

Two-year-olds' sensitivity to subphonemic mismatch during online spoken word recognition

Melissa Paquette-Smith¹ · Natalie Fecher¹ · Elizabeth K. Johnson¹

Published online: 31 August 2016
© The Psychonomic Society, Inc. 2016

Abstract Sensitivity to noncontrastive subphonemic detail plays an important role in adult speech processing, but little is known about children's use of this information during online word recognition. In two eye-tracking experiments, we investigate 2-year-olds' sensitivity to a specific type of subphonemic detail: coarticulatory mismatch. In Experiment 1, toddlers viewed images of familiar objects (e.g., a boat and a book) while hearing labels containing appropriate or inappropriate coarticulation. Inappropriate coarticulation was created by cross-splicing the coda of the target word onto the onset of another word that shared the same onset and nucleus (e.g., to create *boat*, the final consonant of *boat* was cross-spliced onto the initial CV of *bone*). We tested 24-month-olds and 29-month-olds in this paradigm. Both age groups behaved similarly, readily detecting the inappropriate coarticulation (i.e., showing better recognition of identity-spliced than cross-spliced items). In Experiment 2, we asked how children's sensitivity to subphonemic mismatch compared to their sensitivity to phonemic mismatch. Twenty-nine-month-olds were presented with targets that contained either a phonemic (e.g., the final consonant of *boat* was spliced onto the

initial CV of *bait*) or a subphonemic mismatch (e.g., the final consonant of *boat* was spliced onto the initial CV of *bone*). Here, the subphonemic (coarticulatory) mismatch was not nearly as disruptive to children's word recognition as a phonemic mismatch. Taken together, our findings support the view that 2-year-olds, like adults, use subphonemic information to optimize online word recognition.

Keywords Word recognition · Coarticulation · Speech perception · Subphonemic detail

Adult language processing is characterized by an acute sensitivity to fine-grained subphonemic acoustic-phonetic detail. For example, adult word recognition is hindered by misleading coarticulatory information on vowels (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; McQueen, Norris, & Cutler, 1999; Whalen, 1991) and affected by subcategorical variation in consonant duration (e.g., McMurray, Tanenhaus, & Aslin, 2002; Shatzman & McQueen, 2006). Because the extraction of subphonemic detail from speech is thought to optimize adult online word recognition (e.g., Spinelli, McQueen, & Cutler, 2003), it seems likely that developing efficient word recognition abilities in childhood may also involve attention to subphonemic detail. However, to date, there has been very little work examining the development of children's sensitivity to this type of information during online spoken word recognition.

✉ Elizabeth K. Johnson
elizabeth.johnson@utoronto.ca

¹ Department of Psychology, University of Toronto Mississauga, 3359 Mississauga Road North, Mississauga, Ontario, Canada L5L 1C6

Instead, much of the infant speech perception literature has focused on determining how and when infants identify the segments that are contrastive in their native language (e.g., Houston, 2011; Johnson, 2016; Werker & Hensch, 2015; Werker & Tees, 1984). This work has shown that by the time children reach their first birthday, they are already more attentive to segmental contrasts that signal differences between lexical items in their native language than to contrasts that do not (e.g., the speech sounds /l/ and /r/ are phonemically contrastive in English, but not in Japanese). Although there is some variability in how easily infants appear to learn different contrasts (e.g., Narayan, Werker, & Beddor, 2010), in general it seems that the faster infants become attuned to the sounds that signal phonological contrasts in their native language, the stronger their later language abilities (Kuhl, Conboy, Padden, Nelson, & Pruitt, 2005; Tsao, Liu, & Kuhl, 2004). There is also evidence that children, like adults, can use phonemic detail to recognize words as the speech signal unfolds (Fernald, Swingley, & Pinto, 2001; Swingley, 2009; Swingley & Aslin, 2002), and that the efficiency with which children do this gradually improves over the first few years of life (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998).

Although we know a great deal about children's sensitivity to phonemic contrasts in their native language (e.g., the difference between the vowel in *boat* and the vowel in *bat*), we know much less about children's attention to noncontrastive subphonemic variation (e.g., the difference in the realization of the vowel in *boat* and *bone*—although the vowel is phonemically the same in both words, it is colored differently by the articulatory overlap with the following oral vs. nasal coda consonant). Understanding the development of children's sensitivity to subphonemic detail in online word recognition is important because the use of this information is essential to achieving adult-like proficiency in spoken language processing (e.g., McQueen, 2007). Several studies have examined infants' (Curtin, Mintz, & Byrd, 2001; Fowler, Best, & McRoberts, 1990; Johnson, 2003; Johnson & Jusczyk, 2001; McMurray & Aslin, 2005) and young children's (Dietrich, Swingley, & Werker, 2007; Fisher, Hunt, Chambers, & Church, 2001) sensitivity to subphonemic or noncontrastive variation in offline tasks, but very little work has examined whether young children can use subphonemic information to optimize online word recognition. The few studies that do exist on this topic ask the same question: Can children use coarticulatory information to predict upcoming words? The results of these studies have not been clear.

Some studies have suggested that toddlers (Mahr, McMillan, Saffran, Weismer, & Edwards, 2015) and young children (Zamuner, Moore, & Desmeules-Trudel, 2016) can use coarticulatory information in this manner, whereas other studies have found no evidence that 2-year-olds' use of this type of information (Minaudo & Johnson, 2013).

In this study, we investigate 2-year-olds' sensitivity to subphonemic information during online word recognition using a different approach. Rather than asking whether toddlers can use noncontrastive subphonemic information to anticipate upcoming words in the speech stream, we use a child-friendly eye-tracking procedure (also referred to as the Looking-While-Listening Procedure) to ask whether toddlers can detect coarticulatory mismatches in the realization of vowels in familiar words. In Experiment 1, we compare children's recognition of known words when the initial consonant and vowel of the words are identity-spliced with a different token of the same word (e.g., the final C of one token of *boat* was spliced onto the initial CV of another token of *boat*) to instances where these same known words are cross-spliced with a different word (e.g., the final C of *boat* was spliced onto the initial CV of *bone*). Adult studies using a similar methodology have reported that adult word recognition is hindered when words contain inappropriate, or mismatching, coarticulation (Dahan et al., 2001; McQueen et al., 1999; Whalen, 1991). Thus, we reason that if 2-year-olds are sensitive to noncontrastive subphonemic information during online word recognition, then they should identify cross-spliced items (containing inappropriate coarticulation) less efficiently than identity-spliced items (containing appropriate coarticulation). In addition, we explore the possibility that children's ability to detect coarticulatory mismatch may improve with age by comparing performance across two different ages: 24-month-olds and 29-month-olds.

