Onsets and codas in 1.5-year-olds' word recognition

Daniel Swingley
Department of Psychology, University of Pennsylvania, 3401 Walnut St. 302C, Philadelphia, PA 19104

Abstract

Previous tests of toddlers' phonological knowledge of familiar words using word recognition tasks have examined syllable onsets but not word-final consonants (codas). However, there are good reasons to suppose that children's knowledge of coda consonants might be less complete than their knowledge of onset consonants. To test this hypothesis, the present study examined 14- to 21-month-old children's knowledge of the phonological forms of familiar words by measuring their comprehension of correctly-pronounced and mispronounced instances of those words using a visual fixation task. Mispronunciations substituted onset or coda consonants. Adults were tested in the same task for comparison with children. Children and adults fixated named targets more upon hearing correct pronunciations than upon hearing mispronunciations, whether those mispronunciations involved the word's initial or final consonant. In addition, detailed analysis of the timing of adults' and children's eye movements provided clear evidence for incremental interpretation of the speech signal. Children's responses were slower and less accurate overall, but children and adults showed nearly identical temporal effects of the placement of phonological substitutions. The results demonstrate accurate encoding of consonants even in words children cannot yet say.

Keywords

language acquisition; lexical development; word recognition; phonology; speech perception

Introduction

Comprehension of speech requires knowledge of words' phonological forms and a procedure for matching heard speech to these forms during word recognition. Because words' forms and the phonological categories that compose them differ across languages, children learning to recognize words in their native language face a perceptual problem at two levels: discovering language-specific phonological categories, and determining the phonological forms of specific words.

On one view of language development, language-specific phonological categorization begins in infancy, as infants use their analytical abilities to perform a distributional analysis of the speech signal, splitting continuous speech into segments and forming categories by identifying phonetic clusters of these segments (e.g. Kuhl et al., in press; Werker et al., 2007). Studies of infants' categorization show that over the course of the first year, infants become worse at

differentiating similar nonnative sounds (Bosch & Sebastián-Gallés, 2003; Kuhl et al., 1992; Werker & Tees, 1984) and better at categorizing native sounds (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Kuhl et al., 2006). The fact that infants can perform the computations required for language-specific phonetic category learning and category recognition suggests that when learning words in their second year, children should exercise these talents, and should therefore have lexical representations correctly specified in terms of their language's phonological units. Thus, on this view, infants start by learning phonological categories, and then build words from these categories.

An alternative view is that although infants can learn phonetic categories, they should not be credited with anything like the adult phonological system, and that it is a mistake to identify infants' phonetic categories with adults' phonological categories. Indeed, infants might even start with acoustically detailed knowledge of words, but truly phonological learning might require substantial abstraction away from these acoustic representations. This perspective is not tied to only a single theory, but researchers making a distinction between phonetic category learning (performed by infants) and phonological acquisition (constructing a linguistic system in which discrete categories form a set of contrastive oppositions) tend to emphasize the role of the *lexicon* in determining the phonology: two sounds are separate phonological categories in the language because (for example) they help differentiate words the child knows (e.g. Dresher, 2003; Nittrouer, Studdert-Kennedy, & McGowan, 1989; Pisoni, Lively, & Logan, 1994; Storkel, 2002; Walley, 2005; Walley & Flege, 1999). Hence, infant development in speech-sound categorization need not imply that young children represent words in memory as a sequence of phonological categories. If this is the case, a wide range of alternative descriptions of children's lexical knowledge become plausible.

One such description, maintained by the Lexical Restructuring Model (Metsala & Walley, 1998), holds that children's early representations of words are *holistic* (not formed from discrete sound categories), and also *vague* (missing some of the detailed information that is necessary for differentiating words). According to this hypothesis, categorization of speech into phonological categories, and a concomitant increase in the specificity with which individual words are represented, are caused by word learning itself. As the child's vocabulary expands throughout childhood, the lexicon comes to contain more and more sets of similar-sounding words. Through a process that is not understood, the need to differentiate these similar sounding words leads to "restructuring" of holistic, vague words into segmental, precise representations. Thus, children's knowledge of individual words is contingent on the presence of sets of similar-sounding words in the child's vocabulary. On this view, young children learn words, and then construct their phonology out of similarity relations among words.

The present study is one of a number of related studies that address this debate by testing what children know about how familiar words sound. Most of these studies have assessed children's knowledge by looking for differences between children's responses when words are presented in their canonical phonological forms and when they are presented in deviant forms. If children recognize correct forms (e.g., *dog*) more readily than mispronounced forms (e.g., *tog*), children's lexical representations must contain information that favors the correct form over the deviant one. Once this result is found, of course, one may speculate about the processes that led to this knowledge, and about whether this information is represented in terms of phonological features, probability distributions over phonological feature sets, holistic (but not vague) acoustic objects, and so forth; but ruling out the possibility that children's knowledge is vague or underdetermined is a crucial first step.

Here, we compared the effects of mispronouncing the initial and final consonants of familiar monosyllabic words. Previous studies of one-year-olds' word recognition have only tested onset consonants (Bailey & Plunkett, 2002; Ballem & Plunkett, 2005; Fennell & Werker,

2003; Mills et al., 2004; Swingley & Aslin, 2000, 2002; Swingley, Pinto, & Fernald, 1999) or vowels (Mani & Plunkett, 2007; Ramon-Casas, Bosch, Swingley, & Sebastián-Gallés, under review; Swingley & Aslin, 2000, 2002; though see Nazzi, 2005).[1] For example, Swingley and Aslin (2000) showed children pairs of objects, and named one of the objects using its canonical pronunciation (e.g., *ball*) or a deviant pronunciation (e.g., *gall*). Children looked at the named picture more, and more quickly, when it was correctly pronounced than when it was mispronounced. The effect was present for all six tested words (4 consonant substitutions and 2 vowel substitutions). In a number of subsequent explorations of these phenomena (cited above), children have shown either better recognition of correct pronunciations than mispronunciations, or correct differentiation of familiar words differing only in their initial sound. Similarly, in a task that tests children's ability to use novel word labels to group dissimilar objects into categories, Nazzi and Bertoncini (in press) have found similar sensitivity to phonological distinctions appearing in word onset position and coda position. Despite the unanimity of these findings, though, the additional step of testing final (coda) consonants in a word recognition task is useful, for three reasons.

First, even for adults, codas are more difficult to identify than onsets (Redford & Diehl, 1999). In principle, this may contribute to children's generally lower accuracy rates in producing codas in their own speech (e.g. Bernhardt & Stemberger, 1998; Stoel-Gammon & Dunn, 1985). If any part of a word like *dog* were to be poorly specified in children's mental representations, the coda consonant would be a good bet, and therefore may provide a more compelling test of children's knowledge than the onset consonants typically examined (Bowey & Hirakis, 2006).

Second, if children learn phonological forms accurately because of a need to differentiate similar words in the vocabulary (Metsala & Walley, 1998), then codas should be learned accurately later than offsets. This prediction is based on the fact that English monosyllabic words that sound similar to each other (*phonological neighbors*) most often have matching *rimes* (portions from the final vowel onward), not matching onset–vowel portions (de Cara & Goswami, 2002). Thus, as children's vocabularies grow, they learn more sets like *big–pig–dig* than sets like *hat–ham–have* (or, in fact, *bed–bad–bead*). If the need to distinguish similar words is the driving force behind accurate phonological learning, the predominance of "neighborhoods" like *big–pig–dig* should lead to well-specified onsets. As a result, previous tests of children's knowledge of onsets may overestimate children's phonological knowledge.

Third, comparing children's responses to consonant substitutions at the beginnings and ends of words addresses questions about the time-course of children's interpretation of the speech signal. The psycholinguistic literature consistently supports the notion that in adults, speech interpretation is *incremental*: listeners attempt to understand the speech signal as it unfolds, without waiting for lexical, phrasal, or other cues to trigger interpretation (e.g. Marslen-Wilson, 1975; Dahan & Tanenhaus, 2005). Thus, as a listener hears a word like *turtle* in a context like *I saw a…*, he or she briefly considers the possibility that the word is *turtle, turkey, turban*, or any of a range of words consistent with *I saw a tur…*. Initial interpretations are not necessarily absolute; *hurdle* might also receive some consideration by the recognition system, but recognition of words that do not match the signal at onset is typically quite limited (Allopenna, Magnuson, & Tanenhaus, 1998).

Young children interpret speech incrementally too (Swingley et al., 1999). In Swingley et al. (1999), 24-month-olds viewed pictures of (e.g.) a dog and a doll, or a dog and a tree, and heard a sentence like *Where's the doggie?* Children's eye movements were monitored to determine

---

[1]Swingley (2003) tested both word-initial consonants, and syllable-initial, word-medial consonants. Children demonstrated comparable sensitivities to these mispronunciations.

exactly when they recognized whether they were hearing a name for the picture they were looking at. Children fixating a doll while hearing *doggie* were substantially delayed in shifting their fixation to the dog, relative to children initially fixating a tree. This delay was comparable to that shown in a comparison sample of adults. Swingley et al. argued that the delay came about because the speech signal was temporarily compatible with (e.g.) *doll*, leading listeners to consider the doll a likely referent of the spoken word. No such delay was found when children were shown rhyming pictures (e.g., a ball and a doll) and heard one of them labeled. Thus, both children and adults interpreted words incrementally, using the information available from the onset of the target word to guide their interpretation (see Fernald, Swingley, & Pinto, 2001, for similar results at 18 and 21 months).

The question of incremental interpretation is revisited here in even younger children, in an explicit comparison of the temporal consequences of onset and offset mispronunciations of familiar words. Given that children begin "activating" words based on word-initial phonetic information, it is possible that early interpretation can bias children's perception of a word enough to override, delay, or weaken the effects of later-occurring phonetic information. As noted above, children know relatively few words that overlap at onset (such as *dog* and *doll*), so if word recognition were optimized over the child lexicon, it might be most efficient for children to devote more attention to the beginning parts of words than the ends.

Further, there are indications that even for adults, onsets play a greater role than offsets in word identification, particularly for recently learned words. Studies of adults learning miniature artificial "lexicons" have shown greater confusion among words that overlap from their onsets (such as *baf* and *bav*) than words that match in their rimes (such as *geet* and *keet*; Creel, Aslin, & Tanenhaus, 2006; Creel & Dahan, under review; see also Magnuson, Tanenhaus, Aslin, & Dahan, 2003.) In these studies, responses are generally made well after the entire word has been heard, so the dominance of onset-overlap confusions is not simply a consequence of participants making their response before the stimulus is complete. Children may be subject to the same bias, leading to inferior sensitivity or attention to coda consonants relative to onsets.

Other methods examining differences between earlier and later parts of words (though not always contrasting onsets and codas) have yielded mixed results, though when asymmetries are found, they usually show greater weight being given to earlier phonetic information. For example, the phenomenon of "perceptual restoration," in which listeners perceive an altered speech segment as the original, or a missing segment as if it were present, appears to be somewhat weaker early in words, suggesting listeners' more veridical perception of word onsets, or greater bias introduced by lexical context for offsets (Cole & Jakimik, 1980; Marslen-Wilson & Welsh, 1978; Samuel, 1981). Similarly, tasks in which listeners are explicitly told to detect mispronunciations of words have sometimes revealed better mispronunciation detection in onset position than in other positions, in adults (Cole, Jakimik, & Cooper, 1978) and 4- to 5-year-olds (Cole, 1981; Walley, 1987). However, other studies testing both children and adults have failed to find positional effects in either age group (Bowey & Hirakis, 2006; Walley & Metsala, 1990; see also Walley, 1988).

