



Phonological representation in English-learning infants' word comprehension

Caroline Beech^{1,2}, Daniel Swingley²

¹Department of Psychology, Princeton University

²Department of Psychology, University of Pennsylvania

Abstract

Infants begin understanding words within their first year, but few studies have investigated how accurately infants this young represent words' phonological forms during comprehension. This study tested English-learning infants aged 6–8, 9–11, and 12–14 months in an internet-based language-guided-looking task probing recognition of correct pronunciations and initial-consonant mispronunciations of words ($n=97$; 48% female; 80% White, 10% Black, 5% multiracial, 4% Asian; $M_{\text{age}}=10.4$ months; tested 2021–2023). The two older groups recognized correct pronunciations. Mispronunciation reduced target fixation in 9- to 11-month-olds, an effect that was non-significant ($p=0.057$) including all items but significant over items with balanced salience ($OR=1.51$). 12- to 14-month-olds consistently fixated targets less when hearing mispronunciations ($OR=1.59$). These findings set a lower bound on how vague early word representations could be, and are consistent with phonologically accurate encoding of at least some words before the first birthday.

Keywords

phonological representations; word recognition; mispronunciation sensitivity

Speech is a gradient signal, but considering speech in terms of phonological descriptions, or strings of discrete phonemes, helps to explain many aspects of language behavior. For example, in clear speech, a difference in the phonological description of two word forms normally indicates a difference at the lexical level. In English, you could say “dog” with various prosodic contours or a surprisingly long vowel; but if you changed the phonological description by starting “dog” with a long voice onset time, i.e., if you said “tog,” it would provoke a new host of explanatory hypotheses in the listener—perhaps you meant a different word, or you made an error. Adult English speakers know this (e.g., Cole, 1973). But when do small phonological distinctions come to matter in this way among infants? On one view of language development, infants use phonologically specific lexical representations from the start, as soon as they are able to recognize words, and would know that a voicing change implicates a lexical difference. An alternative view holds that there is an initial period in

which infants' representations are vague and holistic before they develop more detailed, phonological representations.

The question of when children have access to accurate, adult-like phonological representations is important for many aspects of language learning, because phonology is implicated in so many basic linguistic processes. A language's phonology defines which word instances count as distinct minimal pairs and which count as the same; it sets limits on what could serve as a morphological marker; and it provides a point of contact between the representations needed for speech comprehension and those needed for speech production. When modeling the early development of these processes, a common starting assumption is that infants can group together instances of the same phonological string, and tell apart instances of different phonological strings. On this kind of account, even in their first months infants encode speech as a sequence of consonants and vowels—not precisely the set of segmental categories given by the native language, as these would not have been mastered yet, but nevertheless as a sequence, rather than a holistic and continuous acoustic object.

This view is often motivated by demonstrations of infants' precocious abilities to perceptually discriminate speech sounds. Infants are able to distinguish some categories (e.g., “ba” vs. “pa”) from birth (e.g., Eimas et al., 1971; Eimas & Miller, 1980), and within the first year, they become attuned to the particular contrasts of their language (e.g., Kuhl et al., 2006; Werker & Tees, 1984). That infants can learn and recognize language-specific phonetic categories might suggest that their first lexical representations would also be specified in terms of their language's phonological units. Listening preference experiments with word forms have found that by 11 months infants prefer (listen longer to) spoken lists of familiar word forms over rare or non-words (see Carbajal et al., 2021 for a meta-analysis), and that in many cases this preference dissipates if the familiar words are realized with phonologically deviant pronunciations (Hallé & de Boysson-Bardies, 1996; Swingley, 2005; Vihman et al., 2004), though there are exceptions, particularly when the deviation is placed into a less salient position compared to other, more prominent positions in the word (e.g., geminates, stressed onsets; e.g., Vihman & Majorano, 2017). Infants at this age also show a familiarity preference for correct pronunciations over mispronunciations when the two are directly pitted against each other (e.g., Ratnag et al., 2023). Although these preference experiments do not necessarily implicate discrete phonological representations, the results are consistent with a conventional vision of early representation in terms of segmental units (for discussion, Swingley, 2022; Vihman, 2022).

