

Mahr, T., McMillan, B. T. M., Saffran, J. R., Ellis Weismer, S., & Edwards, J. (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, 142, 345–350.  
<http://dx.doi.org/10.1016/j.cognition.2015.05.009>

This is the “Accepted Manuscript” version of the article.

Raw data and accompanying analysis scripts: [https://github.com/tjmahr/2015\\_Coartic](https://github.com/tjmahr/2015_Coartic)

## Anticipatory coarticulation facilitates word recognition in toddlers

Tristan Mahr, Brianna T. M. McMillan, Jenny R. Saffran, Susan Ellis Weismer, and Jan Edwards

University of Wisconsin–Madison

## **Abstract**

Children learn from their environments and their caregivers. To capitalize on learning opportunities, young children have to recognize familiar words efficiently by integrating contextual cues across word boundaries. Previous research has shown that adults can use phonetic cues from anticipatory coarticulation during word recognition. We asked whether 18–24 month-olds ( $n = 29$ ) used coarticulatory cues on the word "the" when recognizing the following noun. We performed a looking-while-listening eyetracking experiment to examine word recognition in neutral versus facilitating coarticulatory conditions. Participants looked to the target image significantly sooner when the determiner contained facilitating coarticulatory cues. These results provide the first evidence that novice word-learners can take advantage of anticipatory sub-phonemic cues during word recognition.

## **Keywords**

Word recognition; coarticulation; eye tracking; word learning; lexical development; language development

## **Highlights**

- We report a looking-while-listening eyetracking study with 18–24 month-olds.
- We manipulated the coarticulatory cues on the word “the”.
- Under facilitating coarticulation, the cues predicted the following noun.
- Looking patterns were compared for facilitating vs. neutral coarticulation.
- Toddlers looked to target sooner when “the” contained facilitating coarticulation.

## 1. Introduction

To learn from their environment, young children must be able to process familiar words efficiently. Word recognition mediates toddlers' ability to learn words from caregivers (Weisleder & Fernald, 2013), and efficiency of lexical processing during the first two years predicts vocabulary and working memory later in childhood (Marchman & Fernald, 2008). Grammatical, pragmatic and phonetic contextual cues can constrain word recognition by simplifying the search space, but many such cues to word identification are not word-internal. Therefore, integrating contextual cues across word boundaries is essential for efficient word recognition.

One of the most well established context-sensitive phenomena in phonetics is *coarticulation*: the overlap of articulatory gestures in neighboring sounds. Coarticulation influences the production of sound patterns both within and across word boundaries. Typical English examples include coronal place assimilation (e.g., saying *in case* with a velar nasal consonant) and fronting of /k/ in *keep* (cf. backing and lip-rounding on /k/ in *coop*). A coarticulated sound carries acoustic information about neighboring sounds, introducing redundant and locally coherent information into the speech signal. In this respect, coarticulation provides regularity or "lawful variability" that can support speech perception (Elman & McClelland, 1986).

Indeed, adult listeners access and exploit coarticulatory cues during speech perception and word recognition (Gow, 2002; Gow & McMurray, 2007). Adults are slower to recognize words when there is a mismatch between coarticulatory cues in a vowel and the following consonant (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; McQueen, Norris, & Cutler, 1999; Tobin, Cho, Jennet, & Magnuson, 2010). Conversely, appropriate coarticulation can

facilitate spoken word recognition (e.g., Mattys, White, & Melhorn, 2005). For example, adult English listeners are faster to recognize a noun when the preceding determiner *the* carries information about the onset of the noun (Salverda, Kleinschmidt, & Tanenhaus, 2014).

It is not known whether young children can take advantage of coarticulatory cues during word recognition. Toddlers encode subsegmental details in their lexical representations (Fisher, Church, & Chambers, 2004), so coarticulatory cues should be accessible to these listeners in principle. In addition, toddlers recognize spoken words incrementally, using acoustic cues as they become available as a word unfolds over the speech signal (e.g., Fernald, Swingley, & Pinto, 2001; Swingley, Pinto, & Fernald, 1999). Moreover, toddlers rely on contextual cues when recognizing words produced in fluent speech (Plunkett, 2006). These findings raise an important question: Can young listeners use coarticulatory cues to facilitate recognition of a following word?