One previous study has examined onset and offset consonant substitutions in infants (Swingley, 2005a). That study used a variant of the "headturn preference procedure" (Fernald, 1985; Kemler Nelson, Jusczyk, Mandel, Myers, & Turk, 1995), in which infants' fixation to a visual stimulus triggers continued presentation of auditory materials. Duration of fixation is measured as a proxy for infants' "preference" for a particular sort of speech stimulus. Under these conditions 11-month-olds prefer words over nonwords (Hallé & de Boysson-Bardies, 1994). This preference is not found, or is found in only a restricted subset of trials, when mispronounced words are compared with nonwords, suggesting that infants do know how the tested words should sound (Hallé & de Boysson-Bardies, 1996; Vihman, Nakai, DePaolis, &

Hallé, 2004). Swingley (2005) replicated the preference for words over nonwords in a sample of Dutch 11-month-olds, and showed further that this preference disappeared when the initial consonant or final consonant was substituted with a similar consonant. Thus, infants treated mispronounced versions of words as if they were nonwords, both for onset and coda substitutions. In two additional experiments, infants also preferred correct pronunciations over onset mispronunciations; however, infants as a whole did not prefer correct pronunciations over coda mispronunciations. In the latter case the degree of preference for correct pronunciations was significantly correlated with infants' vocabulary size, measured several months later, suggesting that only some infants, perhaps those with better phonological or word-learning skills, represented the coda consonants accurately. Overall, these headturn preference studies show that for at least some words, 11-month-olds encode phonological features of both onsets and codas of familiar words, though the less consistent effects of coda substitutions suggest weaker lexical representations of codas, or weaker effects of deviant pronunciations of codas.

Here, to test 1.5-year-olds' knowledge of coda consonants and to probe for differences in the temporal consequences of word-initial and word-final consonant substitutions, a referential visual fixation task was used (Cooper, 1974; Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Golinkoff, Hirsh-Pasek, Cauley, & Gordon, 1987). Children were shown pairs of pictures on a large computer monitor, and heard sentences naming one of the pictures (e.g., *Where's the dog?*). On some trials, the target word was mispronounced at onset (Experiment 2A; *Where's the bog?*) or at offset (Experiment 2B; *Where's the dawb?*). Children's eye movements to the pictures were analyzed, and target fixation was compared in correct-pronunciation (CP) and mispronunciation (MP) conditions. If children have only vague, diffuse knowledge of onset or coda consonants, mispronunciation should not have a significant impact on recognition; if children know precisely how the words should sound, mispronunciation is expected to hinder recognition. In addition, if children process speech incrementally, effects of onset mispronunciations should be evident earlier than effects of coda mispronunciations.

To enable detailed predictions of the timing of children's responses, the child experiments were preceded by a comparison study of adults, evaluating mature listeners' interpretation of the same stimulus materials. Although the fidelity of the adults' lexical representations is not in doubt, this test permitted measurement of the sizes of onset and offset mispronunciation effects (to test the possibility that coda substitutions might be less consequential than onset substitutions), and examination of the timing of the effects (to test the possibility that the response to coda substitutions might be delayed because of the biasing context provided by the initial consonant and vowel of the word).

Experiments 1A and 1B: adult listeners

Methods

Overview—Adult native English speakers sat facing a large display screen and were presented with 24 trials. On each trial, two pictures of familiar objects were displayed, and then one was named in a prerecorded sentence. On half of the trials, the object label was mispronounced. Participants gaze at the display was recorded and coded off-line to determine the timing of shifts in fixation on correct pronunciation (CP) and mispronunciation (MP) trials. Participants were told that the primary goal of the experiment was to study language in children, and that their task was to listen to the sentences and look at the pictures.

Participants—The 26 participants were students at the University of Pennsylvania, or peers of such students. They were recruited through classes or by word of mouth, and received either some course credit or a modest fee for participation. All identified themselves as native

speakers of American English. Half of the sample (13) were randomly assigned to the onset-MP condition (Experiment 1A) and half to the offset-MP condition (Experiment 1B).

Visual stimuli—The visual stimuli were digitized photographs of objects on a gray background, presented side by side on a 106-cm Panasonic PHD4P plasma screen driven digitally by a computer. The pictures shown on test trials included a book and a dog, a boat and a duck, and a sock and a cup. Pictures were of similar sizes, averaging about 15 cm wide, and separated by about 45 cm. Participants were seated about 2 m from the screen, with the pictures about 20 cm above eye level. Three short animations were also used as fillers. These included a kangaroo bounding across the screen, a triggerfish blowing a bubble, and a butterfly fluttering to a growing flower.

Auditory stimuli—The speech stimuli were digitally recorded by a female native speaker of American English. Her speaking rate was slow and in an "infant-directed" register. The sentence presented on each trial was always *Where's the [target]?*. Mispronunciations (MPs) of the target words always had the same vowel as the correct pronunciations (CPs), so the onset MPs rhymed with the correct targets, and the offset MPs shared the initial consonant and vowel of the CPs. The spoken target words and their onset and offset mispronunciations were: *boat, poat, boad; book, dook, boop; cup, gup, cub; dog, bog, dawb; duck, guck, dut; sock, zock, sog*. In both experiments, mispronunciations of three words involved changing the place of articulation of the relevant consonant, and three involved changing the voicing of the consonant. Offset consonants included their release. The same tokens of the CPs were used for comparison with both types of MP. The mean target-word durations were: CPs, 1100 ms (sd 130); onset MPs, 1120 ms (sd 80); offset MPs, 1010 ms (sd 60). The mean absolute difference in duration between CPs and their MP variants was 79 ms (sd 104).

The temporal locations at which the MPs were discernably different from their canonical pronunciation were estimated via examination of waveforms and spectrograms by a phonetically trained observer. Onset MPs began essentially at word onset (0ms); offset MPs at about 752 ms, though items varied; the /b/ of *cub* began at around 650 ms, the /p/ of *boop* at around 920 ms, and the other offset MPs fell between. According to the trained observer, the first-appearing distinguishing information for each item pair varied. For *boat, boad* it was in the vowel's second formant transition to the coda; for *book, boop* it was the consonantal burst; for *cup, cub* creak in the vowel of *cup*; for *dog, dob* the consonantal closure and release; for *sock, sog* the second formant transition and the release. Of course, these features might not correspond exactly to those driving participants' responses. These MP-onset times correlated strongly (r = .90) with the author's estimates based only on listening to portions of the stimuli without inspecting waveforms and spectrograms, though listening yielded somewhat later estimates.

Apparatus and procedure—Participants were told that they were being tested in an experiment designed for very young children. They were also informed that they would see pictures and hear sentences referring to the pictures.

The experiments were conducted in a dimly lit room containing a 3-sided booth 1.25 m deep, 1.5 m wide, and 2 m tall. The participant sat on a swivel chair in the open end of the booth, facing the display. Speech stimuli were presented through concealed loudspeakers hidden below the display. Participants' faces were videotaped onto digital videocassettes using a remote-controllable pan/tilt camera centered between the loudspeakers.

The experiment consisted of 24 trials. Each trial began with the simultaneous presentation of two pictures on the screen. About 500 ms later, the speech stimulus began. When the sentence

was complete, the pictures remained on the screen for another 1500 ms. Thus, each trial lasted about four seconds.

Two stimulus orders were created for each experiment. The second order inverted the sequential orders of the trials in the first. Trial type (CP and MP) was quasirandomly ordered so that no immediately consecutive trials were of the same type more than twice. Target side was quasirandomly ordered the same way (but not in tandem with trial type). No picture occurred on any two consecutive trials, and each picture appeared equally often on the left and the right, and appeared as the target equally often on the left and the right. Each picture served as the target and the distracter once in each condition, for a total of 4 presentations per trial-order.

Coding—Videotapes of the participants' faces were stamped with a digital timecode and digitized to Quicktime format. Several highly trained coders used George Hollich's *SuperCoder* software to step through each videorecording frame by frame, noting for each frame whether the participant was looking at the left picture, the right picture, or neither (i.e., in transit between pictures, or fixating off-screen). This response information was integrated with trial-timing information generated by tone pulses aligned with stimulus events, yielding an accurate record of the timing of responses to the target words.

Results and discussion

Following previous research (e.g., Swingley & Aslin, 2000), eye movement results were summarized using a proportion-to-target measure. Considering the window of time from 367 ms to 2000 ms from the onset of the target word on each trial, the proportion of time that participants fixated the named target picture (relative to time fixating either the target or distracter) was computed. This temporal window is the same that has been used in studies of toddlers. The 367 ms starting point reflects the fact that eye movements earlier than about this point probably come too fast to be children's responses to the target word, given the time required for programming an eye movement (e.g. Canfield, Smith, Brezsnyak, & Snow, 1997). The usual choice of 2000 ms as the close of the window was motivated by several datasets in which 18–24-month-olds' performance tended to decline throughout the trial, starting at about 1800 ms. Given that adults respond more quickly than children in this task (e.g., Swingley et al, 1999), this initial analysis window could in principle be shifted earlier in time (to, e.g., 200 ms to 1500 ms). Such manipulations do not change the results in any important way, so here the 367 to 2000 ms window is retained for comparison with the children's results in Experiment 2.

In eyetracking experiments, trial time is typically aligned from trial to trial so that "zero" corresponds exactly to the onset of the target word. Recall that in Experiment 1B, mispronunciations did not occur at a fixed time relative to word onset, but rather at whatever point in time the word's final consonant was realized, with some variation due to variation in word length. Aligning trials by target-word onset thus causes the time of the offset mispronunciation to vary from item to item: at 650 ms for *cub*, at 920 ms for *boop*, and so on. To present a clearer picture of the temporal consequences of offset MPs, here all trials were aligned according to the actual temporal location of the offset mispronunciation, measured as described above. This was accomplished by shifting cells in the time-series matrix of fixation responses forward or backward depending on the duration of the word from onset to the beginning of the MP, so that all the trials lined up at the temporal location of the MP. For simplicity, "zero" in analyses and plots still represents the time of the target words' onset, but this is an average over the 6 items, rather than true word onset; and 752 ms, the average temporal location of the offset MPs, represents the *exact* time at which the mispronunciations occurred.

The statistical analyses turn out essentially the same way whether the items are aligned at word onset or at MP onset.

Participants looked at the target picture more than the distracter in both conditions, over the 367 to 2000 ms time window (Expt. 1A, mean CP target fixation 91.1%; MP, 71.6%, Expt. 1B, CP, 91.5%, MP, 78.1%, all $t > 5$, all p (2-tailed) < .0002). The effects of mispronunciation (i.e., CP fixation performance minus MP performance) were significant in both Experiments considered independently (1A, onset MPs, average effect 19.5%, 95% C.I. = 9.5% – 29.6%; 1B, offset MPs, average effect 13.5%, 95% C.I. = 5.1% – 21.8%).

The data from Experiments 1A and 1B were combined in a single analysis of variance (ANOVA) with experiment (onset-MPs, offset-MPs) as a between-subjects factor and condition (CP, MP) as a within-subjects factor. This analysis revealed a significant effect of condition (CP > MP), and no other effect or interaction. Details are reported in the top half of Table 1.

Fixation proportion analyses—The primary purpose of these experiments was to determine the time-course of adults' responses to onset and offset consonant substitutions. This was evaluated using two complementary methods. First, the moment-by-moment proportion of target fixation for each participant in each condition was computed. The averages over subjects are plotted in Figures 1A and 1B. As the graphs indicate, performance on CP trials (solid line) was nearly identical in the two experiments. However, performance on MP trials was quite different. When words were mispronounced at onset, adults were substantially delayed in fixating the target picture; when words were mispronounced at offset, adults first started looking at the target, but then frequently diverged from the target.

To establish the statistical onset of the MP effect in each experiment, the difference between CP and MP target fixation was computed for each subject, for each 33-ms video frame from 0 ms to 2000 ms. In each experiment, the mean of these difference scores at each moment was divided by the across-subjects standard deviation, yielding a measure of effect size for every frame. For the purpose of drawing a discrete threshold for considering the effect to be present at a given moment, the effect size of 0.60 was chosen. This choice is somewhat arbitrary, but may mark a qualitative change: in both experiments, no effect size preceding the first 0.60 ever exceeded 0.50. Other choices produce similar results. In Experiment 1A, the effect size measure first exceeded 0.60 at 367 ms from word onset. By a single-sample t-test, the effect at time 367 (greater target looking on CP trials than MP trials) was 14.0%, significantly above zero ($t = 2.19$, p (2-tailed) = .024). This was the first frame that was significantly above zero by t-test. Each of the 31 following frames also exhibited effect sizes above 0.60 and significant t-test results. We acknowledge that, as a rule, the computation of large numbers of t-tests risks Type I errors. Here, the tests serve only as one of several possible indicators that the sample of adults consistently behaved differently on CP trials than on MP trials. The fact that the first significant frame was always followed by a long, uninterrupted series of additional significant frames offers some protection against Type I error. In addition, *no* frames exhibited significant effects in the reverse direction, with MP performance exceeding CP performance. Thus, by this effect size metric, the window in which onset mispronunciation effects were found extended from 367 to 1400 ms.