However, there are other reasons to think that in infancy, representations of words might not be composed of discrete phonological units with the correct specifications. Children might instead begin with holistic representations that are not sequences of categorical elements at all, but continuous phonetic chunks, or as Nittrouer (2021) describes them, “recurring patterns of broad spectral structure” (p. 510). Initially, these patterns would be grouped into recognizable (proto-lexical) types according to a similarity space based on acoustically salient features. Children's subsequent refinement of their lexical representations, and convergence toward the language's discrete phonological components, could then be a quite protracted process, driven by the linguistic obligations to differentiate minimal pairs (e.g., Charles-Luce & Luce, 1995; Cui, 2020; Metsala & Walley, 1998; Storkel, 2002), to

articulate clearly (e.g., Vihman & Croft, 2007), and to correctly apportion phonetic variation among affective, pragmatic, syntactic, and lexical sources (e.g., Quam & Swingley, 2010; Werker & Curtin, 2005). On this kind of theory, there is an early developmental period in which words are not represented as sequences of phonological categories, because such representations are not yet necessary.

Computational modeling research has also contributed to this debate by suggesting that identification of the language's speech sound categories may not even be possible before lexical development begins, contrary to the traditional hypothesis that word learning follows phonetic attunement. Phonetic studies of vowel categories in infant-directed speech seem to indicate that language-given phonological categories are not recoverable from analysis of the distribution of vowel tokens in the signal; each vowel type does not present itself as a singular peak of concentrated instances separated by unoccupied low-probability regions. Rather, vowel instances seem to be spread rather thinly over phonetic space, not quite uniformly, but also not in a way that makes categories clear. Computational attempts to mimic infant learning by finding statistical "modes" corresponding to vowel category centers have failed, whereas models that grant the infant some lexical knowledge perform much better (e.g., Feldman et al., 2013; Swingley, 2009a; see also Schatz et al., 2021). These models suggest a "poverty of the stimulus" problem at least partially solved by having word learning and phonological learning go hand in hand developmentally. If this is true, then the very earliest words could not be represented using well-specified phonological categories.

Considering the nature of infant-directed speech and the ecology of real-world language interactions, it is clear that what infants are capable of discriminating perceptually is not necessarily equivalent to what infants represent in memory when they acquire words as lexical items with associated meanings. For instance, while infants can certainly hear the difference between clearly articulated instances of [b] and [p], that does not imply that they represent the word "banana" as a sequence of six phonemes beginning with /b/. They might instead think of "banana" in a very different way, e.g., as the trisyllabic word with a long vowel in the middle. In this case, a one-phoneme substitution, which would signal a lexical distinction to adults, would not be expected to impact infants' word recognition. This could come about because infants do not have true phonological representations for *any* words. Or, infants might represent their most familiar, "best" words (e.g., "mommy") as sequences of canonical consonants and vowels, but not so for slightly less familiar words, simply because what they retain in memory about these words is fuzzy.

While it is quite difficult to conclusively determine whether word representations are holistic or phonologically specified, it is possible to measure whether word forms in memory are more readily evoked using canonical pronunciations than using deviant pronunciations. As noted above, at 11 months, listening preference experiments with lists of unfamiliar word forms vs. correctly pronounced or mispronounced familiar word forms (e.g., "budget" vs. "bubbles" or "mubbles") have been effective for probing infants' representations of words. However, this procedure does not seem to work well before 11 months, with younger infants failing to show a familiarity preference even for correctly pronounced words (e.g., Vihman et al., 2004). In addition, because listening time reflects an infant's preference for the word

list overall, this procedure does not allow the researcher to know which particular words elicit the effect or investigate differences between words.