This question is important given the longstanding debate concerning the nature of early phonological representations. One point of view holds that these representations are under-specified and that children differentiate between words using relatively holistic phonological representations (Charles-Luce & Luce, 1990, 1995; Jusczyk, 1993). Based on a corpus analysis, Charles-Luce and Luce argued that young children do not need the same phonological detail in their lexical representations as adults do because children's phonological neighborhoods are much sparser. Researchers supporting this point of view have hypothesized that children's phonological representations gradually become more detailed as vocabulary size increases (Edwards, Beckman, & Munson, 2004; Metsala, 1999; Werker & Curtin, 2005; Werker et al., 2002). An opposing point of view posits that children's phonological representations are segmental from very early in development (Dollaghan, 1994; Magnuson, Tanenhaus, Aslin, &

Dahan, 2003). This view is supported by studies showing that infants are sensitive to one-feature mispronunciations of familiar words (e.g., Swingley & Aslin, 2000, 2002; K. S. White & Morgan, 2008; see also review in Mayor & Plunkett, 2014). If toddlers use anticipatory coarticulation for word recognition, this finding would provide additional support for the viewpoint that children's phonological representations are well specified even when their vocabularies are relatively small.

In the present study, we investigated whether toddlers took advantage of sub-phonemic anticipatory coarticulatory cues between words. Specifically, we asked whether coarticulatory acoustic cues on the determiner *the* facilitate recognition of the following word. We used a looking-while-listening task (Fernald, Zangl, Portillo, & Marchman, 2008) to determine whether toddlers looked more quickly to a named image in facilitating versus neutral coarticulatory contexts (manipulated within subjects). Crucially, all of the items were cross-spliced to ensure that the recordings were otherwise comparable. We hypothesized that if toddlers are sensitive to coarticulation, we should see earlier recognition of the target noun in facilitating contexts relative to neutral contexts.

## **2. Method**

### *2.1 Participants*

Twenty-nine 18 to 24-month-olds ( $M = 20.8$ , range = 18.1–23.8, 13 male) participated in this study. An additional 11 toddlers were excluded from the analyses due to inattentiveness (10) or having more than 50% missing data during non-filler trials (1). Caregivers completed the short version of the Words and Sentences Form of the MacArthur-Bates Communicative Development Inventory (MBCDI; Fenson et al., 2007).

## 2.2 Materials and Stimuli

We selected target words that are familiar to toddlers in this age group. For the facilitating and neutral items, we presented /d/- and /b/-initial words in yoked pairs: *duck-ball* and *dog-book*. To help maintain interest in the task, we also included filler trials: *cup-sock*, *car-cat*, *cookie-shoe*. Target words were presented in carrier phrases (e.g., *find the \_\_\_\_* or *see the \_\_\_\_*). The durations of the target words ranged from 560 to 850 ms. All stimuli were recorded using child-directed speech.

We manipulated whether the determiner *the* provided coarticulatory cues for the target word by cross-splicing different tokens of *the* with the target words. In the FACILITATING coarticulation items, the determiner contained coarticulatory cues for the subsequent word-onset (/d/ or /b/), as shown in Figure 1. These cues involved bilabial formant transitions for /b/ (i.e., falling first and second formants) and alveolar formant transitions for /d/ (i.e., falling first formant and rising second formant). In the NEUTRAL coarticulation items, the determiner token came from the phrase *the hut*. We used this context because the sequence [əhə] would not provide any coarticulatory cues for /b/ or /d/. Indeed, the vowel in this token showed steady first and second formants, as depicted in Figure 1. The child-directed determiner tokens were approximately 510 ms in duration. They were padded with silence following the vowel so that the duration of each token was 600 ms. Stimuli were cross-spliced in both conditions so that three tokens of *the* were used (i.e., *the*<sup>b</sup> or *the*<sup>d</sup> in facilitating trials, or *the*<sup>ə</sup> in neutral trials). Filler words were also cross-spliced and presented with the neutral determiner token. Stimuli were normalized for amplitude.