In Experiment 2, we compare 2-year-olds' sensitivity to subphonemic and phonemic mismatch during online word recognition. Here, we reason that if children's early representations are overspecified (as has been argued for younger children; e.g., Werker & Curtin, 2005), then both subphonemic and phonemic mismatches might be equally disruptive to 2-year-olds' word recognition.

Experiment 1

The Looking-While-Listening Paradigm was used to investigate how efficiently 2-year-olds recognize familiar words presented with appropriate versus inappropriate

vowel coarticulation. On each of 24 trials, children viewed images of two familiar objects (e.g., a boat and a book) while being asked to look at one of the objects (e.g., “Can you find the boat?”). Two-thirds of the trials contained identity-spliced tokens, in which the coda of the target word was spliced onto the onset and vowel of another token of the same word (so that the coarticulatory cues in the vowel matched the upcoming coda-final consonant(s)). The remaining one third of the trials contained cross-spliced tokens, where the coda of the target word was spliced onto a token of another word containing the same onset and vowel (e.g., the coda-final consonant in *boat* was spliced onto the initial CV of *bone*). As a result, the vowel in these tokens of the target word contained inappropriate coarticulatory cues to the upcoming coda-final consonant.

We predicted that if 2-year-olds are sensitive to subphonemic changes to vowel coarticulation, then we should observe more efficient word recognition during Identity-spliced than Cross-spliced trials. Moreover, if sensitivity to fine-grained phonetic detail improves as children age, then the older 2-year-olds should exhibit a bigger difference in their word recognition performance on Cross-spliced trials compared to Identity-spliced trials than the younger 2-year-olds (i.e., the mismatching coarticulatory cues in the cross-spliced stimuli should hinder word recognition more in the older children than the younger children).

Method

Participants Thirty-six 24-month-old ($M_{\text{age}} = 739$ days; range = 704–790; 16 females) and twenty-four 29-month-old ($M_{\text{age}} = 872$ days; range = 826–910; 14 females) monolingual English-learning children were tested (all had at least 90 % English input). Fourteen additional children were tested but excluded from the study due to disinterest or fussiness (11), parental interference (1), or experimenter error (2).

Materials, apparatus, and procedure Target words consisted of 24 C(C)VC(C) monosyllabic nouns commonly known by 2-year-olds. To facilitate the creation of cross-spliced items, targets were chosen with the constraint that another noun in English had a matching onset and nucleus, but mismatching coda (see Appendix 1 for a list of target words and their splicing counterparts). Note that some of the splicing counterparts were words commonly known by 2-year-olds (e.g., *bite*), but most of them were likely unknown by our participants (e.g., *plague*).

The targets and their splicing pairs were recorded in child-friendly carrier phrases (e.g., “Oh! Can you find the [target]? Isn’t it pretty?” or “Look! Do you like the

[target]? Amazing, eh?”) by a native English-speaking female. Target words always occurred in utterance-final position and were followed by a clear pause and then an ending phrase (e.g., “Amazing, eh?”). Because the articulation for adjacent segments overlap, some criteria were needed for deciding when one segment ended and the next one began. Stop closures were considered part of the coda consonants. Boundaries between vowels and nasals were identified by attending to formant trajectories as well as the point at which there was a marked decrease in intensity. Cross-spliced stimuli were created by splicing the coda of the target word onto the onset and nucleus of its splicing pair. Identity-spliced targets were created by splicing the coda of the target word onto the onset and nucleus of another token of the same target word. To avoid the introduction of splicing artifacts, all splicing was done using Praat (Boersma & Weenink, 2016) at zero crossings so that no pops were audible.

Since cues to the identity of the final consonant can occur in the preceding vowel (e.g., formant transitions, vowel length), an adult rating study was conducted to ensure that adults perceived the cross-spliced stimuli as instances of a subphonemic rather than phonemic mismatch. Using a forced-choice task, we presented the set of tokens to native English-speaking adults ($N = 12$) and asked them to identify whether the word they heard was the target word or the splicing pair that was used to create that word (e.g., the words *boat* and *bone* would appear on the screen, and participants would hear the cross-spliced token of *boat* containing the mismatching coarticulatory information on the vowel). For the cross-spliced items, adults overwhelmingly chose the target word over the splicing pair ($M = 94\%$, $SD = 6.48$), indicating that the cross-spliced items contained subphonemic rather than phonemic changes and were appropriate stimuli for our toddler study.

The visual stimuli consisted of 12 pairs of still images presented side-by-side on a white background. The images were matched in size. The visual complexity of the images was matched as closely as possible by attending to how intricate the images were and by relying on our past experience working with children (e.g., knowing that a 2-year-old will typically find a ball far more interesting than a box, regardless of how visually complex the box is). To make the word recognition task more challenging for children and to encourage them to attend to the vowels rather than just the onset consonants, 11 of the 12 image pairs were matched in their onset (e.g., a boat and a book). Each child saw each image pair twice, with a different object labeled on the

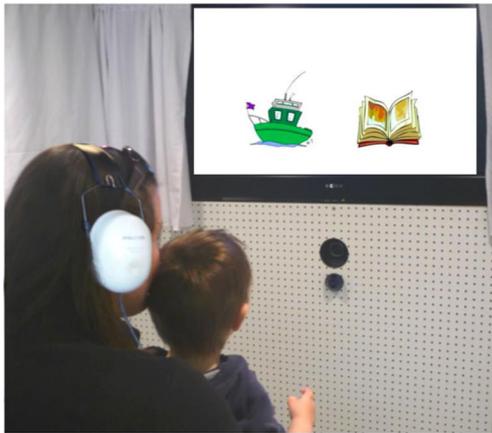


Fig. 1 Using the looking-while-listening paradigm, 24- and 29-month-olds were presented with two side-by-side images of familiar nouns. The target and distractor images were matched in word onset (e.g., boat and book) and were accompanied by a phrase labeling one of the objects

two occasions. Thus, the image that served as the target in one trial served as the distractor in another.