The present study used a test of infants' mispronunciation sensitivity in word recognition developed by Swingley and Aslin (2000, 2002). This extension of the "looking-while-listening" or "language-guided-looking" procedure (Fernald et al., 1998; Swingley, 2012) tests infants' comprehension of, rather than familiarity with, correct pronunciations versus deliberate mispronunciations. On each trial, infants see a pair of images (e.g., a baby and a dog) and, after a brief pause, hear a sentence naming one of the images using either a correct pronunciation or deliberate mispronunciation (e.g., "Where's the dog?" or "Where's the mog?"). Greater target looking during the correct pronunciation trials compared to the mispronunciation trials is taken as a measure of mispronunciation sensitivity, and thus the phonological specificity of infants' lexical representations. One way of thinking about children's performance in this task is that by default, children will look back and forth between the two images, but if they identify a match between what they have heard and what they are looking at, then they should pause longer on that image. If infants' early lexical representations were phonologically specific, then they should perform better with "dog" than with a one-phoneme mispronunciation, as adults do; but if infants instead have vague representations of how words sound, then "dog" and "mog" could be equally good matches.

Many studies have used this procedure to assess the phonological specificity of children's lexical representations in the second year of life and onward. Studies have tested mispronunciation sensitivity in a variety of language populations (e.g., Höhle et al., 2006; Højen et al., 2023; Swingley, 2003; Tao & Qinmei, 2013; Zesiger et al., 2012), in children exposed to bilingual (e.g., Ramon-Casas et al., 2009; Wewalaarachchi et al., 2017) or bidialectal (e.g., Durrant et al., 2015; van der Feest & Johnson, 2016) language input, in children with different vocabulary sizes (e.g., Law & Edwards, 2015), and in different clinical populations (e.g., Curtis et al., 2023; Cychosz et al., 2023). Studies have also explored a number of variables on the stimulus side, including lexical class (e.g., Moore & Bergelson, 2022), age of the talker (e.g., Bernier & White, 2019; Bulgarelli, 2024), licensed (phonetic assimilation or accent-like) versus unlicensed mispronunciations (e.g., Cooper et al., 2023; Singh & Cheng, 2023; Skoruppa et al., 2013; White & Aslin, 2011), and different kinds or sizes of mispronunciation (e.g., Altvater-Mackensen et al., 2014; Mani & Plunkett, 2007, 2010; Quam & Swingley, 2014; Ren et al., 2019; White & Morgan, 2008). Related work has also tested how mispronunciations affect infants' recognition of newly learned words (e.g., Quam & Swingley, 2023; Tsuji et al., 2016) as well as how other kinds of acoustic modifications (e.g., vowel expansion) affect recognition (e.g., Lovcevic et al., 2024).

Swingley and Aslin (2002) used this procedure to test 14- to 15-month-olds, but more recent studies suggest that much younger infants already comprehend some words. For verbs and other more abstract words (e.g., "all gone", "night night"), infants demonstrate above-chance target looking by 9 or 10 or months (Bergelson & Swingley, 2013; Nomikou et al., 2019; Syrnysk & Meints, 2017), and infants as young as 6 to 7 months have succeeded at recognizing concrete nouns (Bergelson & Swingley, 2012, 2015; Tincoff & Jusczyk, 2012). In the present study, we wanted to test infants' mispronunciation sensitivity in these

earliest stages of word recognition. Because infants between 6 and 10 months are still learning the phonetic categories of their particular language (e.g., Werker & Tees, 1984), investigating the nature of infants' lexical representations at these early ages could prove especially important theoretically.