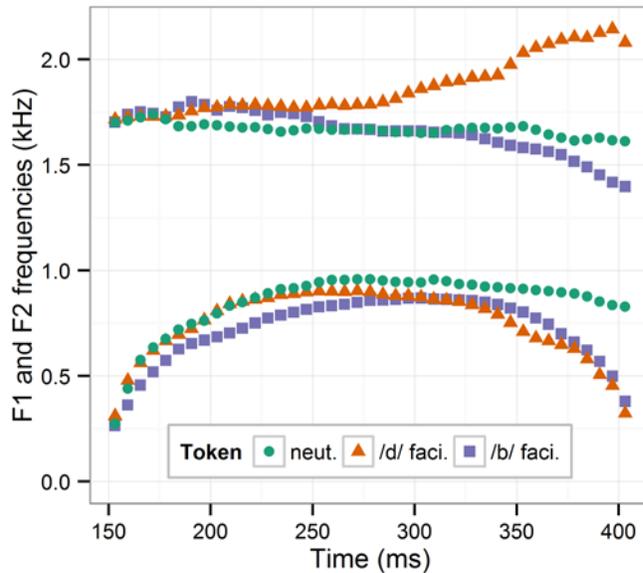


Figure 1. F1 and F2 formants of the tokens of the determiner *the* for the three item types. Note the canonical formant transitions for the facilitating tokens and the steady formant values in the neutral token.

The images were color photographs of the familiar objects, presented in  $600 \times 600$  pixel images over a gray background. The two images were presented on a black computer display with a  $1920 \times 1200$  pixel resolution. The left and right images were centered at pixels (400, 600) and (1520, 600), respectively. The four target words were presented in yoked pairs: *ball-duck* and *book-dog*. Four sets of images were used for each pair.

### 2.3 Procedure and design

The study was conducted on a Tobii T60 XL eyetracker controlled by E-Prime Professional 2.0. Images were counterbalanced for screen-location (left vs. right side), and trials

were pseudorandomly ordered so that filler trials would occur every third trial. Each subject viewed 32 experimental trials (16 of each condition) plus 14 filler trials over two blocks. Before each block, the eyetracker was calibrated using five locations on the screen (four corners plus center). Between blocks, the child watched a short cartoon on the display. During the experiment, a brief animation played onscreen every six trials to keep the child engaged with the task. The child sat on his or her caregiver's lap approximately 60 cm away from the screen. The caregiver's eyes were obscured during the experiment.

We used a gaze-contingent stimulus presentation. First, both images appeared onscreen in silence for 1500 ms. Next, the experiment procedure verified that the child's gaze was being tracked: if the child's gaze was continuously tracked for at least 300 ms, the verbal prompt played. Otherwise, the prompt (e.g., *find the dog*) eventually played after 10 seconds of trying to verify gaze tracking. A reinforcer phrase (e.g., *Look at that!*) played 1000 ms after the end of the verbal prompt followed by 1000 ms of silence. The images then disappeared from the screen during a 500-ms inter-trial interval.

#### *2.4 Data analysis*

The eyetracker recorded the x-y locations of the participant's gaze at a rate of 60 Hz. Gaze coordinates from each eye were averaged together, and these averages were mapped onto the regions of interest (i.e., target and distractor images). We interpolated short windows of missing data (up to 150 ms) if the participant had fixated on the same area of interest immediately before and after the span of missing data.

For our dependent variable for gaze location, we used an empirical logit transformation on the number of looks to the target and distractor images (Barr, 2008; Mirman, 2014). We first down-sampled the eyetracking data into 50-ms bins and added up looks to the target and looks to

the distractor image in each bin. The empirical logit is simply the log-odds of looking to the target image in each bin with 0.5 added to numerator and denominator to avoid division with 0:  $\log((\text{looks to target} + .5) / (\text{looks to distractor} + .5))$ . Empirical logits in each bin were weighted following Barr (2008).

