cole, Jakimik, & Cooper, 1978; Miller & Nicely, 1955; Warner, Smits, McQueen, & Cutler, 2005). And adults are sensitive to sub-phonetic acoustic information as well (semantic activation in lexical tasks is graded as a function of sub-phonetic distance, e.g., VOT; Andruski, Blumstein, & Burton, 1994; McMurray, Tanenhaus, & Aslin, 2002). In the present studies we demonstrate fine-grained sensitivity below the segmental level in 19-month-olds. However, further research is needed to determine the nature of this sensitivity.

We conclude by returning to our main findings: 19-month-olds represent considerable detail about familiar words, demonstrating graded sensitivity to the degree of phonological match between heard tokens and stored representations. These findings suggest striking parallels in the architecture of the system for lexical representation and processing over development. Moreover, the flexible deployment of this fine phonological sensitivity as a function of referential context provides toddlers with an optimal basis for efficient word learning.

Acknowledgments

This research was supported by NIH Grants F 31 DC 007541-01 to K.S.W. and 5 R01 HD32005 to J.L.M. We thank Lauren Wier for help with stimulus preparation, Lori Rolfe for assistance with data collection, and David Richardson for assistance with data coding. We also thank Melanie Soderstrom, Ruth Tincoff, Katherine Demuth, Dan Swingley, and three anonymous reviewers for helpful feedback on previous versions of this manuscript.

Appendix A. Assignment of stimulus pairs to conditions in Experiment 1

Display		Label			
Fam	Unfam	Group 1	Group 2	Group 3	Group 4
keys	garlic	zeys ³	keys ^C	teys ¹	deys ²
book	paint roller	sook ³	book ^C	dook ¹	took ²
bear	hourglass	gear ¹	tear ²	sear ³	bear ^C
cookie	birdie/shuttlecock	dookie ²	vookie ³	cookie ^C	pookie ¹
foot	bottle opener	foot ^C	soot ¹	zoot ²	goot ³
car	barrel	var ³	car ^C	par ¹	dar ²
ball	pump	gall ¹	kall ²	sall ³	ball ^C
bird	pickle	kird ²	sird ³	bird ^C	gird ¹
bottle	accordion	bottle ^C	gottle ¹	kottle ²	sottle ³
shoe	padlock	foo ¹	voo ²	goo ³	shoe ^C
cup	artichoke	cup ^C	tup ¹	bup ²	vup ³
hand	can opener	zand ²	dand ³	hand ^C	fand ¹
balloon	door knocker	balloon ^F	balloon ^F	balloon ^F	balloon ^F
bunny	French horn	bunny ^F	bunny ^F	bunny ^F	bunny ^F
banana	lantern	lantern ^N	lantern ^N	lantern ^N	lantern ^N
flower	bullhorn	bullhorn ^N	bullhorn ^N	bullhorn ^N	bullhorn ^N
telephone	beehive	hive ^N	hive ^N	hive ^N	hive ^N
chair	trophy	trophy ^N	trophy ^N	trophy ^N	trophy ^N

Superscripts indicate condition type (1 = 1 feature, 2 = 2 feature, 3 = 3 feature, C = correct, F = filler, N = novel). Although mispronunciations are notated using English orthography, only onset consonants were modified. Thus, the 1-feature mispronunciation of *bear* as *gear* maintained the same vowel as the original, despite its resemblance to the English word “gear” (long /i/).