During the experiment, the child was seated on his or her caregiver's lap, facing a large TV screen in the center of an Industrial Acoustics Corporation (IAC) sound-attenuated booth (see Fig. 1). A 2 s flashing white star on a black background was presented before each trial to attract the toddler's attention to the center of the screen. Each trial lasted 6 s. The images appeared at the beginning of the trial, and the target words occurred exactly 3 s into the trial. Because the two items depicted on the screen were matched in onset, the average disambiguation point was 102.9 ms ($SD = 80.5$ ms) after word onset. Caregivers were asked to wear headphones and listen to masking music to prevent them from biasing their child's responses. Children's eye movements were recorded for offline coding by a camera situated below the television screen. After the experiment, parents completed the *MacArthur-Bates Communicative Developmental Inventories—Words and Sentences* form (CDI; Fenson et al., 2007).

Design Three experimental lists were created, each containing eight Cross-spliced and 16 Identity-spliced trials. The assignment of words to the Cross-spliced versus Identity-spliced trials was counterbalanced across lists. Each participant was tested on one of two randomized orders of a list and heard every target item once in either an Identity-spliced or a Cross-spliced trial (i.e., no child heard the same word as both cross-spliced and identity-spliced).¹

¹ For six of the 60 participants, one of the identity-spliced targets was mistakenly repeated twice. The repeated trial was excluded from the analysis.

Coding and analysis Each 30 ms frame was coded as a look to the left image, or to the right image, or neither. All coding was done with the audio track disabled so that the coder was blind to both the target location and the trial type. Four randomly selected videos were recoded by a second coder, and reliability was high (Mean $r = .96$, $SD = .04$).

Children's looking behavior was analyzed in the 1 s window of analysis starting 500 ms after target word onset. Although preferential looking studies often use a window of analysis between 1,500–2000 ms in length (e.g., Swingley, 2009; Swingley & Aslin, 2002), we chose to use a shorter, 1,000 ms window of analysis as we expected that any age-related differences between children's looks to identity-spliced and cross-spliced items might be fleeting.² We began our window of analysis 500 ms after target word onset because we deemed that looks prior to this point were unlikely to be driven by recognition of the target item. We based this decision on the fact that listeners need time to program an eye movement (e.g., Swingley & Aslin, 2000), and it takes longer to recognize words in eye-tracking studies when the two images on the screen have the same rather than different onsets (see Dahan et al., 2001, for a similarly timed window of analysis in an adult eye-tracking study using a closely related design).

Results

The looks to target in the 1 s window of analysis beginning 500 ms after target word onset were examined using a weighted empirical-logit regression in a linear mixed effects model (Barr, Gann, & Pierce, 2011). The model was implemented using the lme4 package of the statistical software R 3.2.2 (Bates et al., 2015; R Development Core Team, 2015) using two deviation-coded independent variables, Trial Type (-1: Identity-spliced, 1: Cross-spliced) and age group (-1: 24-month-olds, 1: 29-month-olds). The model included Trial Type, Age, and the Age \times Trial Type interaction as fixed effects. Following the Barr, Levy, Scheepers, and Tily (2013) paper on model selection, we used a maximal structure of random effects including random intercept and Trial Type slopes for participants as well as random intercept, Age, Trial Type, and Age \times Trial Type slopes for items. For the fixed effects we report b , standard error, t values, and p values calculated using Satterthwaite approximations to degrees of freedom and implemented in lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2015).

There was a significant main effect of Trial Type (Identity-spliced vs. Cross-spliced), $b = -0.09$, $SE = 0.04$, $t(345.2) = -2.42$, $p = .016$, and Age (24-month olds vs. 29-month-olds), b

² Note that we are not the first study to use a 1,000 ms window of analysis (e.g., Creel, 2012), and that our results look similar if we use a 2,000 ms window of analysis.

= 0.09, $SE = 0.04$, $t(72.7) = 2.35$, $p = .022$, but no interaction, $b = -0.01$, $SE = 0.04$, $t(179.6) = -0.24$, $p = .814$. The older 29-month-olds showed better recognition of the words, regardless of trial type. However, both the younger and the older children looked longer to the target in Identity-spliced trials than Cross-spliced trials (see Fig. 2). Thus, although we found clear evidence that 2-year-olds are sensitive to subphonemic mismatch, we found no evidence to support our hypothesis that children become more sensitive to this information over the course of the second year of life. We also note that although this experiment was not designed to look at the effects of individual items, for the majority of the items (i.e., 17 out of 24 items) children looked more to the target upon hearing the identity-spliced compared to the cross-spliced tokens (see Fig. 3). In the model, our effects hold regardless of the individual differences between items (which were included as random effects in the model).

Experiment 2

In Experiment 1, we found that both 24- and 29-month-old children readily detect inappropriate vowel coarticulation in familiar words. This could be taken as evidence that children, like adults, are sensitive to coarticulatory information in the speech signal. However, many questions regarding children’s perception of noncontrastive coarticulatory mismatch remain. Although our findings fit with the notion that children

have well-specified representations of familiar words, well-specified words are not necessarily adult-like. Indeed, it is possible that children’s representations could be overspecified, such that inappropriate coarticulation may disrupt word recognition as much as, for example, a familiar word spliced with an inappropriate vowel (for a discussion of possible overspecification in early lexical representations, see Singh, White, & Morgan, 2008; Werker & Curtin, 2005). That is, children may weigh coarticulatory mismatch as strongly as a vowel mismatch where, for example, the CV of *boat* is replaced with the CV from *bait*.