We used weighted empirical-logit growth curve analysis (Mirman, 2014) to model how the probability of fixating on the target image changed over time and under different coarticulatory conditions. Time was modeled using linear, quadratic and cubic orthogonal polynomials. Condition was coded using the neutral condition as the reference level, so that the condition parameters described how the growth curve in the facilitating condition differed from the neutral condition. Models included participant and participant-by-condition random effects, because we expected participants to vary randomly in their ability to access coarticulatory information. Analyses were performed in R (vers. 3.1.3) with the `lme4` package (vers. 1.1.7; Bates, Mächler, Bolker, & Walker, 2014). Because it is computationally and theoretically difficult to estimate the degrees of freedom for mixed-effects models, we analyzed *t*-scores assuming a Gaussian distribution (i.e.,  $t > \pm 1.96$  was considered significant). Raw data and analysis scripts are available at [https://github.com/tjmahr/2015\\_Coartic](https://github.com/tjmahr/2015_Coartic).

### **3. Results**

Overall looking patterns are presented in Figure 2. Accuracy hovers around chance performance over the course of the determiner and approximately 250 ms into the target word. Accuracy increases from 250 to 1000 ms, and after 1000 ms accuracy begins to plateau then decline. Time clearly predicts accuracy; the probability of looking to target increases as the word unfolds. Importantly, coarticulatory information also predicts accuracy because participants have a noticeable head-start on the facilitating items. In order to formally estimate how these growth

curves differ, we modeled the portion of the data from 200 to 1000 ms after word-onset. We chose this window because the facilitating curve steadily grows from 200 to 1000 ms.