A recent meta-analysis (Von Holzen & Bergmann, 2021), combining data from the previous twenty years of mispronunciation sensitivity experiments, concluded there was robust evidence that infants are sensitive to mispronunciations in word recognition, and that across studies, this ability does not seem to be modulated by age. This result might suggest that infants have mature, well-specified phonological representations of familiar words even in the earliest stages of word recognition. However, only 1 out of the 32 papers included in the meta-analysis (Bergelson & Swingley, 2018) tested infants under 12 months, leaving open the possibility that infants have vaguer, more inclusive representations of words' forms before this point. Bergelson and Swingley (2018) tested the effect of vowel mispronunciations (e.g., "banoona" vs. "banana") in order to compare effects of phonological vowel changes against effects of hearing an unfamiliar talker, for infants from 6 to 14 months of age. Their design was between-participants (i.e., one group of infants heard all correct pronunciations and a separate group of infants heard all mispronunciations), which for present purposes is not ideal because individual performance is variable in this age range, but overall, their results suggested that very young infants may not precisely represent the vowels in early-learned words. The youngest group of infants (6- to 7-month-olds) looked to the target just as much when hearing a mispronunciation ("banoona") as a previously tested group (Bergelson & Swingley, 2012) had when hearing a correct pronunciation ("banana"). Here we build on this work, investigating mispronunciation sensitivity in very young infants for a second time, but using a within-subjects design and instead testing consonants, which have been proposed to anchor lexical processing (e.g., Nazzi & Cutler, 2019).

Present Work

In this work, we sought to investigate the age at which infants come to accurately represent the consonants within familiar words. Following substantial previous work with older infants and children, we presented infants with two images at a time and compared their target looking in correct pronunciation trials (e.g., "Look at the dog!") and mispronunciation trials ("Look at the mog!"). In addition, to assess successful recognition of the correct pronunciations, we tested whether infants in each age group looked more at a given image when it was named with a correct pronunciation than when the other image on the screen was named. Per our pre-registration (available on the Open Science Framework, <https://osf.io/xgcqv/>), we tested infants from three different age groups: 6 to 8 months, 9 to 11 months, and 12 to 14 months. Based on several previous studies, we predicted that the older infants (12 to 14 months) would successfully recognize correct pronunciations, and would exhibit worse performance in response to mispronunciations. For the younger infants (6 to 8, and 9 to 11 months), we also expected successful recognition of correct pronunciations (e.g., Bergelson & Swingley, 2012), and aimed to determine whether these two younger age groups would demonstrate a sensitivity to mispronunciations. While several studies suggest a sensitivity to mispronunciations among infants in the second year of life, the theoretical

and empirical case for phonological representations of words in younger infants is much less clear-cut.

While not the primary focus of this study, this study also represented a test of the feasibility of conducting these kinds of experiments over the internet. Web-based infant eye-tracking experiments have been gaining traction in recent years (e.g., Chuey et al., 2024) particularly in response to the COVID-19 pandemic during which the data for this study were collected. Prior studies have adapted several procedures for the web, including violation of expectation (Scott et al., 2017), syntactic bootstrapping using preferential looking (Scott et al., 2017), headturn preference or central fixation (Li et al., 2023), anticipatory looking (Steffan et al., 2024), and the novel word learning Switch task (Bulgarelli & Bergelson, 2022; Nguyen et al., 2024). This study joins a handful of other studies assessing infants' comprehension of familiar words over the internet (see e.g., Bacon et al., 2021; Nguyen et al., 2024; Weaver et al., 2024 for internet-based looking-while-listening with 24-month-olds and 14- to 18-month-olds), and, to our knowledge, was the first to test infants' comprehension of mispronunciations online.