In Experiment 1B, the effect size measure first exceeded 0.60 at 1133 ms, or 367 ms following the mispronunciation. The effect at 1133 ms was 6.0%, significantly above zero ($t = 2.49$, p (2-tailed) = .013). This was the first frame that was significantly different from zero by t-test, and was followed by a series of 36 consecutive frames that also exhibited effect sizes above 0.60 and significant t-test outcomes. Thus, the window in which offset mispronunciation effects were found extended from 1133 to 2333 ms (the end of the trial).

By this analysis of target fixation proportions, mispronunciation reduced listeners' target fixation starting at 367 ms from the onset of the mispronunciation: at 367 ms for word-onset mispronunciations, and at 1133 ms (367 ms after the coda mispronunciation) for word-offset mispronunciations. There was no indication that participants were slower to detect offset MPs than onset MPs (relative to the timing of the MP).

Having isolated more precisely the time windows in which the MP effects began and ended for onset and offset MPs, the experiments may be compared for just these restricted windows using the same Anovas reported above (see Table 1, "restricted window"). Here, the offset-MP window was cut short slightly to make the analysis windows the same length for the two experiments (i.e., the 32 frames from 1133 ms to 2167 ms inclusive). This analysis finds once again that the strength of the MP effects in the two experiments was similar: a difference of 26.9% in target looking for onset MPs, and 24.1% for coda MPs. These effects were found in all six items, in both positions.

The analysis also revealed a marginal tendency for performance in the restricted window to be better in Experiment 1B (mean, 83.7%) than Experiment 1A (mean, 76.3%). This effect arose because in Experiment 1B, the restricted analysis window was later in the trial, after participants had heard the target word and had had some time to shift to it already.

Analyses of discrete shifts in fixation—Although proportion-of-fixation statistics can provide sensitive measures of performance in this procedure, there are limits to what they reveal. One problem is the non-independence of the proportions from frame to frame. A gradually accumulating target fixation score (or a condition effect like the MP effects tested here) might appear to emerge at a given moment only by finally surpassing a significance criterion, without actually being qualitatively different at that moment. Another property of proportion-to-target analyses is that they measure summary states rather than discrete events. As a result (for example), proportions might remain flat because listeners do not shift their gaze, or because shifts to the target are balanced by shifts away from it. Furthermore, the proportional measures do not reveal whether mispronunciation effects come largely from shifts away from the *target* upon hearing an mispronunciation of its label, or from shifts away from the *distractor* less often, or later, upon hearing an MP. These are different decisions. Defection from the target suggests that listeners do not consider the MP a reasonable label for the fixated target object. Less-frequent rejection of the distracter upon hearing an MP suggests that listeners do not recognize the MP as a familiar word and are uncertain about whether the novel form can apply to the distracter picture.

Considerations like these have prompted a number of analysis strategies in the eyetracking literature (e.g., Dahan & Gaskell, 2007; Dahan, Tanenhaus, & Salverda, 2007; Fernald et al., 2001; Mirman, Dixon, & Magnuson, in press; Swingley et al., 1999). Here, we present a new analysis of the timing of listeners' refixations. The idea behind the analysis is to compute the likelihood of shifting away from the distracter or target at each moment in time, for each condition. Thus, the analysis concerns discrete events.

Figure 2 shows listeners' changes in fixation across the trial, for each experiment and each condition. Two sorts of refixation are shown: shifts away from the distracter (D→T), and shifts away from the target (T→D). The lines on the graph represent the number of changes in fixation occurring in a given frame t (every 33 ms), divided by the number of trials on which participants were fixating the relevant object on frame $t - 1$. Each line was smoothed using a 10-frame moving average. (Note that smoothing makes the overall pattern of performance easier to visualize, but also smears the results temporally to some degree. This is what explains the extremely early divergence of the D→T and T→D lines, slightly before listeners could have initiated eye movements in response to the signal. In Experiment 1B, early divergence of these

lines was also due to the temporal adjustment of each trial to align the offset MPs; at the time zero plotted on the graph, a few of the 12 test words had already begun.)

Performance on CP trials is plotted in black circles and plus signs. Listeners were much more likely to shift away from the distracter (circles) than away from the target (plus signs) when hearing correct pronunciations. Indeed, by about 850 ms after target-word onset, shifting from the distracter was almost entirely saturated, with fewer than 10 CP trials left in each experiment on which listeners were fixating the distracter picture (and as a result, probabilities are not reliable; this is indicated on the plots by "graying out" the lines where the denominators in the probability computations concerned < 10 trials).

Patterns of shifting on MP trials differed in the two experiments. When words were mispronounced at onset (Experiment 1A), shifts from the distracter (solid line) and from the target (dashed line) did not differ in their likelihood until about 500 ms. When words were mispronounced at offset (Experiment 1B), shifting patterns in the CP and MP conditions were essentially identical until about 900 ms. By that time, listeners were looking at the target on most trials, so the comparison of interest in Experiment 1B concerns the relative probability of shifts away from the target on CP and MP trials.

The statistical reliability of the differences between CP and MP trial performance was evaluated by calculating, for each subject, the likelihood of each kind of refixation (from the distracter vs. from the target) for each trial type (CP, MP). These likelihoods were computed over a series of 200-ms windows starting at the location of the mispronunciation for each item: from 0 to 200 ms; 100 to 300; 200 to 400; and so forth. Windows in which CP and MP shifting probabilities differed consistently from zero over subjects are marked on the graph using horizontal lines near the x axis. The criterion was the effect size measure described previously (greater than 0.60 or less than -.60), which here included windows significant at $p < 0.05$ in paired two-tailed t-tests. (Most of the marked windows were significant the alpha level of 0.01.) All effects reaching significance by this criterion (or indeed more relaxed criteria) were those in which shifts away from the distracter were more likely on CP than MP trials, and in which shifts away from the target were more likely on MP than CP trials.

As indicated in Figure 2, these effects of mispronunciation emerged at about 300 ms for onset MPs (the 200–400 ms window) and at about 1067 ms for coda MPs (200–400 ms after the offset MP), a difference of 767 ms. This difference closely matched the difference obtained in the analysis of proportions, and also corresponded exactly to the predicted difference as estimated from the acoustic characteristics of the stimulus words.

Effects of onset and offset mispronunciations were of similar magnitudes. Prior experiments using eyetracking showed larger competition effects of onset overlap (e.g., looking at a beetle when hearing *beaker*) than of rhyme overlap (looking at a speaker when hearing *beaker*; Allopenna et al., 1998). These studies might be taken to show that word onsets are more important than offsets in recognition, and that as a result we should have expected much larger effects of onset MPs than coda MPs. The latter result was a possible outcome of the present study, but was not found. The reason for this concerns the difference between hearing a deviant pronunciation of an existing word (*tog*), and hearing a pronunciation that is fully consistent with more than one displayed picture (*bea*…). In a neutral context (Dahan & Tanenhaus, 2004), listeners hearing *bea*… have little basis for anticipating one frequent *bea* word over another, and, in experiments, often look at a referent that turns out to be the wrong one. The present study did not offer a suitable referent for such sequences as *gup* or *zock*, which do not match words children knew, and certainly did not correspond completely to either picture. In the terms of the "activation" metaphor of word recognition, the onset mispronunciations should not have strongly activated any words, and thus only weakly activated their CP forms in the

lexicon. Thus, the very large "competitor" effects characteristic of onset-overlapping word sets were not expected. Larger effects of onset or coda mispronunciations could not be ruled out *a priori*, but they would have other causes if found.

In sum, adults' fixations to named target pictures were evaluated. When words were mispronounced at onset, adults looked less at the target pictures than when words were correctly pronounced, a difference that began immediately after the words' onset. Mispronunciations at word offset also yielded less target looking than correct pronunciations. This difference began several hundred milliseconds after word onset, but almost immediately after the onset of the mispronunciation. This result provides a baseline hypothesis for the analysis of children given the same task, namely that if children are like adults, they should reveal effects of both sorts of mispronunciations, with a delay in the impact of offset MPs comparable to the acoustic delay in the MPs' realizations.

Experiments 2A and 2B: children

Methods

Except as noted, methods in Experiment 2 were the same as in Experiment 1. Participants were children 14–22 months of age rather than adults.

Participants—Children were recruited via letters sent to parents in the Philadelphia area. In Experiment 2A (onset mispronunciations), the 60 children in the final sample ranged in age from 447 to 665 days (about 14 months 21 days or 14;21, to 21;26), with a mean of 540 days (17;23) and sd of 65 days. Half were girls; the ages of boys and girls were similar (means, 532 days vs. 549 days). An additional 27 children were tested but not included in the final sample because they became fussy (did not complete at least 6 trials in each condition), and two were tested but excluded due to experimenter error (1) or parental interference (1). In Experiment 2B (offset mispronunciations), the 36 children in the final sample ranged from 428 to 670 days (about 14;0 to 22;1), with a mean of 531 (17;14) and sd of 72 days. Half were girls, and the ages of boys and girls were similar (means, 522 days vs. 539). An additional 18 children were tested but not included because they became fussy (n=17) or because the parent peeked at the screen during the test (n=1).[2]

Visual stimuli—The visual stimuli were identical to those tested in Experiment 1, with a single exception. Some (28) children in Experiment 2A viewed a picture of a toy boat as the *boat* item; the remainder viewed the same modern yacht that was used in Experiments 1 and 2B. Although children recognized the toy boat somewhat better than the yacht, the impact of this change on the overall pattern of results was negligible; the outcomes of the analyses do not change in any significant way if this item is excluded.

For 28 children in Experiment 2A, the pictures were paired as follows: the dog with the sock, the boat with the duck, and the book with the cup. For all other participants the pictures were paired as in Experiment 1, wherein the dog and book were presented on the same trials. This pairing of *dog* with *book* might be expected to enhance the effects of mispronunciations, because the MPs *dook* and *bog* were initially compatible with the names of the distracter pictures (Swingley et al., 1999). Of course, children would have to know the onsets of either *book* or *dog* to take advantage of this overlap, but the fact that the overlap existed complicates

---

[2]Experiment 2A was run as two separate experiments, the first with 32 children and the second with 28. The second of these experiments was initiated following the completion of Experiment 2B to examine a nonsignificant difference between responses to onset place-of-articulation MPs and onset voicing MPs. This nonsignificant difference in the first study was not replicated in the second. Because the procedures were nearly identical in these two onset-MP studies, they are collapsed here into the single Experiment 2A. It is this historical sequence, rather than an a priori design decision, that led to the larger number of children in 2A than 2B. The results of all analyses are substantively the same if only the first 32 children in Experiment 2A are included.

comparison of this pair's results against the others. Any differences in effects that may be traceable to this change are noted below.

Auditory stimuli—In Experiment 1, participants heard one sentence on each trial. In Experiment 2, to help maintain children's interest in the procedure, the *Where's the [target]?* sentence was followed by either *Do you like it?* or *Can you find it?*.

Evaluation of neighborhood properties of the stimulus words—As noted previously, deCara and Goswami (2002) pointed out that English words that sound similar to each other (i.e., phonological neighbors) tend to be *rime* neighbors, sharing their vowel and coda, as opposed to *onset-overlapping* neighbors, sharing their onset and vowel. To determine whether this was also true of the specific words chosen as test items in the present study, the neighborhood characteristics of these words were examined in a corpus of 14 mothers' American English infant-directed speech (Brent & Siskind, 2001). From the set of over 487,000 word tokens, pronunciations of each word type were taken from the Callhome English Lexicon (Kingsbury, Strassel, McLemore, & MacIntyre, 1997). Words not contained in this dictionary (mostly polysyllabic, idiosyncratic babytalk words like *eensie* and *fussies*) were excluded, leaving 6506 word types. Each of these was checked against each of the six target words, counting the number of dictionary words that could be converted into a target word by adding, deleting, or substituting a single onset or coda sound. This analysis revealed that all six words had both onset and rime neighbors in the corpus. For all target words except *cup* and *duck*, each word had more rime neighbors than onset neighbors; *cup* and *duck* were tied at 3 and 9 neighbors of each sort, respectively. Considering only neighbors with corpus frequencies above one, all targets except *cup* had more rime neighbors (mean = 5.5) than onset neighbors (mean = 3.5), a significant difference (95% C.I. = 0.01 – 3.99). Lists of the neighbors of each target are given in Table 2. Thus, although the numerical advantage of rime neighbors of our target words was not overwhelming, it was present in this infant-directed speech corpus. If rime neighbors highlight onsets and therefore lead to onsets' accurate phonological specification, stronger effects of onset mispronunciations than offset mispronunciations would be expected.