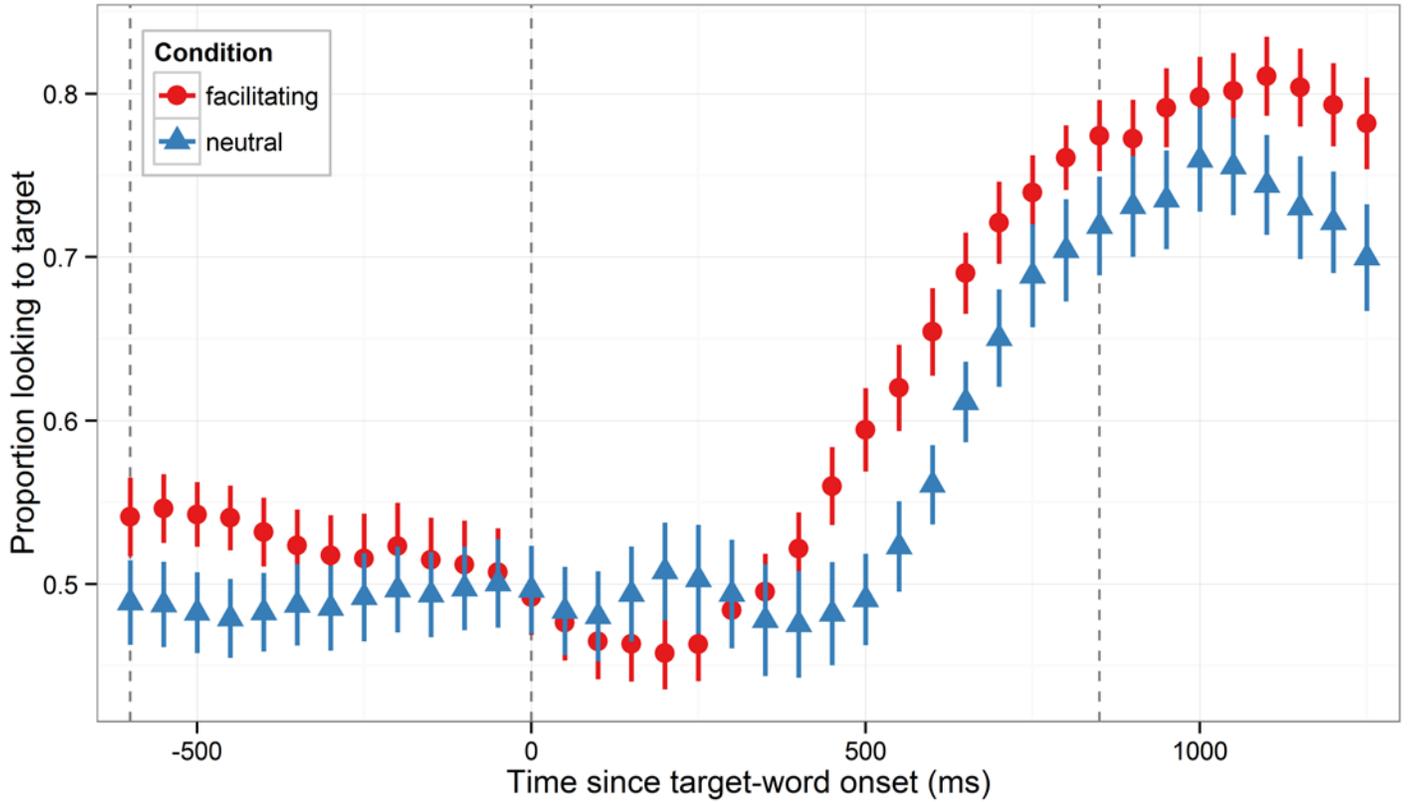


Figure 2. Proportion looking to target from onset of *the* to 1250 ms after target-word onset in the two conditions. Symbols and error bars represent observed means  $\pm$ SE. Dashed vertical lines mark onset of *the*, target-word onset, and target-word offset.

Model estimates are depicted in Figure 3, and complete model specifications are given in Appendix 1. The log-odds of looking to target in the neutral condition over the entire analysis window were estimated by the intercept term [ $\gamma_{00} = 0.438$ ; as a proportion: .608]. The linear, quadratic, and cubic orthogonal time terms were all significant, confirming a curvilinear, sigmoid-shape change in looks to target over time.

There was a significant increase in accuracy in the facilitating condition [ $\gamma_{01} = 0.21$ ;  $SE = 0.1$ ;  $t = 2.05$ ;  $p = .04$ ] such that the overall proportion of looking to target increased by .049. There was a significant effect of condition on the quadratic term [ $\gamma_{21} = -0.5$ ;  $SE = 0.18$ ;  $t = -2.81$ ;  $p = .005$ ]. These effects can be interpreted geometrically: The larger intercept increases the overall area under the curve, and the reduced quadratic effect decreases the bowing on the center of the curve, allowing the facilitating curve to obtain its positive slope earlier than the neutral curve. There was not a significant effect of condition on the linear term [ $\gamma_{11} = 0.54$ ;  $SE = 0.37$ ;  $t = 1.44$ ;  $p = .15$ ], indicating that the overall slopes of the growth curves did not differ significantly. These condition effects result in the two curves being roughly parallel at the center of the analysis window but with points phase-shifted by 100 ms.

Participant-level variables were tested by comparing nested models. There was no effect of vocabulary size on the intercept ( $\chi^2(1) = 2.9$ ,  $p = .091$ ), nor did vocabulary size interact with the condition effect ( $\chi^2(2) = 3.1$ ,  $p = .22$ ). There was also no effect of age on the intercept term ( $\chi^2(1) = 2.6$ ,  $p = .1$ ), nor did age interact with condition ( $\chi^2(2) = 2.7$ ,  $p = .26$ ). Model fit did not significantly improve when vocabulary size or age were allowed to interact with Time or Time-by-Condition parameters.