In Experiment 2, we explored this possibility by comparing 29-month-olds’ recognition of words that contain inappropriate coarticulation to words that contain a phonemically different vowel. We will henceforth refer to these two conditions as *Subphonemic Mismatch* and *Phonemic Mismatch*. Subphonemic mismatches were created in the same way as they were in Experiment 1. To create the phonemic mismatches, the coda of a target word was cross-spliced onto the onset and nucleus of a word with a different vowel (e.g., the final consonant of *boat* was spliced onto the initial CV of the word *bait*). As outlined in the *Method* section, the procedure in Experiment 2 differed in several key respects from Experiment 1. Most importantly, rather than seeing two familiar objects on the screen, on each mismatch trial participants saw one familiar object and

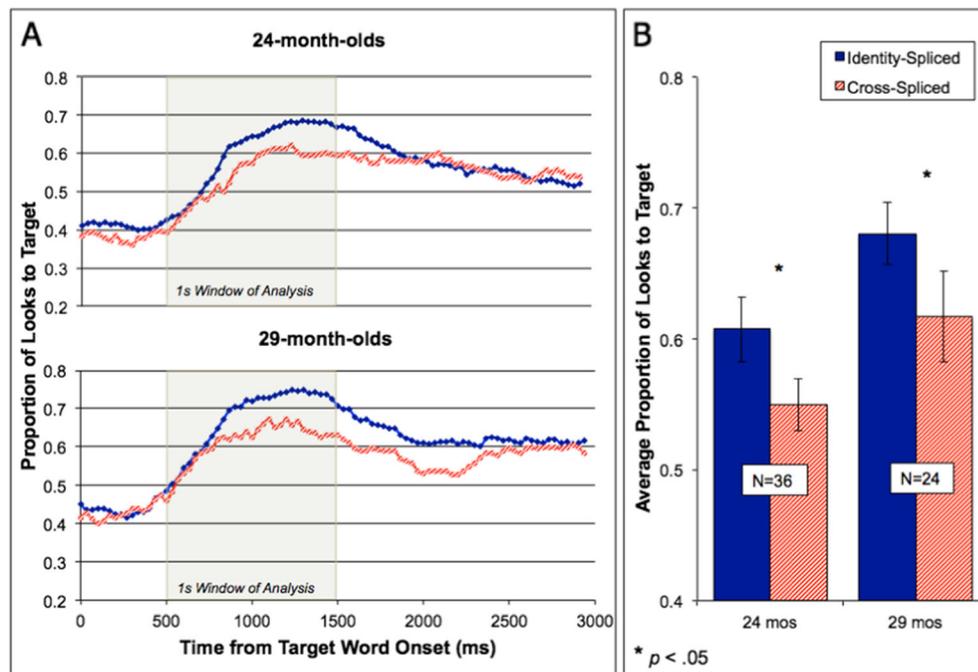


Fig. 2 Panel A shows the mean proportion of looks to the target after target word onset (0 ms) in the Identity-spliced and Cross-spliced trials for the 24-month-olds and 29-month-olds. Panel B shows the mean

proportion of looks to the target in the 1 s window of analysis beginning 500 ms after word onset for both trial types

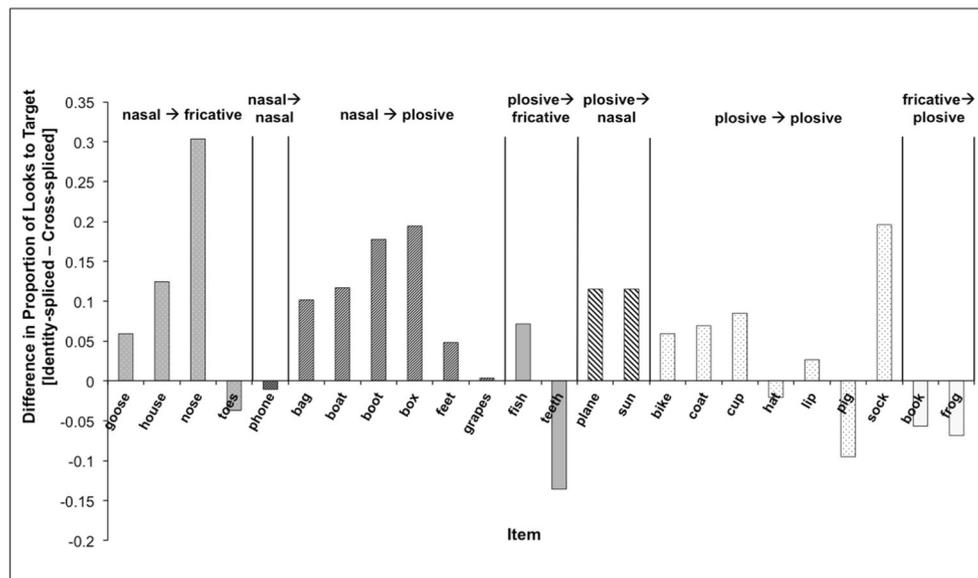


Fig. 3 Mean difference between looks to target in the Identity-spliced versus Cross-spliced trials by item. *Note:* This experiment was not originally designed to look at the effects of individual items. Each bar represents the difference between the looks to the target for the two-thirds

of participants that heard the identity-spliced token ($n = 40$) minus the looks to the target for the third that heard the cross-spliced token of that item ($n = 20$)

one novel object. We reasoned that if children perceived the labels provided in the Phonemic Mismatch trials to be an unacceptable pronunciation of a familiar word, then they should look to the novel object as a possible referent. For example, if children saw a boat and a novel object, and heard the label *bait*, then they might look to the novel object as a possible referent for the word *bait* (assuming they do not know the word *bait*; see White & Morgan, 2008, for use of a similar design to examine children's sensitivity to phonemic mispronunciations). We predicted that if children are categorizing subphonemic mismatches in an adult-like manner, then word recognition should be more disrupted when target words contain a phonemic mismatch compared to a subphonemic mismatch. However, if children's representations are overspecified, then a subphonemic mismatch may be just as disruptive as a phonemic mismatch.

Method

Participants Twenty-four 29-month-old ($M_{\text{age}} = 893$ days; range = 851–914; 11 females) monolingual English-learning children were tested (all had at least 90 % English input). The data from three additional participants were excluded from the study before coding due to a diagnosed language difficulty (1) and fussiness (2).

Materials, apparatus, and procedure The targets and their subphonemic and phonemic splicing pairs were recorded by the same female monolingual English speaker who recorded the materials for Experiment 1.

A subset of the target words from Experiment 1 were chosen to be targets in Experiment 2, and the remaining words were used as fillers. Rather than reusing a subset of the stimuli recorded for Experiment 1, the entire set of tokens was re-recorded in a single recording session to ensure that all Experiment 2 materials were matched in recording quality. The subphonemic mismatches were spliced in the same way as they were in Experiment 1. The phonemic mismatches were created by splicing the coda of a target word onto the onset and nucleus of a word with a different vowel (e.g., the final consonant of *boat* was spliced onto the initial consonant and vowel of the word *bait*). The words selected for the phonemic mismatches were either nonsense words or words that are unlikely to be known by 29-month-olds. For a complete list of target words and their phonemic and subphonemic splicing pairs, see Appendix 2.