Method

Participants

Participants were 97 children: 32 6- to 8-month-olds (mean age = 6;23, 53% female), 32 9- to 11-month-olds (mean age = 10;11, 47% female) and 33 12- to 14-month-olds (mean age = 14;03, 45% female), tested from April 2021 through June 2023. This sample size ($n=32$ per group) was selected so as to yield approximately 80% power (two-tailed t -test) based on the effect size observed in Swingley and Aslin (2002). Because data collection occurred online at a time convenient for the parents, data collection closed after one supernumerary 14-month-old joined the sample, resulting in 33 participants in the oldest age group. An additional 78 infants were tested but excluded from the final sample because of equipment failures (e.g., computer failed to record sound well enough to identify stimulus timing information, $n=22$), parental interference with the task ($n=2$), or because infants failed to contribute at least 8 trials to each condition ($n=54$). 32 experimental trials (16 per condition) were presented to each infant. Following our pre-registration, trials were excluded when infants failed to fixate the images for a minimum of 2/3 of a second in the analysis window ($n=371/3104$ total trials among retained infants); when trials belonged to a sequence of two or more consecutive trials where the infant only ever fixated one side of the screen ($n=63/3104$ total trials among retained infants); or when technological problems or background noise made it impossible to confirm the timing of stimulus presentation ($n=12/3104$ total trials among retained infants). Among the 54 participants excluded for insufficient trials, 50 were due to insufficient fixation duration, 3 due to a combination of insufficient fixation duration and prolonged fixation to only one side of the screen, and 1 due to trials with uncertain stimulus presentation timing.

Participants were recruited from the Philadelphia area primarily through study advertisements emailed to their parents on our behalf by the Children's Hospital of Philadelphia. Once their parents expressed interest in the study, infants were eligible to participate if they were typically developing, with no reported ear or hearing problems, had

been born full-term, and were being raised in a monolingual English environment (at least 75% English exposure, based on parental estimate). In the final sample, participants were predominantly mid to high SES (84% of parents had a college degree or higher). 10% of participants' parents identified their child as Black or African American, 80% White, 5% more than one race, and 4% Asian. These demographic characteristics are very similar to the 2024 Census-reported racial demographics for Pennsylvania, where the lab is based, and approximately equal numbers of male and female children were included in the sample, making it representative in terms of gender. The Institutional Review Board approved the study, and parents provided informed written consent prior to testing.

Materials

The 16 target words we tested are displayed in Table 1. These words were selected because they are easily depicted using still images; are frequent in speech to infants, as assessed using both the Brent and Siskind (2001) corpus and a subset of the SEEDLingS corpus (Bergelson, 2017) (e.g., median frequency of at least 5 across different dyads); and because they were commonly reported to be understood by older infants, according to WordBank's (Frank et al., 2017) data for American English-learning children (e.g., at least 30% of 12- to 14-month-olds across different demographic backgrounds reported to understand).

For each correct pronunciation, we created a corresponding mispronunciation by changing the word's initial consonant (or consonant cluster, in the case of "spoon") by at least two features (e.g., place and manner of articulation). Items were presented in yoked pairs, such that e.g., shoe was always the distracter when ball was the target and vice versa.

Visual stimuli. The visual stimuli were photographs of objects against a solid gray background, presented side by side on participating families' computers. Each image occupied most of a square bounding box 1/4 the width of the screen (e.g., 320 by 320 pixels on a 1280-by-800-pixel laptop screen). Infants were seated at eye level with the pictures, on their parent's lap or in a high-chair or other supported position. Stimulus presentation across trials was pseudo-randomized such that the target pictures appeared on the left and the right equally often, but never more than twice in a row on the same side, and not covarying with condition (correct pronunciation vs. mispronunciation) or carrier phrase ("Look at the" vs. "Find the").

Auditory stimuli. To create the auditory stimuli, we recorded a female native speaker of American English who produced each sentence in clear, child-directed speech, using the same prosody for the correct pronunciations and the mispronunciations. We recorded multiple instances of each sentence and selected the stimuli so that correct pronunciations and their corresponding mispronunciations were as closely matched as possible, qualitatively and in duration ($M = 0.70$ s, $SD = 0.10$ vs. $M = 0.69$ s, $SD = 0.09$). Audio editing was used to remove pops or clicks from the recording, adjust the amplitude of waveform, and splice together different recording takes if one provided a better matched realization of the target word while another provided a better matched realization of the carrier phrase ("Look at the" or "Find the"). During the experiment, each sentence was preceded by a bell sound signaling the start of the trial followed by 2 seconds of no speech,

with two metronome clicks embedded within this period (e.g., “*Ding! Click click...* Look at the ball!”). These punctate sounds allowed us to confirm the timing of stimulus presentation when reviewing participants’ recordings.