Children's ability to pronounce the target words—Following procedures described later, parent report data on children's own pronunciations of the target words was gathered from 69 of the participants. Children's ability to say each of the six words (whether correctly or not) ranged over words from 29.0% (for *boat*) to 60.9% (*dog*), with a mean of 44.7%. Thus, a small majority of the test trials in the experiment concerned words that children were not yet attempting to say, according to their parents. Counting only words children were reported to say *correctly*, the analogous mean was 27.1%. Children's reported errors in pronunciation were strongly weighted toward coda consonants as opposed to onsets; considering all reported mispronunciations, coda errors were more than four times more frequent than onset errors. This pattern held for all six words; over the six words the mean proportion of coda errors was .827 (sd = .179).

Apparatus and procedure—Parents were asked to complete a MacArthur-Bates Communicative Development Inventory (Words and Sentences CDI; Fenson et al., 1994) before coming to the laboratory; most complied (91 of 96). Most of these (68) completed a modified version of the CDI that allowed parents to indicate separately whether their child understood each word or whether their child said each word. When parent and child arrived at the lab, the experimenter played with the child for a few minutes while discussing the procedure of the study. Then they were escorted into the room containing the presentation booth. Parents were instructed to try to keep their child on their lap and facing the monitor throughout the procedure. Parents were also asked to refrain from speaking, and to close their eyes and bend their head downward throughout the trials. (Parents not closing their eyes when the first trial

began were instructed to do so; thus, parents were blind to target side.) Children first saw a simple animation of a goldfish and a toy duck. Once children were oriented to the screen, the first trial was initiated.

Trials presented to children were paced more slowly than the trials presented to adualts. The time given to view the pictures before the speech stimulus began was 2 s rather than 500 ms, and following the added second sentence, the pictures remained on screen for 2 s rather than 1.5 s. Each trial lasted about 6 seconds (mean, 6.12 s, sd .20 s), and the entire procedure lasted 4 min.

Four stimulus orders were used. Two were the orders employed in Experiment 1, and two more reversed the left and right screen positions of the pictures in the first and second orders.[3]

After the experimental procedure, if participants were willing, an informal assessment of children's pronunciation of the 6 target words was conducted. For each word, parents were asked whether their child could say the word. If so, parents were asked to try to reproduce or otherwise characterize their child's pronunciation. (Some children were willing to speak the words on demand, but they were rare; we relied primarily on parent report.) These data were available for 67 of the 96 children (69.8%).

Coding—Coding was done as in Experiment 1. Trials on which children were not fixating the screen for at least 20 frames in the first 3 seconds after target-word onset were discarded, which affected about 4.5% of the trials.

Results and discussion

Analyses first determined whether children responded differently to correct pronunciations (CPs) and mispronunciations (MPs) in Experiments 2A and 2B. The second question was whether the timing of the effects differed for onset and coda MPs. Third, we examined whether the strength of mispronunciation effects varied with children's age or vocabulary size.

As described above, in previous studies we have characterized children's performance using a temporal window extending from 367 ms to 2000 ms from the onset of the spoken target word (e.g., Swingley & Aslin, 2000). However, this window may not capture the effects of mispronunciations occurring word-finally. Such effects might be revealed well after the actual MP in the signal, but cannot occur before the MP. Because Experiment 1 showed that the timing of differential reactions to MPs and CPs was closely aligned with the timing of the MPs themselves, the first analysis of Experiments 2A and 2B used the same window (367–2000 ms), but tied to the location of the mispronunciation rather than the onset of the word. Thus, anchored from word onset, the first analysis window for Experiment 2A extended from 367 to 2000 ms, and the window for 2B extended from 1133 to 2767 ms from word onset.

Children looked at the target more than the distracter when hearing a CP in both experiments (2A, mean target looking 61.2%, C.I. 8.3% – 14.0%; 2B, mean 67.0%, 12.9% – 21.2%). Target looking given MPs was significantly above chance only in Experiment 2B (2A, mean 52.3%, -0.6% – 5.3%; 2B, mean 54.0%, 0.5% – 7.5%). The slight advantage for Experiment 2B, which was shown somewhat less consistently by adults, makes sense: while Experiment 2A's analysis window of 367–2000 ms from word onset included the period in which children were first hearing the word and recognizing what it meant, the 1133–2767 window of Experiment 2B did not include this early period. Considering the 367–2000 window for the children in Experiment 2B, target fixation on CP trials was comparable to that found in 2A (mean target looking 62.9%, C.I. 9.0% – 16.7%).

---

[3]Due to an error, 3 children in 2A and two in 2B were tested in an order that was missing 1 MP and 1 CP trial.

An Anova including data from both 2A and 2B showed significantly reduced target looking on mispronunciation trials relative to correct-pronunciation trials. The apparent superiority of children in Experiment 2B, which was a consequence of 2B's later analysis window, was significant by subjects but not items (see Table 3). As in the adult study, the lack of interaction between Experiment and Condition suggests that children were disrupted to similar degrees by onset and offset MPs. The effects of mispronunciations were independently significant in each experiment (Expt. 2A, onset MPs, average effect 8.8%, 95% C.I. = 4.9% – 12.8%; 2B, offset MPs, average effect 6.8%, 95% C.I. = 1.7% – 11.9%).

Children's moment-by-moment target fixation is plotted in Figures 3A and 3B. As the Figures make clear, children's behavior while hearing the spoken words was very similar for CPs and offset MPs: target fixation rose from 50%, starting at approximately 500 ms from target onset. The effects of onset MPs were similar to the effects shown by adults, with a prolonged period in which target fixation remained around 50%. Offset MPs, too, elicited similar responses from children and adults, with a precipitous late decline in target fixation. The differences between children and adults concern quantitative features of the graphs, not their overall forms. Adults' performance was substantially better than children's (note that the scales on the y-axes have a wider range for the adults), and children's responses were much slower than adults'.

Detailed analyses of fixation proportions—To evaluate the timing of the effects more rigorously, the difference between CP and MP target fixation was computed for each subject in each experiment. As in Experiment 1, the means of these differences over subjects were divided by the moment-by-moment standard deviations, to get a measure of effect size. Because children's responses were more variable than adults', a threshold of 0.30 rather than 0.60 was used to identify the temporal location of the mispronunciation effects. In Experiment 2A, the first frame in which this measure exceeded the threshold was at 800 ms. All subsequent frames up to and including 2033 ms were also above 0.30. In no case did any frame present a value below -0.10. The MP effect (CP-MP for each child) significantly exceeded zero at 800 ms (mean, 7.2%; $t(59)= 2.79$, $p< .005$) and this frame was followed by a long series of frames on which the effect was significant by t-test. Thus, the effect was found to take place consistently over participants from 800 ms to 2033 ms (a 1233 ms span). In Experiment 2B, the first frame exceeding 0.30 was at 1767 ms, the first of a series of frames above 0.30 extending to and including 2767 ms. The effect at 1767 ms was significantly greater than zero (mean, 7.1%, $t(35)= 1.85$, $p< .05$), and once again this frame was followed by a long series of frames in which the effect exceeded chance levels. From 2800 to 3000 ms the effect size hovered around 0.30 (mean, 0.297), so to equalize the duration of the restricted test windows in Experiments 2A and 2B, the effect in Experiment 2B was considered to extend from 1767 to 3000 ms.

These restricted windows of analysis were used to evaluate the sizes of the effects for each MP location. As shown in Table 3, the condition effect over subjects was significant (which was necessarily true, given the selected windows), as was the effect over items (not trivially true). Most importantly, the effects of mispronunciation were equally large whether words were mispronounced at onset or offset, as shown by the lack of interaction. Thus, there were no indications that offset MPs were less influential than onset MPs in altering the course of children's fixations. An additional analysis using the restricted temporal windows compared first-half and second-half performance, and found no difference overall (F(1,94)= 1.2, p > .25) and no interactions with condition or experiment (all $F < .1$).

Analyses of discrete shifts in fixation—To evaluate in detail the eye movements that gave rise to the summary plots in Figure 3, an analysis of changes in fixation was conducted, as in Experiment 1. Children's shifts in fixation over time are plotted in Figure 4. In each condition, shifts away from the Distracter (dots and solid lines) and away from the Target (pluses and dashed lines) are shown, arrayed over time from the onset of the target word to

3000 ms afterward. Condition differences in which effect sizes (mean condition difference divided by *sd* of the difference) were greater than 0.30 or less than -0.30 are shown using horizontal lines spanning their temporal window. (Effect sizes were used as cutoffs rather than significance levels because the two mispronunciation types were not tested with the same number of participants. In all cases but one, paired t-tests yielded 2-tailed significance levels below 0.025; the remaining case was marginal at 0.075.) These differences in shifting probabilities were computed as in Experiment 1, but using 300 ms rather than 200 ms windows. (Narrower windows were feasible in the first experiment because adults were more consistent in their responses.)

Children's shifting patterns were similar to the adults' in several ways. Upon hearing correct pronunciations, children's likelihood of shifting increased quickly if they were fixating the distracter (dotted lines) and decreased if they were fixating the target (pluses). In Experiment 2A, children hearing onset MPs shifted equally to and from the target for more than a full second from target onset. Children were more likely to leave the distracter upon hearing a CP than an MP starting at about 650 ms from word onset (i.e., the 500–800 ms time bin); children were more likely to defect from the target upon hearing an MP than a CP starting at about 750 ms (the 600–900 ms time bin). In Experiment 2B, children hearing offset MPs were, for the most part, already looking at the target by that time. (Nonetheless, children were marginally more likely to shift away from the distracter upon hearing a CP than an MP at about 750 ms; i.e., in the 600–900 ms bin; p = .058.) Effects were concentrated in the shifts away from the target, where most children were looking when the MP happened (just as the adults had been). These shifts were more likely for MPs than CPs starting at about 1617 ms (i.e., the 1467–1767 ms time bin, which corresponds to 700–1000 ms after the MP itself).

The fact that children showed an impact of mispronunciation on recognition about 100 ms later for codas than for onsets (relative to when the mispronunciation occurred) might be taken as evidence that children were "garden-pathed" by the initial portions of the target words and were therefore delayed in detecting the mispronunciation. However, this small difference was not statistically robust. To evaluate possible temporal differences in children's responses to onset and coda MPs, fixation data from Experiments 2A and 2B were aligned at the MP's start in each trial, and moment-by-moment differences in CP and MP responding were computed for each child. Unpaired comparison of these difference scores across experiments revealed no significant differences between the conditions from 0 to 1500 ms after the onset of the mispronunciation. Thus, our best estimates of the timing of children's sensitivity to mispronunciations do not show any reliable difference between onset and offset MPs beyond the difference that may be predicted from the temporal position of the MPs themselves. In this respect, children were not different from adults.

Individual differences—The next set of analyses concerned individual differences in sensitivity to mispronunciation. These analyses were computed over proportion-to-target means for the temporal window in which MP effects were shown in children as a whole in each experiment, i.e. from 800 to 2033 ms for onset MPs, and from 1767 to 3000 ms for offset MPs. None of the conclusions about individual variation in MP effect sizes changes if other similar windows are used.

Children's reported spoken vocabulary sizes ranged from 0 to 574 (Experiment 2A) and from 0 to 538 (2B). These counts showed no strong correlations with mispronunciation effect sizes, using the restricted analysis windows for each experiment: onset, $r = .07$; offset, $r = .13$; both *ns*. Comprehension vocabulary size was also not correlated with the mispronunciation effect: onset, $r = -.12$; offset, $r = .16$; both *ns*. Vocabulary size did predict other aspects of performance. Using the 367–2000 ms temporal window traditionally used to assess word recognition in this procedure (e.g. Swingley & Aslin, 2000), both experiments revealed significant correlations

between target fixation proportions and spoken vocabulary size (Experiment 2A, CP $r(54) = .36$, $p < .01$; MP $r(54) = .25$, $p = .06$; Experiment 2B, CP $r(33) = .33$, $p = .05$; MP $r(33) = .18$, $p > .20$). The analogous correlations using comprehension vocabulary were somewhat higher (2A, CP $r(40) = .36$, $p < .02$; MP $r(40) = .50$, $p < .001$; 2B, CP $r(24) = .53$, $p < .002$; MP $r(24) = .38$, $p < .025$). These results are consistent with previous research on toddlers (Bailey & Plunkett, 2002; Killing & Bishop, 2008; Swingley, 2003; Swingley & Aslin, 2000, 2002). Thus, although children's vocabulary size was associated with fixation behavior, there were no indications that differential responses to correct and deviant pronunciations of words depended upon children's command of a large vocabulary.