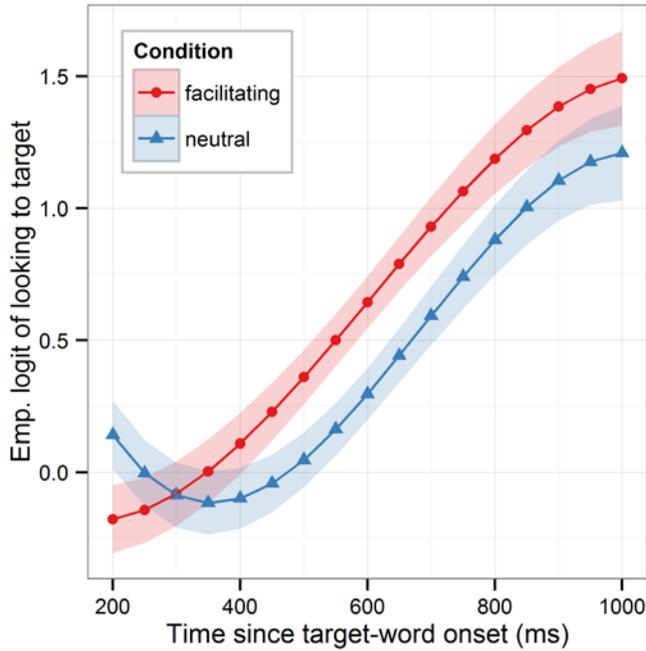


Figure 3. Growth curve estimates of looking probability during analysis window. Symbols and lines represent model estimates, and ribbon represents  $\pm$ SE. Empirical logit values on y-axis correspond to proportions of .5, .62, .73, .82. Note that the curves are essentially phase-shifted by 100 ms.

#### 4. Discussion

The present study provides the first evidence that toddlers take advantage of coarticulatory cues across word boundaries when recognizing familiar words. Participants on average looked to a named image approximately 100 ms earlier when the determiner *the* contained coarticulatory cues about the onset of the following noun. These results indicate that novice word-learners can take advantage of anticipatory coarticulatory information across word boundaries to support recognition of familiar words. Salverda, Kleinschmidt, and Tanenhaus

(2014) demonstrated a similar coarticulatory advantage with adult listeners. Our results show that listeners can take advantage of these cues in earliest stages of lexical development.

The results of this study have consequences for theories of lexical development and word recognition. In particular, these results do not support the hypothesis that early lexical representations are underspecified—that is, encoding just enough phonetic detail to differentiate a word from competing words. In this respect, our findings agree with numerous mispronunciation studies by demonstrating that novice-word learners have detailed phonetic information in their lexical representations.

This study does provide support for models of word recognition in which sub-phonemic features may provide useful information that can constrain inferences across word boundaries. In a lexical activation model (e.g. TRACE 1.0 as in Elman & McClelland, 1986), acoustic feature-detectors can capture coarticulatory information and trigger earlier activation of the subsequent onset consonant and the target word. Similarly, in a Bayesian framework, coarticulatory information increases the likelihood of a given sound and target word, permitting earlier inference about the identity of the word.

These findings also have methodological implications for word-recognition research. There is rich phonetic detail in the speech signal, and children can take advantage of this information. As a result, researchers cannot assume that lexical processing begins after target-word onset. If an experimental design explores how listeners process and integrate phonetic cues over time, investigators should analyze or control for coarticulatory dependencies over word boundaries.

In this study, we found that children 18–24 months of age could use anticipatory coarticulatory cues to facilitate word recognition. It will be of interest to determine if these

results generalize to even younger children or to children with language impairments, given that delayed word learning is a characteristic of virtually all language disorders. Because none of the developmental measures (age and vocabulary size) predicted sensitivity to coarticulation in this study, we might hypothesize that toddlers do not *develop* sensitivity to coarticulation or *gradually learn* to exploit these anticipatory cues. Put another way, knowledge about coarticulation and the contextual covariation of speech sounds may be an integral part of children's representations of words from the earliest stages of word learning.

## **Acknowledgements**

This research was supported by NIDCD grant R01 DC012513 to Susan Ellis Weismer, Jan Edwards, & Jenny R. Saffran, NICHD grant R37-HD037466 to Jenny R. Saffran, a grant from the James F. McDonnell Foundation to Jenny R. Saffran, NIDCD Grant R01-02932 to Jan Edwards, Mary E. Beckman, & Benjamin Munson, NICHD Grant 2-T32-HD049899 to Maryellen MacDonald, and NICHD Grant P30-HD03352 to the Waisman Center. We thank Eileen Haebig, Franzo Law II, Erin Long, Courtney Venker, and Matt Winn for help with many aspects of this study.