Similar to Experiment 1, the visual stimuli consisted of 12 pairs of still images presented side-by-side on a white background. To test whether the mismatches were prominent enough to signal a novel label, the subphonemic and phonemic targets were depicted alongside an image of a novel object (e.g., a garlic press) instead of a familiar object. Filler trials consisted of two images of known objects.

Design Four experimental lists were created, each containing five Phonemic (vowel) Mismatch trials, five Subphonemic (coarticulatory) Mismatch trials, and 14 filler trials. The assignment of words to the Phonemic Mismatch versus Subphonemic Mismatch trials was counterbalanced across

lists. Each participant heard the 10 target items once with either a phonemic mismatch or a subphonemic mismatch.

Coding and analysis Coding was done in the same manner as in Experiment 1. Four randomly selected videos were recoded by a second coder, and reliability was high (Mean $r = .98$, $SD = .02$).

Results

We know that children have a strong tendency to fixate the known object when there is a known and a novel object on the screen (Schafer, Plunkett, & Harris, 1999; White & Morgan, 2008). Thus, similar to White and Morgan (2008), we accounted for these strong baseline preferences by comparing the looks to the target in the baseline time period (in this case, the 3 s window before word onset) to the 3 s time period after word onset (critical window) for both of our trial types (Subphonemic vs. Phonemic Mismatch). We ran a weighted empirical-logit regression in a linear mixed effects model (Barr et al., 2011) using the lme4 package of the statistical software R 3.2.2 (Bates et al., 2015; R Development Core Team, 2015). Before running the model, we deviation coded the independent variables Time Window (-1: baseline, 1: after word onset), Trial Type (-1: Subphonemic Mismatch, 1: Phonemic Mismatch). The model included Time Window, Trial Type and the Time Window \times Trial Type interaction as fixed effects. Following the Barr et al. (2013) paper on model selection, we used a maximal structure of random effects, including random intercept and Time Window, Trial Type, and Time Window \times Trial Type slopes for participants as well as for items. For the fixed effects we report b , standard error, t tests, and p values calculated using Satterthwaite approximations to degrees of freedom and implemented using the lmerTest package (Kuznetsova et al., 2015). Because the Time Window (before and after word onset) was entered into the model, we were no longer interested in the main effect of Trial Type but rather the interaction between Trial Type and Time Window. As expected, there was a main effect of Time Window, $b = 0.17$, $SE = 0.06$, $t(8.95) = 2.82$, $p = .020$, and Trial Type, $b = -0.17$, $SE = 0.05$, $t(15.02) = -3.53$, $p = .003$. Most importantly, there was a significant interaction between Time Window and Trial Type, $b = -0.10$, $SE = 0.04$, $t(18.80) = -2.20$, $p = .041$. The interaction indicates that children increased their looks more to the target in the critical window after word onset when there was a subphonemic compared to a phonemic mismatch (see Figs. 4 and 5). This supports our hypothesis that although a coarticulatory mismatch is noticeable, it does

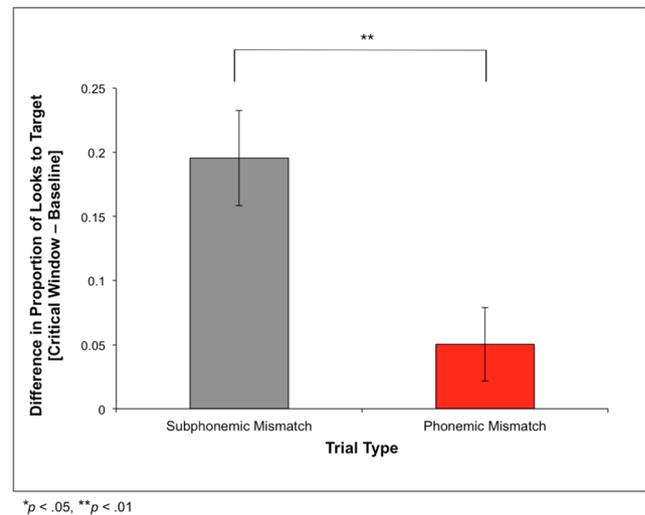


Fig. 4 Increase from baseline in the proportion of looks to target for the Subphonemic Mismatch and Phonemic Mismatch trials

not disrupt children's word recognition as much as a phonemic mismatch. To determine if children were looking at the novel distractor more during the Phonemic Mismatch trials, the weighted empirical-logit regression in a linear mixed effects model described above was rerun with looks to the distractor as the dependent variable (instead of looks to the target). Here, we found no main effect of Time Window, $b = 0.04$, $SE = 0.06$, $t(9.58) = 0.71$, $p = .495$, but there was a main effect of Trial Type, $b = 0.16$, $SE = 0.04$, $t(13.25) = 3.73$, $p = .002$. Most importantly, there was a significant interaction between Time Window and Trial Type, $b = 0.12$, $SE = 0.05$, $t(11.86) = 2.55$, $p = .026$, indicating that children looked significantly more to the distractor when there was a phonemic mismatch compared to when there was a coarticulatory mismatch.

General discussion

In this study, we asked two questions: (1) are 2-year-olds sensitive to inappropriate coarticulation (i.e., subphonemic mismatch) during online word recognition (Experiment 1), and (2) if so, how does the effect of a subphonemic mismatch compare to the effect of a phonemic mismatch (Experiment 2)? Our results clearly indicate that by 24 months of age, children are already sensitive to subphonemic mismatches during online word recognition. Moreover, we have shown that although 2-year-olds readily detect a subphonemic mismatch in the speech signal, this sort of mismatch does not disrupt word recognition nearly as much as a phonemic mismatch. These findings lead us to conclude