Sensitivity to mispronunciations also did not appear to be contingent on children's knowledge of similar-sounding words (minimal pairs) exemplifying the tested contrasts. This possibility was evaluated by considering the vocabulary data provided by the 61 parents who evaluated their child's comprehension of each word on the CDI. The Inventory contains several minimal pairs like *bump–dump* and *take–tape*. In principle, knowledge of such words could trigger the insight in children that the sounds varying in such pairs (viz., the onsets [b], [d] and the codas [k], [p]) are contrastive in the language. Some of the tested mispronunciations, such as *poat* and *boad*, could benefit from this learning process in principle; others, namely *zock*, *guck*, *dawb*, and *cub*, involved phonological substitutions whose contrastive significance was not exemplified by any word pairs on the CDI.

Each child's knowledge of word pairs potentially contributing to awareness of a given phonological contrast was examined. For the most part, parents' estimates of children's vocabularies included few pairs of words that exemplified the tested contrasts. The most commonly known pairs involved the sounds *b–d* in syllable onset position: almost half of the children (45%) knew word pairs like *ball–doll*, which might have helped them reject *dook* as a pronunciation of *book* or *bog* for *dog*. A quarter (23%) knew *big* and *pig*, a pair potentially relevant for the item *poat*. But for the remaining nine stimulus items, an average of 5% of children knew any word-pairs that might have led them to discover the tested contrast via generalization from familiar word pairs. The pairs and the proportion of children's CDIs indicating knowledge of both words in each pair were as follows: blue/glue, .034; big/pig, .230; bump/dump, .046; ball/doll, .425; dry/try, .069; do/two, .046; keys/peas, .345; kick/pick, .115; could/good, .023; a(u)nt/and, .092; bat/bad, .080; write/ride, .046; take/tape, .034; coke/coat, .011; sick/sit, .069; wake/wait, .092; like/light, .069; knock/not, .080; pick/pig, .115. Note that these proportions were not evenly distributed across children; for example, most of the children who knew *bump* and *dump* also knew *ball* and *doll*. Although the CDI may underestimate vocabulary knowledge, particularly in older children, these results do not support a strong role for minimal pairs in helping to refine children's knowledge of the words that were tested.

Another mechanism that would in principle be effective in refining children's knowledge of words is parental correction. When children pronounce words in noncanonical ways, parents sometimes correct them (e.g., Chouinard & Clark, 2003), and children often struggle to improve their pronunciation and comment explicitly on their efforts (Clark, 1978). It is likely that correction is helpful when it is provided; whether it is *necessary* for the performance seen in the present studies can be evaluated by examining fixation responses only for trials on which children could not say the words they were being tested on. Parents cannot correct the pronunciations of words children do not say. Among the 66 children in the sample with report data, the percentage of children who were *not* saying each of the target words was: *boat*, 56; *book*, 38; *cup*, 53; *dog*, 32; *duck*, 40; *sock*, 50. Analyses of responses over only the words children were not reported to say included 29 children tested on onset mispronunciations and 26 tested on offset mispronunciations. (Eleven children were saying all six words, leaving no responses to examine in this analysis). Using the restricted temporal windows established as

described above, mispronunciation effects were still strong when limiting trials to words children could not yet say. For onset MPs, the mean effect was 16.0% (CI = 7.0 to 25.0); for offset MPs, 7.6% (CI = 0.0 to 15.2). Thus, children's encoding of onset and coda consonants does not depend upon children's ability to say the words, nor parental correction.

Children's responses to correct pronunciations and mispronunciations might be affected by the type of mispronunciation used. Recall that in each experiment, three of the tested MPs involved alterations of voicing (e.g. *cup–gup* or *sock–sog*) and three tested alterations of place of articulation (e.g. *book–dook* or *duck–dut*). Although the two phonological features were tested using different words, and thus cannot be compared directly, the MP effects in each experiment were not restricted to one or the other sort of MP. In Experiment 2A, onset MPs disrupted performance for place changes (mean effect, 14.0%; $t(59) = 5.1$, $p < .0001$, 95% C.I. 8.5–19.5) and voicing changes (mean effect, 8.7%, $t(59) = 3.0$, $p < .005$, C.I. 2.9–14.6). Likewise, in Experiment 2B, coda MPs had similar effects for both types of change (place, 13.4%, $t(35) = 3.2$, $p < .005$, C.I. 4.8–22.0; voice, 10.8%, $t(35) = 2.6$, $p < .015$, C.I. 2.5–19.2). These results match the findings of White and Morgan (2008).

As mentioned in Methods, Expt. 2A was carried out in two parts, with 32 children in the first and 28 in the second. The second part was motivated by a trend in the first part, in which a marginally greater effect of place-of-articulation MPs than voicing MPs was observed. This trend may have arisen because in the initial study the pictures *dog* and *book* were paired together on the screen. For these items, the MPs *bog* or *dook* were initially phonetically consistent with the distracter pictures. If children could detect this match when hearing MPs, they might have lingered on the distracter longer for these two items than for the others, leading to a larger effect of mispronunciation for these two items (and, in turn, an inflated estimate of the effects of place-of-articulation changes relative to voicing changes). The 28 children in the follow-up study saw *dog* paired with *sock* and *book* with *cup*); thus, for these children, no MPs overlapped phonetically with the distracter pictures' names at onset. Inspection of the onset-MP effects in these two parts of the experiment confirmed that the MP effect with the rearranged pairs was significantly smaller than the effect with the original, onset-overlapping picture pairs (rearranged: mean effect 10.4%, C.I. 2.4–18.5; overlapping: mean effect 23.3%, C.I. 13.3–33.3). This (serendipitous) result is consistent with the broader conclusion that children interpret speech incrementally as the signal unfolds (Swingley et al., 1999); it appears that even very young children can use the first consonant of a word to guide their interpretation.

The majority of the children showed better performance for CP targets than MP targets (44 of 60 in 2A, 26 of 36 in 2B). In neither experiment was the size of the MP effect (i.e., CP - MP) significantly correlated with children's age: onset, $r = .13$, $p > .2$; offset, $r = 0.28$, $p = .091$. Still, development in word recognition is known to take place over the tested age range of 14 to 22 months (Fernald et al., 1998). To examine whether the effects changed over the course of the second year, the children were grouped by age using a median split for each experiment: those children 525 days and older (just over 17 months) in Experiment 2A, and at 526 days in 2B. The mean age of the younger group was 479 days (almost 16 months); the older group, 595 days (19 and a half months).

To examine the effects in younger and older one-year-olds, age-group was added as a factor in an anova comparing CP and MP performance over the windows established above in each experiment. In Experiment 2B, testing offset MPs, there was neither a main effect of age group nor an interaction (age, $F(1,34) < 1$; age X condition, $F(1,34) = 1.03$, *ns*). In Experiment 2A, there was a main effect of age ($F(1,58) = 7.45$, $p < .01$) and an interaction of age and condition ($F(1,58) = 5.40$, $p < .025$). The interaction arose because the younger children's effects of condition (CP, MP) were smaller than those of the older children. However, considering each age group alone, effects of onset MPs were significant in both groups (younger children, CP

59.0%, MP 52.8%, t(29)= 2.23, *p*(2-tailed)< .05, 95% C.I.= 0.50 to 11.9; older children, CP 70.2%, MP 54.7%, t(29)= 5.33, *p*< .0001, C.I.= 9.6 to 21.6).

The fact that effects were larger for older children than younger children in Experiment 2A is a novel result. It came about because older children performed much better than younger ones on CP trials, and because this increase in CP trial performance with age was not consistently matched by a corresponding increase in MP trial performance (as had been found by Swingley & Aslin, 2000). It is possible that floor effects were responsible for this pattern in the present study. Among the younger children in Experiment 2A, 13/30 (43%) fixated the target picture 55% or less on CP trials, considering the traditional 367–2000 ms analysis window. Given that chance performance is 50% target looking, these children could not show large mispronunciation effects. Among the older children, only 6/30 (20%) looked at the target less than 55%. Thus, here the interaction of condition and age may have originated in a group of younger children whose poor CP performance left no "room" for an MP effect. This was not the case in Swingley and Aslin's (2002) study of 14–15 month olds, wherein only 12/50 (24%) of children fell below the 55% mark on CP trials. The difference in performance between the present study and Swingley & Aslin's study may be due to the selection of words in the present study, which was constrained to monosyllables with onset and coda consonants. The items used here may have been less well known by some of the younger participants. Thus, the conclusion we draw is that on the whole, both younger and older one-year-olds know the phonological characteristics of familiar words well enough to be hindered in recognition when those words are pronounced with a substituted consonant either in onset or coda position.

General Discussion

A range of considerations suggested that detection of mispronounced coda consonants might be delayed in development. Word-final consonants are, in general, less clearly articulated; they are heard only after perception of the initial parts of the word has led children to consider an interpretation; and they enjoy less of the (hypothesized) benefit of membership in dense phonological neighborhoods. Yet there were no signs here that small phonological changes in coda consonants had any less effect than changes in onset consonants. Even with an age range spanning most of the second year (i.e., 14 – 21 months), and with spoken vocabulary sizes extending from zero to several hundred, children showed significant mispronunciation effects that were uncorrelated with vocabulary size.

The timing of adults' and children's responses to mispronunciations revealed the continuous nature of speech perception throughout life: as the signal unfolds over time, listeners attempt to connect what they have heard so far to a linguistic interpretation. Clearly, these connections are made more quickly and reliably in adults than in children; adults detected mispronunciations virtually immediately, while children showed delays of several hundred milliseconds. Neither group provided evidence that listeners, having made an interpretation, remained wedded to it, as would be shown by weak or significantly delayed effects of offset mispronunciations relative to onset mispronunciations. Under the present testing conditions, phonological substitutions in both positions had similar effects on recognition, suggesting that even very young children's lexical representations are adequate, in principle, for supporting the phonological function of lexical contrast.

What does mispronunciation sensitivity mean for children's interpretation? It is important to note that we cannot determine whether children recognized that the MPs were not only poor realizations of their source CPs, but also potential words in their own right. Even when children reject (e.g.) a picture of a dog upon hearing *dawb*, this does not guarantee that children also have determined that *dawb* is another word. Children's default behavior while looking at the screen is to scan the images, one after the other. This behavior is altered by the speech stimulus:

when the spoken word matches the fixated object, children maintain their gaze; when it refers to a nonmatching object, children quickly shift their gaze to find that object (even if it turns out not to be the other image either; Swingley & Fernald, 2002). If neither situation applies because the word is hard to recognize, children search, perhaps attempting to satisfy an expectation that the spoken word will match one of the presented objects. Under conditions of uncertainty, search is a reasonable strategy given the ease of refixation (e.g., Droll & Hayhoe, 2007). Children hearing MPs might refixate as a low-cost way to find out if shifting yields the expected feeling of recognition, without explicitly considering the possibility that the MP is a new word.

This perspective on the meaning of gaze shifts in this experimental paradigm opens the door to reconsideration of the role of the lexicon in the development of speech-sound categories, given the present results and prior studies using eye movements. We have consistently found that children respond more accurately to canonical pronunciations of words than to phonologically deviant pronunciations, even in the youngest children we can measure (Swingley, 2005a; Swingley & Aslin, 2002, and here), even among children who do not appear to know any phonological neighbors of the words we are testing (Swingley, 2003), and even when we test single-feature mispronunciations of coda consonants (Experiment 2B). This body of work in general, and the present studies in particular, appear to contradict predictions of the Lexical Restructuring Model (LRM; e.g., Metsala & Walley, 1998), most directly in its assertion that in general words' forms are represented with only enough phonetic detail to tell apart the words already in the child's lexicon.