## References

- Barr, D. J. (2008). Analyzing “visual world” eyetracking data using multilevel logistic regression. *Journal of Memory and Language*, 59(4), 457–474. doi:[10.1016/j.jml.2007.09.002](https://doi.org/10.1016/j.jml.2007.09.002)
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). *lme4: linear mixed-effects models using Eigen and S4*. Retrieved from <http://CRAN.R-project.org/package=lme4>
- Charles-Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children’s lexicons. *Journal of Child Language*, 17(1), 205–215. doi:[10.1017/S0305000900013180](https://doi.org/10.1017/S0305000900013180)
- Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children’s receptive vocabularies. *Journal of Child Language*, 22(3), 727–735. doi:[10.1017/S0305000900010023](https://doi.org/10.1017/S0305000900010023)
- 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(5–6), 507–534. doi:[10.1080/01690960143000074](https://doi.org/10.1080/01690960143000074)
- Dollaghan, C. A. (1994). Children’s phonological neighbourhoods: half empty or half full? *Journal of Child Language*, 21(2), 257–271. doi:[10.1017/S0305000900009260](https://doi.org/10.1017/S0305000900009260)
- Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children’s production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47(2), 421–36. doi:[10.1044/1092-4388\(2004\)034](https://doi.org/10.1044/1092-4388(2004)034)
- Elman, J. L., & McClelland, J. L. (1986). Exploiting lawful variability in the speech wave. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability of speech processes* (pp. 360–385). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- 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: Brookes.
- 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(4), 1003–15. doi:[10.1111/1467-8624.00331](https://doi.org/10.1111/1467-8624.00331)
- Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: using eye movements to monitor spoken language comprehension by infants and young children. In I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), *Developmental psycholinguistics: on-line methods in children’s language processing* (pp. 97–135). Amsterdam: John Benjamins Publishing Company.
- Fisher, C., Church, B. A., & Chambers, K. E. (2004). Learning to identify spoken words. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 3–40). Cambridge, MA: MIT Press.

- Gow, D. W., Jr. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28(1), 163–179. doi:[10.1037/0096-1523.28.1.163](https://doi.org/10.1037/0096-1523.28.1.163)
- Gow, D. W., Jr., & McMurray, B. (2007). Word recognition and phonology: the case of English coronal place assimilation. *Papers in Laboratory Phonology*, 9, 173–200.
- Jusczyk, P. W. (1993). From general to language-specific capacities: the WRAPSA model of how speech perception develops. *Journal of Phonetics*, 21, 3–28.
- Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General*, 132(2), 202–227. doi:[10.1037/0096-3445.132.2.202](https://doi.org/10.1037/0096-3445.132.2.202)
- Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, 11(3), F9–16. doi:[10.1111/j.1467-7687.2008.00671.x](https://doi.org/10.1111/j.1467-7687.2008.00671.x)
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: a hierarchical framework. *Journal of Experimental Psychology: General*, 134(4), 477–500. doi:[10.1037/0096-3445.134.4.477](https://doi.org/10.1037/0096-3445.134.4.477)
- Mayor, J., & Plunkett, K. (2014). Infant word recognition: insights from TRACE simulations. *Journal of Memory and Language*, 71(1), 89–123. doi:[10.1016/j.jml.2013.09.009](https://doi.org/10.1016/j.jml.2013.09.009)
- 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(5), 1363–1389. doi:[10.1037/0096-1523.25.5.1363](https://doi.org/10.1037/0096-1523.25.5.1363)
- Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, 91(1), 3–19. doi:[10.1037//0022-0663.91.1.3](https://doi.org/10.1037//0022-0663.91.1.3)
- Mirman, D. (2014). *Growth curve analysis and visualization using R*. Boca Raton, FL: Chapman & Hall/CRC.
- Plunkett, K. (2006). Learning how to be flexible with words. In Y. Munakata & M. Johnson (Eds.), *Processes of change in brain and cognitive development: attention and performance* (Vol. XXI, pp. 233–248). Cambridge, MA: MIT Press.
- Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. *Journal of Memory and Language*, 71(1), 145–163. doi:[10.1016/j.jml.2013.11.002](https://doi.org/10.1016/j.jml.2013.11.002)
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76(2), 147–66. doi:[10.1016/S0010-0277\(00\)00081-0](https://doi.org/10.1016/S0010-0277(00)00081-0)

Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*(5), 480–484. doi:[10.1111/1467-9280.00485](https://doi.org/10.1111/1467-9280.00485)

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. *Cognition*, *71*(2), 73–108. doi:[10.1016/S0010-0277\(99\)00021-9](https://doi.org/10.1016/S0010-0277(99)00021-9)

Tobin, S., Cho, P. W., Jennet, P., & Magnuson, J. S. (2010). Effects of anticipatory coarticulation on lexical access. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd annual meeting of the cognitive science society* (pp. 2200–2205). Austin, TX: Cognitive Science Society.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–52. doi:[10.1177/0956797613488145](https://doi.org/10.1177/0956797613488145)

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*(2), 197–234. doi:[10.1080/15475441.2005.9684216](https://doi.org/10.1080/15475441.2005.9684216)

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: effects of age and vocabulary size. *Infancy*, *3*(1), 1–30. doi:[10.1207/15250000252828226](https://doi.org/10.1207/15250000252828226)

White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, *59*(1), 114–132. doi:[10.1016/j.jml.2008.03.001](https://doi.org/10.1016/j.jml.2008.03.001)

## Appendix: Model Summary

$Y_{ijk}$  estimates the log-odds of looking to target image at time-bin  $t$  for Child  $j$  in Condition  $k$ .

Linear, quadratic and cubic Time terms are orthogonal polynomials.

$$Y_{ijk} = \alpha_{0jk} + \beta_{1jk} * \text{Time}_t + \beta_{2jk} * \text{Time}_t^2 + \beta_{3jk} * \text{Time}_t^3 + R_{ijk}$$

$$\alpha_{0jk} = \gamma_{00} + \gamma_{01} * \text{Condition} + U_{0j} + W_{0jk} + \zeta_{0jk}$$

$$\beta_{njk} = \gamma_{n0} + \gamma_{n1} * \text{Condition} + U_{nj} + W_{njk} + \zeta_{njk}, \text{ for each parameter Time}^n$$

Fixed Effects	Estimate	SE	$t$	$p$
Intercept ( $\gamma_{00}$ )	0.438	0.092	4.785	< .001
Time ( $\gamma_{10}$ )	1.841	0.320	5.751	< .001
Time <sup>2</sup> ( $\gamma_{20}$ )	0.521	0.128	4.074	< .001
Time <sup>3</sup> ( $\gamma_{30}$ )	-0.433	0.108	-4.008	< .001
Facilitating Cond. ( $\gamma_{01}$ )	0.211	0.103	2.050	.04
Time $\times$ Facilitating Cond. ( $\gamma_{11}$ )	0.537	0.374	1.438	.15
Time <sup>2</sup> $\times$ Facilitating Cond. ( $\gamma_{21}$ )	-0.503	0.179	-2.809	.005
Time <sup>3</sup> $\times$ Facilitating Cond. ( $\gamma_{31}$ )	0.196	0.140	1.405	.16

(Using normal approximation for p values)

Random Effects		Variance	SD	Correlations			
Child	Intercept ( $U_{0j}$ )	0.089	0.299	1.00			
	Time ( $U_{1j}$ )	0.947	0.973	.69	1.00		
	Time <sup>2</sup> ( $U_{2j}$ )	0.011	0.104	.40	-.38	1.00	
	Time <sup>3</sup> ( $U_{3j}$ )	0.056	0.237	-.57	-.99	.52	1.00
Child $\times$ Condition	Intercept ( $W_{0jk}$ )	0.151	0.388	1.00			
	Time ( $W_{1jk}$ )	1.962	1.401	-.10	1.00		
	Time <sup>2</sup> ( $W_{2jk}$ )	0.410	0.640	-.08	.33	1.00	
	Time <sup>3</sup> ( $W_{3jk}$ )	0.233	0.483	.18	-.33	.09	1.00
Residual	$R_{ijk}$	0.332	0.577				