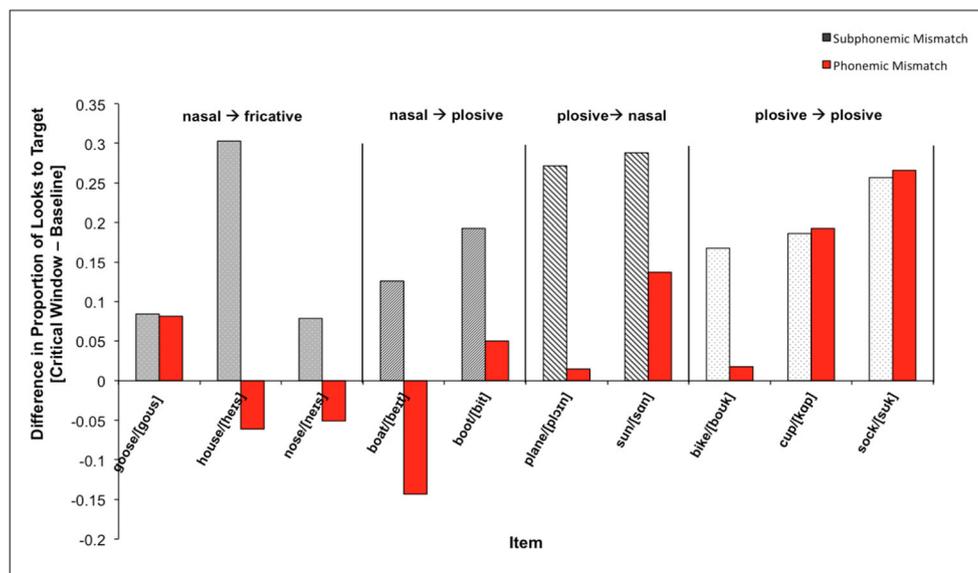


Fig. 5 Increase from baseline in the proportion of looks to target after word onset for the Subphonemic Mismatch and Phonemic Mismatch trials by item. Note that children heard each word produced with either

a subphonemic or a phonemic mismatch, thus, each bar represents the mean increase in the proportion of looks to target in the subset of the sample ($n = 12$) that heard that particular token

that toddlers may already process subphonemic information in the speech signal in a relatively mature manner.

Much work in the area of developmental speech perception has been aimed at understanding when and how children learn to focus their attention on the speech sounds that signal lexical contrasts in their native language (e.g., Narayan et al., 2010; Werker & Tees, 1984). Indeed, learning to “ignore” contrasts that do not signal lexical differences is often seen as a crucial step towards acquiring the native language phonology, but adult research has shown that subphonemic detail in the speech signal carries very useful information that can facilitate rapid decoding of the speech signal (e.g., Shatzman & McQueen, 2006). Thus, it seems reasonable to ask how well and when in development children detect this information during online word recognition. In Experiment 1, we examined 2-year-olds’ sensitivity to a specific type of subphonemic detail: coarticulatory mismatch (e.g., the sort of mismatch that occurs when an anticipatory velar gesture for a nasal consonant is present in a vowel, but no nasal consonant follows). We predicted that sensitivity to subphonemic detail might increase with age. However, we found that both 24- and 29-month-olds readily detected subphonemic mismatch in familiar words, suggesting that children’s sensitivity to subphonemic mismatch is in place long before they develop an extensive lexicon.

Given these results, one could speculate that sensitivity to coarticulatory mismatch may be present from birth. That is, infants may have an inborn understanding of how speech articulators are generally coordinated when speakers vocalize. At the same time, one could also speculate that sensitivity to coarticulatory mismatch is (at least partially) driven by

experience listening to the speech signal. Perhaps we would have seen a change in sensitivity to subphonemic mismatch over the course of development if we had tested slightly younger children (e.g., 18-month-olds rather than 24-month-olds). It is also possible that children might demonstrate greater sensitivity to coarticulatory mismatch in high frequency words than in newly learned words. If this were the case, then perhaps vocabulary size might have been a better predictor of sensitivity to subphonemic detail than age (e.g., Van Heugten, Krieger, & Johnson, 2015). However, when we examine the relationship between children’s sensitivity to coarticulatory mismatch and the size of children’s vocabulary in Experiment 1, we find no support for this hypothesis, $r(56) = .02$, $p = .874$.³

Although the results of Experiment 1 suggest that children have well-specified representations of familiar words, these representations are not necessarily adult-like (see Singh et al., 2008, for a discussion of possible overspecification in early lexical representations). Indeed, it is possible that children’s representations could be overspecified, such that inappropriate coarticulation may disrupt word recognition as much as, for example, a familiar word spliced with an inappropriate vowel. If this were the case, overattention to subphonemic detail could actually slow down word recognition by toddlers. In Experiment 2, we addressed this issue by presenting 2-year-olds with labels that contained either a phonemic mismatch (i.e., containing the wrong vowel or diphthong, such as *bait* for *boat*) or a subphonemic mismatch. Our results demonstrate

³ Three children were excluded from this analysis because their caregivers did not provide a vocabulary form.

that word recognition was far more disrupted when the target word contained a phonemic mismatch than when it contained a subphonemic mismatch, suggesting that children may treat noncontrastive subphonemic changes as less of a deviation from the canonical pronunciation of a word than a phonemic mispronunciation. This finding supports the notion that 2-year-olds may be using subphonemic information in an adult-like manner as the speech signal unfolds.

A key methodological difference between Experiment 1 and Experiment 2 was that the former experiment presented children with two known objects whereas the latter presented them with one known object and one novel object. Experiment 2 was designed in this fashion so that we could test whether children would consider words with phonemic (but not subphonemic) mismatches as labels for the novel object rather than just unusual pronunciations of familiar words. Indeed, we found that when children heard a phonemic mismatch, their looks to the novel distractor increased, whereas when they heard a subphonemic mismatch, their looks to the distractor decreased. This is evidence that although the subphonemic mismatches hindered children's recognition of the words, they were not enough of a deviation from the canonical pronunciation to elicit looks to the novel distractor. However, when children heard a phonemic mismatch, they looked more towards the novel distractor, indicating that these types of mismatches might be perceived as novel words.