However, the fact that this claim is not supported by word recognition studies in toddlers does not mean that the development of phonological interpretation is unaffected by the characteristics of the developing lexicon. Here we suggest two ways to reconcile the present results with both the data supporting the LRM, and experiments showing language-specific phonetic tuning in infancy. One possibility is that phonological neighborhood structure exerts its influence at a metalinguistic level. When children learn words that sound similar, this may help children fine-tune their expectations regarding the degree to which phonological variation signals lexical difference. A number of studies have shown that small phonological changes to words are not reliably taken by young children as indications that a new word is being offered (Merriman & Schuster, 1991; Stager & Werker, 1997; Swingley & Aslin, 2007; White & Morgan, 2008). In some of these studies, when children were shown a familiar object and an unfamiliar object, and heard a deviant pronunciation of the familiar object's name, they seemed to assume that the speaker referred to the familiar object. In addition, at 18 months, children found a novel word difficult to learn if it overlapped phonologically with a very familiar word whose referent was not present (Swingley & Aslin, 2007). Younger children under some teaching conditions have trouble learning two similar-sounding words in one session (Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002; but see Fennell, 2006). Finally, in a series of studies Nazzi and colleagues have shown that while 20-month-olds can use consonantal differences between novel word labels (like and [pize] and [tize]) to group objects together, they fail if words differ in their vowels only (e.g., [pize] and [paze]; Nazzi, 2005, Nazzi & Bertoncini, in press; Nazzi & New, 2007).

The child's problem in all of these cases is not one of perceptual discrimination; the children in several of these studies have demonstrated their ability to tell the words apart. The problem is children's *interpretation* of the variability they can perceive. The interpretive process of deciding that a phonetic form is a word to be learned may be affected by phonological neighborhoods in the lexicon. (Similar accounts have been offered for the development of phonological awareness; e.g., De Cara & Goswami, 2003.) When children know many words that sound quite similar to one another, presentation of a novel phonological neighbor may lead them to reflect upon the possibility that the speaker intends to refer to an unfamiliar word.

This predicts better novel-neighbor learning among children who already knew many neighbors (evidently more neighbors than are typically known at 18 months; Swingley & Aslin, 2007). But it does not necessarily require greater sensitivity to mispronunciations in word recognition.

Another way to reconcile the LRM and the present results is to suppose that phonological categorization does indeed improve with development into middle childhood, and that phonological neighborhood structure is a crucial component of this improvement; however, the infant or young child's starting point is more advanced than assumed under LRM. Children's knowledge of many words is not vague. But their phonological representations may be better characterized as probability distributions over segments or features, rather than lists of phonetic categories, particularly for words children have not encountered very often (Swingley, 2007). On such an account, mispronunciation effects reveal that correct pronunciations match stored knowledge best (i.e., CPs match the highest-probability features in the representation), but in principle such effects leave room for further development in the form of sharpening probability distributions until each word's representation is effectively a list of segments each with a very high probability.

More generally, the notion that segmental phonology develops in infancy from analysis of the speech signal (e.g., Kuhl et al., in press) and the theory that phonology develops in early childhood from the lexicon (e.g., Storkel, 2002) have always been viewed as contradictory. This apparent disagreement comes from two assumptions that are probably incorrect. The first assumption is that language-specific phonological tuning developmentally precedes word learning. In fact, infants begin to learn words at the same time that they begin to learn their language's phonetic categories – on current evidence, between 6 and 12 months (words: Bortfeld et al., 2005; Jusczyk & Hohne, 1997; Swingley, 2005b; Tincoff & Jusczyk, 1999; phonetic categories: Bosch & Sebastián-Gallés, 2003; Cheour et al., 1998; Kuhl et al., 1992; Polka & Werker, 1994; Werker & Tees, 1984). Thus, right from the beginning, words have the potential to inform phonetic category learning in much the same way that words may inform mature phonetic categorization (McClelland & Elman, 1986).

A second common assumption with little support is that by early childhood, speech perception is no longer a significant limiting problem in language acquisition. Until very recently (e.g., Kuhl et al., 2005), infants' pattern of success in responding to native contrasts, and failure otherwise, was often interpreted as evidence that infants "solve" the speech perception problem in the first year. But studies of older children, using more sensitive and more demanding tasks, have revealed important changes in speech categorization into middle childhood (e.g., Hazan & Barrett, 2000; Nittrouer, 2002; Walley & Flege, 1999), as well as broad individual differences in children's phonological performance, differences that are linked to linguistic disorders like dyslexia (e.g., Swan & Goswami, 1997).

These developmental changes in childhood do not imply that the learning that begins in infancy is irrelevant. Infants begin to optimize the categorization of speech sounds according to the needs imposed by their language, and this perceptual tuning helps young children learn and recognize words. But the developmental process also involves improved interpretation of the linguistic implications of phonological variation, and additional refinement in phonetic categorization. A full account of phonological development will require investigation of the interpretation of speech beyond infancy, particularly in the understudied period between 12 months and the early school years. The present experiments place on stronger footing the viewpoint that children do have phonetic representations that are consistent with the canonical pronunciations of familiar words. Further developmental change, then, does not appear to consist of refinements that supply missing phonetic features. Instead, development may be a

matter of better interpretation of the linguistic significance of perceptible phonetic variation, perhaps by virtue of better confidence in phonetic categorization.

Acknowledgments

References

Allopenna PD, Magnuson JS, Tanenhaus MK. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory and Language 1998;38:419–439.

Bailey TM, Plunkett K. Phonological specificity in early words. Cognitive Development 2002;17:1265–1282.

Ballem KD, Plunkett K. Phonological specificity in children at 1;2. Journal of Child Language 2005;32:159–173. [PubMed: 15779881]

Bernhardt, BH.; Stemberger, JP. Handbook of phonological development: From the perspective of constraint-based nonlinear phonology. San Diego, CA: Academic Press; 1998.

Bortfeld H, Morgan JL, Golinkoff RM, Rathbun K. Mommy and me: Familiar names help launch babies into speech-stream segmentation. Psychological Science 2005;16:298–304. [PubMed: 15828977]

Bosch L, Sebastián-Gallés N. Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. Language and Speech 2003;46:217–243. [PubMed: 14748445]

Bowey JA, Hirakis E. Testing the protracted lexical restructuring hypothesis: The effects of position and acousticphonetic clarity on sensitivity to mispronunciations in children and adults. Journal of Experimental Child Psychology 2006;95:1–17. [PubMed: 16546204]

Brent MR, Siskind JM. The role of exposure to isolated words in early vocabulary development. Cognition 2001;81:B33–B44. [PubMed: 11376642]

Canfield RL, Smith EG, Brezsnyak MP, Snow KL. Information processing through the first year of life. Monographs of the Society for Research in Child Development 1997;62(2) Serial Number 250.

Cheour M, Ceponiene R, Lehtokoski A, Luuk A, Allik J, Alho K, Naatanen R. Development of language-specific phoneme representations in the infant brain. Nature Neuroscience 1998;1:351–353.

Chouinard MM, Clark EV. Adult reformulations of child errors as negative evidence. Journal of Child Language 2003;30:637–669. [PubMed: 14513471]

Clark, EV. Awareness of language: Some evidence from what children say and do. In: Sinclair, A.; Jarvella, RJ.; Levelt, WJM., editors. The child's conception of language. Heidelberg: Springer-Verlag; 1978. p. 17-43.

Cole RA. Perception of fluent speech by children and adults. Annals of the New York Academy of Sciences 1981;379:92–109.

Cole RA, Jakimik J, Cooper WE. Perceptibility of phonetic features in fluent speech. Journal of the Acoustical Society of America 1978;64:44–56. [PubMed: 712001]

Cole RA, Perfetti CA. Listening for mispronunciations in a children's story: The use of context by children and adults. Journal of Verbal Learning and Verbal Behavior 1980;19:297–315.

Cooper RM. The control of eye fixation by the meaning of spoken language. Cognitive Psychology 1974;6:84–107.

Creel SC, Aslin RN, Tanenhaus MK. Acquiring an artificial lexicon: segment type and order information in early lexical entries. Journal of Memory and Language 2006;54:1–19.

Creel SC, Dahan D. The influence of spoken words' temporal structure on paired-associate learning. under review.

Dahan D, Gaskell MG. Temporal dynamics of ambiguity resolution: Evidence from spoken-word recognition. Journal of Memory and Language 2007;57:483–501. [PubMed: 18071581]

Dahan D, Tanenhaus MK. Continuous mapping from sound to meaning in spoken-language comprehension: immediate effects of verb-based thematic constraints. Journal of Experimental Psychology: Learning, Memory, and Cognition 2004;30:498–513.

Dahan D, Tanenhaus MK. Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. Psychonomic Bulletin & Review 2005;12:453–459. [PubMed: 16235628]

Dahan, D.; Tanenhaus, MK.; Salverda, AP. The influence of visual processing on phonetically driven saccades in the 'visual world' paradigm. In: Gompel, Rv; Fischer, M.; Murray, W.; Hill, R., editors. Eye-movements: A window on mind and brain. Oxford: Elsevier; 2007. p. 471-486.

De Cara B, Goswami U. Similarity relations among spoken words: the special status of rimes in english. Behavior Research Methods, Instruments, and Computers 2002;34:416–423.

De Cara B, Goswami U. Phonological neighbourhood density: effects in a rhyme awareness task in five-year-old children. Journal of Child Language 2003;30:695–710. [PubMed: 14513474]

Dresher, BE. On the acquisition of phonological contrasts. In: Kampen, Jv; Baauw, S., editors. Proceedings of the generative approaches to language acquisition conference. Vol. 1. Utrecht: Landelijke Onderzoekschool Taalwetenschap; 2003. p. 27-46.

Droll J, Hayhoe M. Trade-offs between working memory and gaze. Journal of Experimental Psychology:Human Perception and Performance 2007;33:1352–1365. [PubMed: 18085948]

Fennell, CT. Proceedings of the 30th annual boston university conference on language development. Somerville, MA: Cascadilla Press; 2006. Infants of 14 months use phonetic detail in novel words embedded in naming phrases; p. 178-189.

Fennell CT, Werker JF. Early word learners' ability to access phonetic detail in well-known words. Language and Speech 2003;46:245–264. [PubMed: 14748446]

Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ. Variability in early communicative development. Monographs of the Society for Research in Child Development 1994;59(5) Serial Number 242.

Fernald A. Four-month-old infants prefer to listen to motherese. Infant behavior and development 1985;8:181–195.

Fernald A, Pinto JP, Swingley D, Weinberg A, McRoberts GW. Rapid gains in speed of verbal processing by infants in the second year. Psychological Science 1998;9:228–231.

Fernald A, Swingley D, Pinto JP. When half a word is enough: Infants can recognize spoken words using partial acoustic-phonetic information. Child Development 2001;72:1003–1015. [PubMed: 11480931]

Golinkoff RM, Hirsh-Pasek K, Cauley KM, Gordon L. The eyes have it: Lexical and syntactic comprehension in a new paradigm. Journal of Child Language 1987;14:23–45. [PubMed: 3558524]

Hallé PA, de Boysson-Bardies B. Emergence of an early receptive lexicon: infants' recognition of words. Infant Behavior and Development 1994;17:119–129.

Hallé PA, de Boysson-Bardies B. The format of representation of recognized words in the infants' early receptive lexicon. Infant Behavior and Development 1996;19:463–481.

Hazan V, Barrett S. The development of phonemic categorization in children aged 6-12. Journal of Phonetics 2000;28:377–396.

Jusczyk PW, Hohne EA. Infants' memory for spoken words. Science 1997;277:1984–1986. [PubMed: 9302291]

Kemler Nelson DG, Jusczyk PW, Mandel DR, Myers J, Turk A. The headturn preference procedure for testing auditory perception. Infant Behavior and Development 1995;18:111–116.

Killing SEA, Bishop DVM. Move it! Visual feedback enhances validity of preferential looking as a measure of individual differences in vocabulary in toddlers. Developmental Science 2008;11:525–530. [PubMed: 18576960]

Kingsbury, P.; Strassel, S.; McLemore, C.; MacIntyre, R. Callhome American English Lexicon (Pronlex). Philadelphia: Linguistic Data Consortium; 1997.

Kuhl PK, Conboy BT, Coffey-Corina S, Padden D, Rivera-Gaxiola M, Nelson T. Developmental phonetic perception: Native Language Magnet Theory Expanded (NLM-e). Philosophical Transactions of the Royal Society of London, B. in press.

Kuhl PK, Conboy BT, Padden D, Nelson T, Pruitt J. Early speech perception and later language development: implications for the 'critical period'. Language Learning and Development 2005;1:237–264.