Appendix B. Assignment of stimulus pairs to conditions in Experiment 2

Display		Label		
Fam	Unfam	Group 1	Group 2	Group 3
keys	garlic	heys ^M	geys ^V	teys ^P
book	paint roller	dook ^P	vook ^M	pook ^V
cookie	birdie/shuttlecock	pookie ^P	hookie ^M	gookie ^V
phone	bottle opener	sone ^P	pone ^M	vone ^V
car	barrel	gar ^V	par ^P	har ^M
bird	pickle	vir ^M	pir ^V	gir ^P
bottle	accordion	pottle ^V	gottle ^P	vottle ^M
cup	artichoke	gup ^V	tup ^P	hup ^M
bunny	trophy	vunny ^M	punny ^V	dunny ^P
bear	French horn	bear ^C	bear ^C	bear ^C
ball	door knocker	ball ^C	ball ^C	ball ^C
shoe	padlock	shoe ^C	shoe ^C	shoe ^C
hand	can opener	hand ^C	hand ^C	hand ^C
banana	lantern	lantern ^N	lantern ^N	lantern ^N
flower	bullhorn	bullhorn ^N	bullhorn ^N	bullhorn ^N
foot	beehive	hive ^N	hive ^N	hive ^N
chair	pump	pump ^N	pump ^N	pump ^N
balloon	hourglass	hourglass ^N	hourglass ^N	hourglass ^N

Superscripts indicate condition type (V = voice, P = place, M = manner, C = correct, N = novel).

Appendix C. Assignment of stimulus pairs to conditions in Experiment 3

Display		Label		
Fam	Unfam	Group 1	Group 2	Group 3
car	abacus	car ^C	car ^C	car ^C
balloon	pump	pump ^N	pump ^N	pump ^N
bird	paint roller	zird ^{PM}	kird ^{PV}	fird ^{VM}
bottle	artichoke	kottle ^{PV}	fottle ^{VM}	zottle ^{PM}
hand	accordion	tand ^{PM}	zand ^{PV}	gand ^{VM}
shoe	bullhorn	shoe ^C	shoe ^C	shoe ^C
foot	door knocker	boot ^{VM}	koot ^{PM}	zoot ^{PV}
banana	lantern	lantern ^N	lantern ^N	lantern ^N
sock	barrel	vock ^{PV}	dock ^{VM}	pock ^{PM}
hat	bottle opener	gat ^{VM}	tat ^{PM}	zat ^{PV}
ball	can opener	ball ^C	ball ^C	ball ^C
dog	hourglass	sog ^{VM}	vog ^{PM}	kog ^{PV}
fish	birdie/shuttlecock	zish ^{PV}	bish ^{VM}	tish ^{PM}
door	horseshoe	voor ^{PM}	koor ^{PV}	soor ^{VM}
flower	beehive	hive ^N	hive ^N	hive ^N

Superscripts indicate condition type (PV = place + voice, PM = place + manner, VM = voice + manner, C = correct, N = novel).

References

- Anderson, J. L., Morgan, J. L., & White, K. S. (2003). A statistical basis for speech sound discrimination. *Language and Speech*, 46, 155–182.
- Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, 52, 163–187.
- Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language*, 52, 339–362.
- Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development*, 17, 1265–1282.
- Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language*, 32, 159–173.
- Barton, D. (1976). Phonemic discrimination and the knowledge of words in children under three years. *Papers and Reports on Child Language Development*, 11, 61–68.
- Barton, D. (1980). Phonemic perception in children. In G. Yeni-Komshian, J. Kavanaugh, & C. Ferguson (Eds.), *Child phonology: Perception* (Vol. 2, pp. 97–116). New York: Academic Press.
- Benki, J. R. (2003). Analysis of English nonsense syllable recognition in noise. *Phonetica*, 60, 129–157.
- Bortfeld, H., & Morgan, J. L. (submitted for publication). Early word recognition may be stress-full.
- Brown, C., & Matthews, J. (1997). The role of feature geometry in the development of phonemic contrasts. In S. Hannahs & M. Young-Scholten (Eds.), *Focus on phonological acquisition* (pp. 67–112). Philadelphia: John Benjamins Publishing Company.
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language*, 17, 205–215.
- Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, 22, 727–735.
- Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language*, 30, 441–469.
- Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America*, 64, 44–56.
- Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, 37, 463–480.
- Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28, 125–127.
- Dollaghan, C. A. (1994). Children's phonological neighbourhoods: Half empty or half full? *Journal of Child Language*, 21, 257–271.
- Eilers, R. E., & Oller, D. K. (1976). The role of speech discrimination in developmental sound substitutions. *Journal of Child Language*, 3, 319–329.
- Eimas, P. D. (1974). Auditory and linguistic processing of cues for place of articulation by infants. *Perception and Psychophysics*, 16, 513–521.
- Eimas, P. D., & Miller, J. L. (1980). Discrimination of information for manner of articulation. *Infant Behavior and Development*, 3, 367–375.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, 171, 303–306.
- Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, 46, 245–264.
- 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.
- Golinkoff, R. M., Mervis, C. V., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical

- principles framework. *Journal of Child Language*, 21, 125–155.
- Greenberg, J. H., & Jenkins, J. J. (1964). Studies in the psychological correlates of the sound system of American English. *Word*, 20, 157–177.
- Halberda, J. (2003). The development of a word-learning strategy. *Cognition*, 87, B23–B34.
- Hallé, P. A., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, 19, 463–481.
- Hollich, G. (2003). Available from <http://hincapie.psych.purdue.edu/Splitscreen/index.html>.
- Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1570–1582.
- Jarvis, L. H., Merriman, W. E., Barnett, M., Hanba, J., & Van Hartsma, K. S. (2004). Input that contradicts young children's strategy for mapping novel words affects their phonological and semantic interpretation of other novel words. *JSLHR*, 47, 392–406.
- Kay-Raining Bird, E., & Chapman, R. S. (1998). Partial representations and phonological selectivity in the comprehension of 13- to 16-month olds. *First Language*, 18, 105–127.
- Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, 255, 606–608.
- Lieberman, P., & Blumstein, S. E. (1988). *Speech physiology speech perception and acoustic phonetics. Part of Cambridge series in speech science and communication*. Cambridge, UK: Cambridge University Press.
- Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, 57, 252–272.
- Markman, E. M., Wasow, J. L., & Hansen, M. B. (2003). Use of the mutual exclusivity assumption by young word learners. *Cognitive Psychology*, 47, 241–275.
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86, B33–B42.
- 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: Lawrence Erlbaum Associates.
- Milberg, W., Blumstein, S. E., & Dworetzky, B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society*, 26, 305–308.
- Miller, J. L., & Eimas, P. D. (1983). Studies on the categorization of speech by infants. *Cognition*, 13, 135–165.
- Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some english consonants. *Journal of the Acoustical Society of America*, 27, 338–352.
- Nazzi, T. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. *Cognition*, 98, 13–30.
- Pater, J., Stager, C. L., & Werker, J. F. (2004). The lexical acquisition of phonological contrasts. *Language*, 80, 361–379.
- Pruden, S. M., Hirsh-Pasek, K., Golinkoff, R. M., & Hennon, E. A. (2006). The birth of words: Ten-month-olds learn words through perceptual salience. *Child Development*, 77, 266–280.
- Schafer, G., Plunkett, K., & Harris, P. L. (1999). What's in a name? Lexical knowledge drives infants' visual preferences in the absence of referential input. *Developmental Science*, 2, 187–194.
- Schvachkin, N. K. (1973). The development of phonemic speech perception in early childhood. In C. Ferguson & D. Slobin (Eds.), *Studies of child language development* (pp. 91–127). New York: Holt, Rinehart, and Winston (originally published in 1948).
- Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, 51, 173–189.
- Singh, L., White, K. S., & Morgan, J. L. (2008). Building a phonological lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, 4, 157–178.
- 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.
- Storkel, H. L. (2002). Restructuring of similarity neighborhoods in the developing mental lexicon. *Journal of Child Language*, 29, 251–274.
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76, 147–166.
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13, 480–484.
- Swingle, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, 54, 99–132.
- 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. (2005). Speech perception in childhood. In D. B. Pisoni & R. E. Remez (Eds.), *Handbook of speech perception* (pp. 449–468). Oxford: Blackwell.
- Warner, N., Smits, R., McQueen, J. M., & Cutler, A. (2005). Phonological and statistical effects on timing of speech perception: Insights from a database of Dutch diphone perception. *Speech Communication*, 46, 53–72.
- 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.