Based on our findings from Experiments 1 and 2, we conclude that 2-year-olds likely possess adult-like sensitivity to coarticulatory mismatch. However, additional work could be done to more fully support this assertion. For example, adult studies have shown both the target label and its splicing partner present in the visual array at the same time (Beddor, McGowan, Boland, Coetzee, & Brasher, 2013; Dahan et al., 2001). In our study, the splicing partner was never shown on the screen (e.g., children were shown a boat and a book, not a boat and a bone, when hearing a cross-spliced token consisting of the initial CV of *bone* and the final C of *boat*). Future research is needed to examine whether children behave in the same way when the splicing partner is visually present on the screen. Another aspect of our study that differs from many adult studies is that our study was not designed to ask whether the lexical status of the splicing pair might impact children's behavior in the same way that it impacts adult behavior. Adult studies have shown that due to lexical competition, subphonemic mismatches are more disruptive to word recognition if the target is cross-spliced with another real word rather than a nonsense word (e.g., *neck* is harder to recognize when spliced with *net* than when spliced with *nep*; Dahan

et al., 2001). Although our splicing pairs were all real words in English, many of them were likely unknown by our 2-year-old participants. Thus, although our study was not designed to examine how lexical status of the splicing partner impacts sensitivity to coarticulatory mismatch, we can at least investigate whether our findings might support the idea that children behave like adults in this respect. Based on parental report, we identified the four targets cross-spliced with words that the children in our study were most likely to know (i.e., *bite*, *cut*, *lid*, *bone*), and the 10 targets cross-spliced with words that the children in our study were least likely to know (i.e., *teak*, *plague*, *goon*, *fiend*, *sod*, *ban*, *hound*, *gnome*, *hack*, *tone*). The remaining words, that parents gave the most mixed responses for in terms of whether their children knew them or not, were excluded from the analysis (e.g., *Coke*). This admittedly post-hoc analysis revealed no evidence that the subphonemic mismatches hindered 2-year-olds' performance more when the nontarget word used for cross-splicing was known versus when it was not known, $t(59) = 0.37$, $p = .710$. Interestingly, however, when we limited our analysis to the 29-month-olds, we found a trend in the expected direction, with recognition of targets cross-spliced with familiar words being harder to recognize than those cross-spliced with unfamiliar words, $t(23) = 1.83$, $p = .080$.⁴ Thus, one could speculate that the processing of subphonemic mismatch may be more adult-like in older children. Future studies should explore this possibility with stimuli specifically designed to address this question (e.g., Do children recognize *bike* faster when it is cross-spliced with the nonword *bipe* than when it is cross-spliced with the known word *bite*?).

To conclude, our study is the first to examine toddlers' sensitivity to subphonemic versus phonemic mismatch during online word recognition. Contrary to Minaudo and Johnson (2013), our findings provide support for the claim that 2-year-olds use coarticulatory information to facilitate online word recognition (see also Mahr et al., 2015). This study further shows that word recognition in 2-year-olds is far more disrupted by a phonemic mismatch than a subphonemic mismatch, supporting the notion that children's sensitivity to coarticulatory mismatch is fairly mature early on. An important goal for future work will be to better understand how infants and toddlers learn the information status of subphonemic patterns in speech, and how this information is handled by the emerging proto-lexicon.

⁴ Paired-sample t tests were run using empirical-logit transformed data from Experiment 1. Tests compared the mean difference between identity-spliced and cross-spliced tokens in trials where the splicing partner was known compared to when it was unknown.

Appendix 1

Table 1 Experiment 1 stimuli

Target item	Cross-spliced pair onset→ offset	Type of mismatch	Duration cross-spliced target (s)	Duration identity-spliced target (s)	Distractor item
goose	goo(n) → (goo) se	nasal → fricative	0.97	0.82	grapes
house	hou(nd) → (hou) se	nasal → fricative	0.82	0.7	hat
nose	gno(me) → (no) se	nasal → fricative	0.65	0.72	lips
toes	to(ne) → (to) es	nasal → fricative	0.8	1.03	teeth
phone	foa(m) → (pho) ne	nasal → nasal	0.64	0.6	feet
bag	ba(n) → (ba) g	nasal → plosive	0.63	0.65	box
boat	bo(ne) → (boa) t	nasal → plosive	0.65	0.57	book
boot	boo(m) → (boo) t	nasal → plosive	0.56	0.46	bike
box	bo(nd) → (bo) x	nasal → plosive	0.93	0.83	bag
feet	fie(nd) → (fee) t	nasal → plosive	0.81	0.71	phone
grapes	grai(ns) → (gra) pes	nasal → plosive	0.89	0.8	goose
fish	fi(g) → (fi) sh	plosive → fricative	0.67	0.58	frog
teeth	tea(k) → (tee) th	plosive → fricative	0.64	0.64	toes
plane	pla(gue) → (pla) ne	plosive → nasal	0.65	0.65	pig
sun	su(b) → (su) n	plosive → nasal	0.61	0.52	sock
bike	bi(te) → (bi) ke	plosive → plosive	0.72	0.71	boot
coat	co(ke) → (coa) t	plosive → plosive	0.5	0.55	cup
cup	cu(t) → (cu) p	plosive → plosive	0.63	0.68	coat
hat	ha(ck) → (ha) t	plosive → plosive	0.56	0.55	house
lips	li(ds) → (li) ps	plosive → plosive	0.71	0.68	nose
pig	pi(t) → (pi) g	plosive → plosive	0.44	0.54	plane
sock	so(d) → (so) ck	plosive → plosive	0.73	0.62	sun
book	bu(sh) → (boo) k	fricative → plosive	0.50	0.51	boat
frog	fro(st) → (fro) g	fricative → plosive	0.66	0.71	fish

Appendix 2

Table 2 Experiment 2 stimuli

Target item	Subphonemic splicing pair onset→offset	Type of mismatch	Duration subphonemic target (s)	Phonemic splicing pair onset→offset	Duration phonemic target(s)	Novel distractor item
goose	goo(n) → (goo) se	nasal → fricative	0.79	/goo(s)/ → (goo) se	0.72	gourd
house	hou(nd) → (hou) se	nasal → fricative	0.85	/her(s)/ → (hou) se	0.83	ski binding
nose	gno(me) → (no) se	nasal → fricative	0.83	/ner(s)/ → (no) se	1.05	belay device
boat	bo(ne) → (boa) t	nasal → plosive	0.69	/ber(t)/ → (boa) t	0.69	garlic press
boot	boo(m) → (boo) t	nasal → plosive	0.76	/bi(t)/ → (boo) t	0.73	typewriter
plane	pla(gue) → (pla) ne	plosive → nasal	1.03	/plɒ(n)/ → (pla) ne	1.15	corkscrew
sun	su(b) → (su) n	plosive → nasal	0.56	/sa(n)/ → (su) n	0.69	scuba flippers
bike	bi(te) → (bi) ke	plosive → plosive	0.76	/bo(k)/ → (bi) ke	0.76	cassette tape
cup	cu(t) → (cu) p	plosive → plosive	0.64	/ka(p)/ → (cu) p	0.80	hard drive
sock	so(d) → (so) ck	plosive → plosive	0.83	/so(k)/ → (so) ck	0.66	waffle iron