Kuhl PK, Stevens E, Hayashi A, Deguchi T, Kiritani S, Iverson P. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Developmental Science 2006;9:F13–F21. [PubMed: 16472309]

Kuhl PK, Williams KA, Lacerda F, Stevens KN, Lindblom B. Linguistic experience alters phonetic perception in infants by 6 months of age. Science 1992;255:606–608. [PubMed: 1736364]

Magnuson JS, Tanenhaus MK, Aslin RN, Dahan D. The time course of spoken word learning and recognition: Studies with artificial lexicons. Journal of Experimental Psychology: General 2003;132:202–227. [PubMed: 12825637]

Mani N, Plunkett K. Phonological specificity of consonants and vowels in early lexical representations. Journal of Memory and Language 2007;57:252–272.

Marslen-Wilson WD. Sentence perception as an interactive parallel process. Science 1975;189:226–228. [PubMed: 17733889]

Marslen-Wilson WD, Welsh A. Processing interactions during word-recognition in continuous speech. Cognitive Psychology 1978;10:29–63.

McClelland JL, Elman JL. The TRACE model of speech perception. Cognitive Psychology 1986;18:1–86. [PubMed: 3753912]

Merriman WE, Schuster JM. Young children's disambiguation of object name reference. Child Development 1991;62:1288–1301. [PubMed: 1786716]

Metsala, JL.; Walley, AC. Spoken vocabulary growth and the segmental restructuring of lexical representations: precursors to phonemic awareness and early reading ability. In: Metsala, JL.; Ehri, LC., editors. Word recognition in beginning literacy. Mahwah, NJ: LEA; 1998. p. 89-120.

Mills D, Prat C, Zangl R, Stager C, Neville H, Werker J. Language experience and the organization of brain activity to phonetically similar words: ERP evidence from 14- and 20-month-olds. Journal of Cognitive Neuroscience 2004;16:1452–1464. [PubMed: 15509390]

Mirman D, Dixon JA, Magnuson JS. Statistical and computational models of the visual world paradigm: Growth curves and individual differences. Journal of Memory and Language. in press.

Nazzi T. Use of phonetic specificity during the acquisition of new words: differences between consonants and vowels. Cognition 2005;98:13–30. [PubMed: 16297674]

Nazzi T, Bertoncini J. Phonetic specificity in early lexical acquisition: new evidence from consonants in coda positions. Language and Speech. in press.

Nazzi T, New B. Beyond stop consonants: consonantal specificity in early lexical acquisition. Cognitive Development 2007;22:271–279.

Nittrouer S. Learning to perceive speech: How fricative perception changes, and how it stays the same. Journal of the Acoustical Society of America 2002;112:711–719. [PubMed: 12186050]

Nittrouer S, Studdert-Kennedy M, McGowan RS. The emergence of phonetic segments: Evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. Journal of Speech and Hearing Research 1989;32:120–132. [PubMed: 2704187]

Pisoni, DB.; Lively, SE.; Logan, JS. Perceptual learning of nonnative speech contrasts. In: Goodman, JC.; Nusbaum, HC., editors. The development of speech perception. Cambridge, MA: MIT Press; 1994. p. 121-166.

Polka L, Werker JF. Developmental changes in perception of nonnative vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance 1994;20:421–435. [PubMed: 8189202]

Ramon-Casas M, Bosch L, Swingley D, Sebastián-Gallés N. Vowel categorization during word recognition in bilingual toddlers. under review.

Redford MA, Diehl RL. The relative perceptual distinctiveness of initial and final consonants in CVC syllables. Journal of the Acoustical Society of America 1999;106:1555–1565. [PubMed: 10489711]

Samuel AG. Phonemic restoration: Insights from a new methodology. Journal of Experimental Psychology: General 1981;110:474–494. [PubMed: 6459403]

Stager CL, Werker JF. Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature 1997;388:381–382. [PubMed: 9237755]

Stoel-Gammon, C.; Dunn, C. Normal and disordered phonology in children. Austin, TX: Pro-Ed; 1985.

Storkel HL. Restructuring of similarity neighbourhoods in the developing mental lexicon. Journal of Child Language 2002;29:251–274. [PubMed: 12109371]

Swan D, Goswami U. Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis. Journal of Experimental Child Psychology 1997;66:18–41. [PubMed: 9226932]

Swingley D. Phonetic detail in the developing lexicon. Language and Speech 2003;46:265–294. [PubMed: 14748447]

Swingley D. 11-month-olds' knowledge of how familiar words sound. Developmental Science 2005a; 8:432–443. [PubMed: 16048516]

Swingley D. Statistical clustering and the contents of the infant vocabulary. Cognitive Psychology 2005b; 50:86–132. [PubMed: 15556130]

Swingley D. Lexical exposure and word-form encoding in 1.5-year-olds. Developmental Psychology 2007;43:454–464. [PubMed: 17352552]

Swingley D, Aslin RN. Spoken word recognition and lexical representation in very young children. Cognition 2000;76:147–166. [PubMed: 10856741]

Swingley D, Aslin RN. Lexical neighborhoods and the word-form representations of 14-month-olds. Psychological Science 2002;13:480–484. [PubMed: 12219818]

Swingley D, Aslin RN. Lexical competition in young children's word learning. Cognitive Psychology 2007;54:99–132. [PubMed: 17054932]

Swingley D, Fernald A. Recognition of words referring to present and absent objects by 24-month-olds. Journal of Memory and Language 2002;46:39–56.

Swingley D, Pinto JP, Fernald A. Continuous processing in word recognition at 24 months. Cognition 1999;71:73–108. [PubMed: 10444905]

Tincoff R, Jusczyk PW. Some beginnings of word comprehension in 6-month-olds. Psychological Science 1999;10:172–175.

Vihman MM, Nakai S, DePaolis RA, Hallé P. The role of accentual pattern in early lexical representation. Journal of Memory and Language 2004;50:336–353.

Walley AC. Young children's detections of word-initial and -final mispronunciations in constrained and unconstrained contexts. Cognitive Development 1987;2:145–167.

Walley AC. Spoken word recognition by young children and adults. Cognitive Development 1988;3:137–165.

Walley, AC. Speech perception in childhood. In: Pisoni, DB.; Remez, RE., editors. Handbook of speech perception. Oxford: Blackwell; 2005. p. 449-468.

Walley AC, Flege JE. Effect of lexical status on children's and adults' perception of native and nonnative vowels. Journal of Phonetics 1999;27:307–332.

Walley AC, Metsala JL. The growth of lexical constraints on spoken word recognition. Perception and Psychophysics 1990;47:267–280. [PubMed: 2326149]

Werker JF, Fennell CT, Corcoran KM, Stager CL. Infants' ability to learn phonetically similar words: effects of age and vocabulary size. Infancy 2002;3:1–30.

Werker JF, Pons F, Dietrich C, Kajikawa S, Fais L, Amano S. Infant-directed speech supports phonetic category learning in English and Japanese. Cognition 2007;103:147–162. [PubMed: 16707119]

Werker JF, Tees RC. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behavior and Development 1984;7:49–63.

White KS, Morgan JL. Sub-segmental detail in early lexical representations. Journal of Memory and Language 2008;59:114–132.
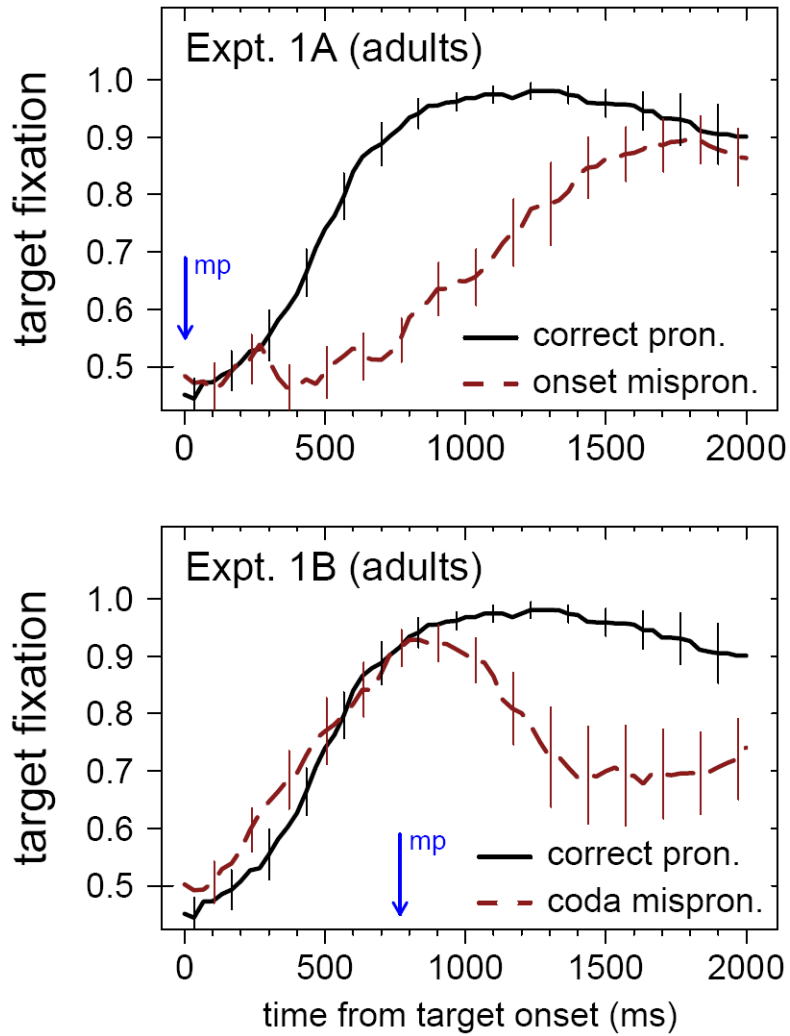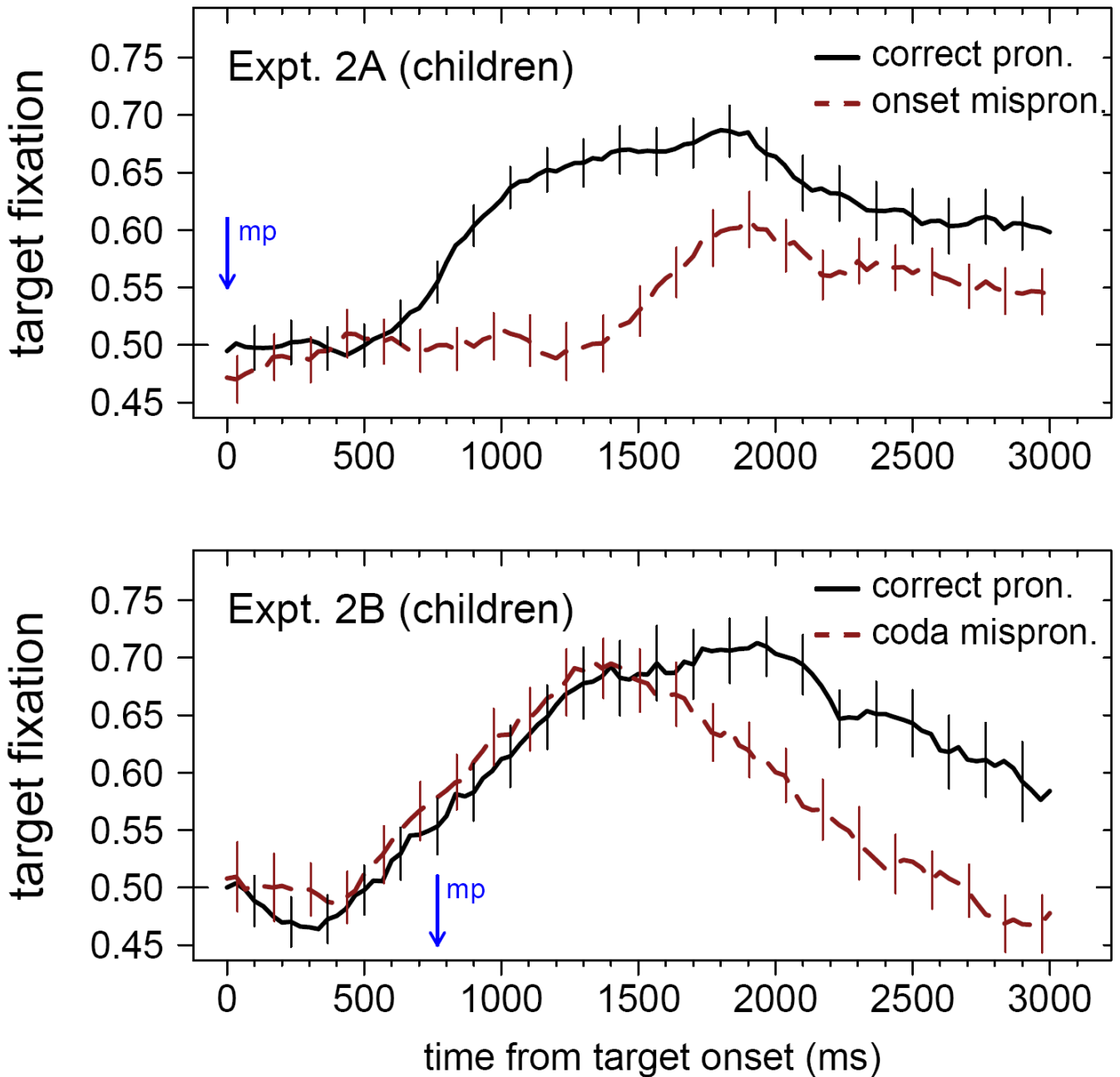
Figure 1.
Results from Experiments 1A and 1B, showing the timecourse of adults' fixation to the named target picture upon hearing a correct pronunciation or a mispronunciation. The upper panel shows results from Experiment 1A, which tested onset mispronunciations, while the lower panel shows results from Experiment 1B, which tested coda (offset) mispronunciations. Solid lines show fixation proportions for correct pronunciations and dashed lines for mispronunciations. Vertical error bars show standard errors of subject means in each condition for every fourth 33-ms timeframe. The vertical arrow in each panel shows the temporal location of the mispronunciations ("mp").
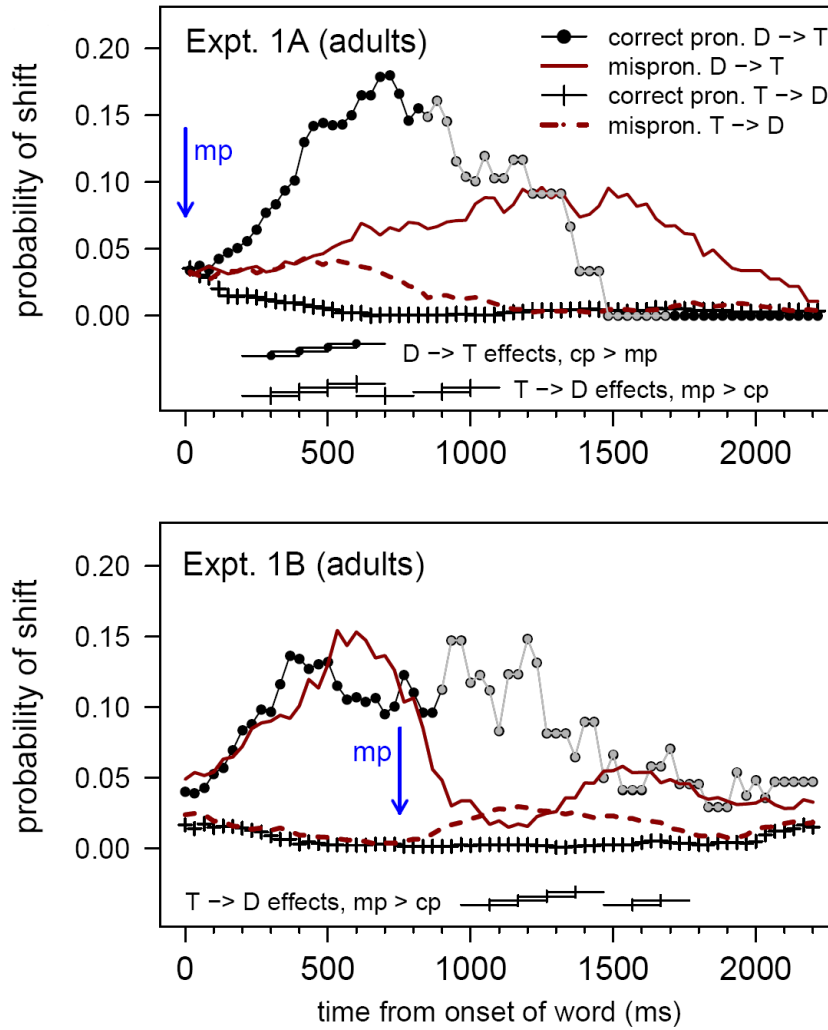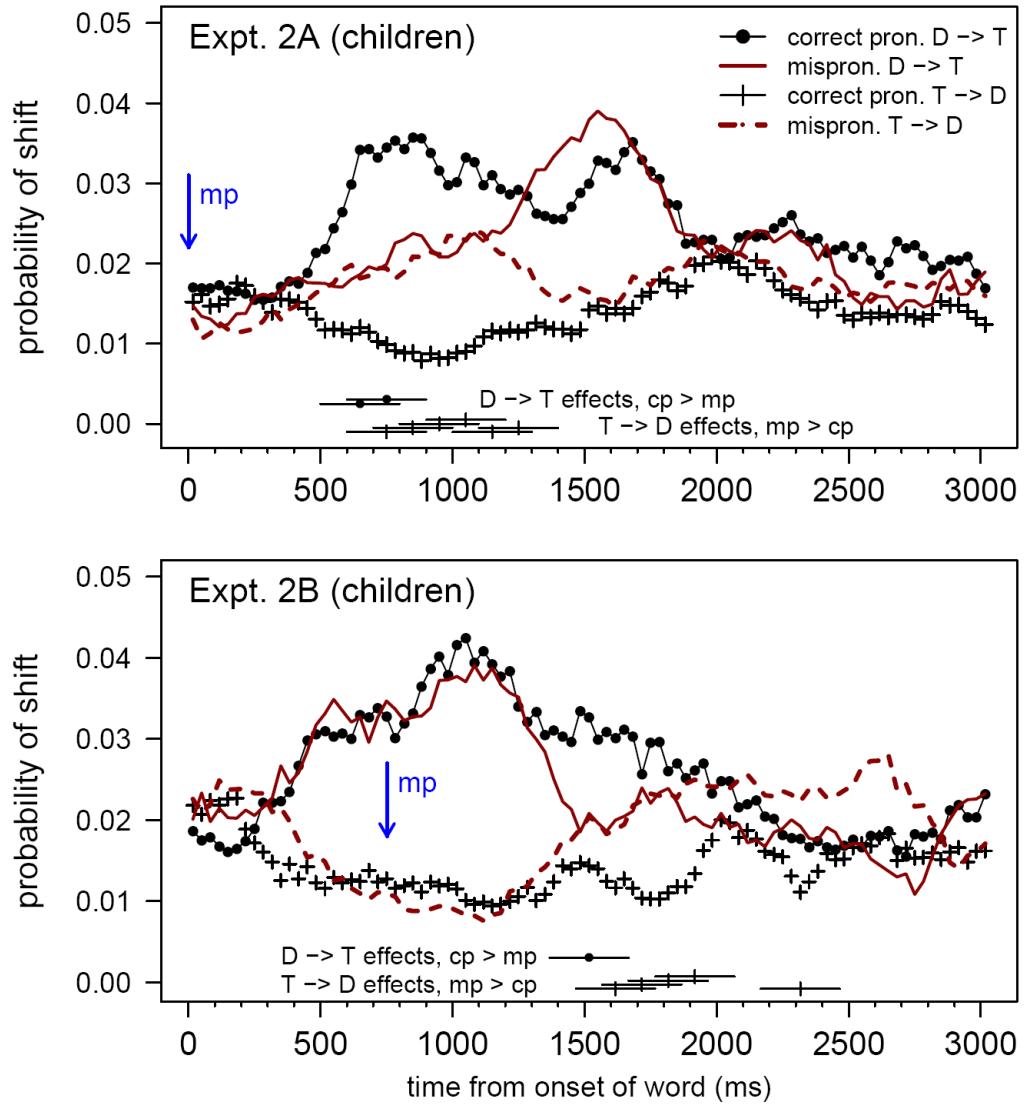
Figure 2.
Results from Experiments 1A and 1B, showing moment-by-moment changes in the probability of shifts from the distracter to the target (D→T) or vice versa (T→D) in each condition. The upper panel shows results from Experiment 1A, the lower panel results from Experiment 1B. Horizontal lines along the x-axis mark 200-ms bins in which the likelihood of shifting was significantly greater in one pronunciation condition than another ($p < .01$). Gray points indicate times at which fewer than 10 trials entered into the probability computation. "Correct pronunciation" is abbreviated as "cp", and "mispronunciation" as "mp."

Figure 3.
Results from Experiments 2A and 2B, showing the timecourse of children's fixation to the named target picture upon hearing a correct pronunciation or a mispronunciation. The solid line shows fixation proportions for correct pronunciations, the dashed line mispronunciations. Vertical error bars show standard errors of subject means for each condition for every fourth 33-ms timeframe. The vertical arrow in each panel shows the temporal location of the mispronunciations ("mp").

Figure 4.
Results from Experiments 2A and 2B, showing moment-by-moment changes in the probability of shifts from the distracter to the target (D→T) or vice versa (T→D) in each condition. Horizontal lines along the x-axis mark 500-ms bins in which the likelihood of shifting was significantly greater in one pronunciation condition than another ($p < .01$). "Correct pronunciation" is abbreviated as "cp", and "mispronunciation" as "mp."

Table 1

Anova results for Experiments 1A and 1B. The "restricted window" corresponds to the time from 367 to 1400 ms after target-word onset in Experiment 1A, and from 1133 to 2167 ms after target-word onset in Experiment 1B.

	by participants			by items			min F'		
	df	F1	p	df	F2	p	df	minF'	p
367 to 2000 ms window									
Condition (correct, mispron.)	1,24	30.28	<.0001	1,5	64.64	<.0005	1,25	20.62	<.0005
Expt. 1A vs 1B	1,24	<1		1,5	1.87	>.2	1,24	<1	
Expt. 1A vs 1B X Condition	1,24	1.01	>2	1,5	<1		1,20	<1	
restricted window									
Condition (correct, mispron.)	1,24	42.43	<.0001	1,5	42.52	<.002	1,25	21.24	<.0003
Expt. 1A vs 1B	1,24	2.52	.125	1,5	10.75	<.02	1,29	2.04	.164
Expt. 1A vs 1B X Condition	1,24	1.01	>.2	1,5	<1		1,29	<1	

Table 2

Phonological neighbors of the tested target words, as identified in the Brent & Siskind child-directed speech corpus. Words are ordered by frequency. Neighbors with frequencies of 100 or more are given in boldface, and words with a frequency of only one are given in italics.

Target	onset neighbors	rime neighbors
boat	**bowl**, both, bone, Bo, board, bored	coat, goat, throat, wrote, float, note, moat, *quote*
book	bush, bull	**look**, **took**, cook, hook, shook, brook
cup	**come**, cut, cub	**up**, yup, *pup*
dog	doll	frog, hog, log, *bog*
duck	**done**, **does**, dump, dum, dumb, dust, dub, *dug*, *dove*	**truck**, **yuck**, **stuck**, cluck, suck, tuck, struck, snuck, luck
sock	sopped, psalms	**knock**, **block**, rock, clock, lock, tock, *shock*, *crock*, *cock*

Table 3

Anova results for Experiments 2A and 2B. The "restricted window" corresponds to the time from 800 to 2033 ms from the onset of the target words in Experiment 2A, and from 1767 to 3000 ms from the onset of the target words in 2B.

	by participants			by items			min F'		
	df	F1	p	df	F2	p	df	minF"	p
367 to 2000 ms from MP location									
Condition (correct, mispron.)	1,94	27.30	<.0001	1,5	21.29	<.01	1,15	11.96	<.005
Expt. 2A vs 2B	1,94	15.10	<.0005	1,5	3.57	.117	1,8	2.89	.127
Expt. 2A vs 2B X Condition	1,94	<1		1,5	<1		1,10	<1	
restricted window									
Condition (correct, mispron.)	1,94	41.62	<.0001	1,5	28.88	<.005	1,14	17.05	<.002
Expt. 2A vs 2B	1,94	<1		1,5	<1		1,64	<1	
Expt. 2A vs 2B X Condition	1,94	<1		1,5	<1		1,93	<1	