References

- Barr, D. J., Gann, T. M., & Pierce, R. S. (2011). Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, 137, 201–207. doi:10.1016/j.actpsy.2010.09.011
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H.,...Grothendieck, G. (2015). Linear mixed-effects models using “Eigen” and S4 [Computer software]. Retrieved from <http://lme4.r-forge.r-project.org/>
- Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *Journal of the Acoustical Society of America*, 133, 2350–2366. doi:10.1121/1.4794366
- Boersma, P., & Weenink, D. (2016). *Praat: Doing phonetics by computer*. Retrieved from <http://www.praat.org>
- Creel, S. C. (2012). Preschoolers’ use of talker information in on-line comprehension. *Child Development*, 83, 2042–2056. doi:10.1111/j.1467-8624.2012.01816.x
- Curtin, S., Mintz, T. H., & Byrd, D. (2001). Coarticulatory cues enhance infants’ recognition of syllable sequences in speech. In A. H. J. Do, L. Dominguez, & A. Johansen (Eds.), *Proceedings of the 25th Annual Boston University Conference on Language Development (Vol. 1)* (pp. 191–201). Somerville, MA: Cascadilla Press.
- Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16, 507–534. doi:10.1080/01690960143000074
- Development Core Team, R. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Vienna.
- Dietrich, C., Swingle, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 16027–16031. doi:10.1073/pnas.0705270104
- Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates communicative development inventories: User’s guide and technical manual* (2nd ed.). Baltimore, MD: Paul H. Brookes.
- Fernald, A., Pinto, J. P., Swingle, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in the speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9, 228–231. doi:10.1111/1467-9280.00044
- Fernald, A., Swingle, D., & Pinto, J. P. (2001). When half a word is enough: Infants can recognize spoken words using partial phonetic information. *Child Development*, 72, 1003–1015. doi:10.1111/1467-8624.00331
- Fisher, C., Hunt, C., Chambers, K., & Church, B. (2001). Abstraction and specificity in preschoolers’ representations of novel spoken words. *Journal of Memory and Language*, 45, 665–687. doi:10.1006/jmla.2001.2794
- Fowler, C. A., Best, C. T., & McRoberts, G. W. (1990). Young infants’ perception of liquid coarticulatory influences on following stop consonants. *Perception & Psychophysics*, 48, 559–570. doi:10.3758/BF03211602
- Houston, D. M. (2011). Infant speech perception. In R. Seewald & M. Tharpe (Eds.), *Comprehensive handbook of pediatric audiology* (pp. 47–62). San Diego, CA: Plural.
- Johnson, E. K. (2003). *Word segmentation during infancy: The role of subphonemic cues to word boundaries* (Unpublished doctoral dissertation). The Johns Hopkins University, Baltimore, MD.
- Johnson, E. K. (2016). Constructing a proto-lexicon: An integrative view of infant language development. *Annual Review of Linguistics*, 2, 391–412. doi:10.1146/annurev-linguistics-011415-040616
- Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language*, 44, 548–567. doi:10.1006/jmla.2000.2755
- Kuhl, P. K., Conboy, B. T., Padden, D., Nelson, T., & Pruitt, J. (2005). Early speech perception and later language development: Implications for the “critical period”. *Language Learning and Development*, 1, 237–264. doi:10.1207/s154733411ld0103&4_2
- Kuznetsova, A., Brockhoff, B., & Christensen, R. H. B. (2015). Tests in linear mixed effects models [Computer software]. Retrieved from <https://cran.r-project.org/web/packages/lmerTest/index.html>
- Mahr, T., Mcmillan, B. T. M., Saffran, J. R., Weismer, S. E., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, 142, 345–350. doi:10.1016/j.cognition.2015.05.009
- McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition*, 95(2), B15–B26. doi:10.1016/j.cognition.2004.07.005
- McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, 86(2), B33–B42. doi:10.1016/S0010-0277(02)00157-9
- McQueen, J. M. (2007). Eight questions about spoken-word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics* (pp. 37–53). Oxford, UK: Oxford University Press. doi:10.1093/oxfordhb/9780198568971.013.0003
- McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1363–1389. doi:10.1037/0096-1523.25.5.1363
- Minado, C., & Johnson, E. (2013). Are two-year-olds sensitive to anticipatory coarticulation? *Proceedings of Meetings on Acoustics* (Vol. 19). doi:10.1121/1.4800296
- Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, 13, 407–420. doi:10.1111/j.1467-7687.2009.00898.x
- Schafer, G., Plunkett, K., & Harris, P. L. (1999). What’s in a name? Lexical knowledge drives infants’ visual preferences in the absence of referential input. *Developmental Science*, 2(2), 187–194. doi:10.1111/1467-7687.00067
- Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics*, 68, 1–16. doi:10.3758/BF03193651
- Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, 4, 157–178. doi:10.1080/15475440801922131
- Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, 48, 233–254. doi:10.1016/S0749-596X(02)00513-2

- Swingle, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language, 60*, 252–269. doi:10.1016/j.jml.2008.11.003.Onsets
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*, 147–166. doi:10.1016/S0010-0277(00)00081-0
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science, 13*, 480–484. doi:10.1111/1467-9280.00485
- Tsao, F.-M., Liu, H.-M., & Kuhl, P. K. (2004). Speech perception in infancy predicts language development in the second year of life: A longitudinal study. *Child Development, 75*, 1067–1084. doi:10.1111/j.1467-8624.2004.00726.x
- van Heugten, M., Krieger, D. R., & Johnson, E. K. (2015). The developmental trajectory of toddlers' comprehension of unfamiliar regional accents. *Language Learning and Development, 11*, 41–65. doi:10.1080/15475441.2013.879636
- Werker, J., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*, 197–234. doi:10.1207/s15473341l1d0102_4
- Werker, J. F., & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology, 66*, 173–196. doi:10.1146/annurev-psych-010814-015104
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49–63. doi:10.1016/S0163-6383(84)80022-3
- Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics, 50*, 351–360. doi:10.3758/BF03212227
- White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language, 59*, 114–132. doi:10.1016/j.jml.2008.03.001
- Zamuner, T.S., Moore, C., & Desmeules-Trudel, F. (2016). Toddler's sensitivity to within-word coarticulation during spoken word recognition: Developmental differences in lexical competition. *Journal of Experimental Child Psychology*. doi:10.1016/j.jecp.2016.07.012