Trial timing. On each trial, the audio file started playing at the same time the images appeared on the screen, giving participants just over 2000 ms of image preview time (2 seconds plus the length of the bell sound) before the start of speech. Images remained on the screen for 4000 ms after target-word offset.

Procedure

All participants were tested at home over the internet using the PCIBex online experiment platform which presents stimuli and records video and audio using the computer’s webcam and microphone (Zehr & Schwarz, 2018). Parents first video-called with the lab manager to learn about the study, provide written consent, and answer demographic questions about their child. Then, parents were given a link to complete the experiment at their convenience. Thus, parents’ personal computers were used for experimental presentation and participant recording to the web. Parents were told to seat their child in front of the computer screen so that the child’s face would be clearly visible and centered in the webcam recording, and asked to refrain from speaking, pointing, or otherwise directing their child during the experiment. After the experiment, parents were provided monetary compensation for their time and asked to complete the MacArthur-Bates CDI Words and Gestures (Fenson et al., 1994).

The experiment consisted of five calibration trials (in which a flower-shaped color wheel appeared on the left or right side of the screen), four warm-up trials featuring words and images not used in the rest of the experiment (e.g., a fish image and a sock image accompanied by the sentence “Where’s the sock?”), and 32 experimental trials (see Figure 1 for an example). Over these 32 trials, each of the 16 target words (Table 1) was presented once as a correct pronunciation (e.g., “ball”) and once as a mispronunciation (e.g., “chall”). Half the words appeared first in their correctly pronounced form and half appeared first in their mispronounced form, with trials featuring the same pair of images spread out throughout the experiment. Participants were randomly assigned to one of four presentation orders so that which trials were correct pronunciation trials versus mispronunciation trials and the overall sequence of trials was counterbalanced across participants. These presentation orders and all stimuli are available online on the Open Science Framework.

Data Preparation

Participant recordings were manually coded offline by the first author and trained research assistants. We examined the recordings frame by frame (1 frame = 33 ms) using ELAN (2022) and coded, for each frame, whether the infant’s eye gaze was toward the left side of screen, the right side of screen, away from the screen, or in transition. Left and Right codes were then replaced with Target and Distracter codes based on target image side for that trial, in order to compute infants’ proportion looking to the target image.

Identifying the timing of stimulus presentation required unique considerations compared to in-lab experiments. Because the PCIBex platform produced a video recording for each trial and the onset of webcam recording relative to actual stimulus presentation onset varied slightly by trial, we coded each trial individually and reassigned video frames to times in pre-processing based on when they occurred in time relative to two metronome clicks in the audio stimuli. In each of the audio stimuli files (which played at the start of each trial and coincided with the appearance of the visual stimuli), the second metronome click occurred exactly 1.5 seconds in, so the video frame in the participant recording 1.5 seconds before this click was considered the start of the trial, and any video frames recorded before the stimuli began were excluded. Trials where the metronome clicks were not identifiable in the recording were excluded from the sample. The webcam-recorded videos also had slightly variable frame rates (across and within videos) but were re-encoded as a constant 30 frames per second for coding in ELAN.

Following previous work (e.g., Bergelson & Swingley, 2012) and our pre-registration, we analyzed target and distracter looking from 367 ms to 3500 ms after target-word onset as the primary analysis window. Target-word onset was computed arithmetically based on the onset of stimuli presentation or start of the trial (identified in the participant recordings as described above) and the known timing of the start of the target word in each audio stimuli file.

Analysis

As the dependent measure in our analyses, we computed the proportion looking to each image (time spent looking to that image, divided by total time spent looking to either of the two images on the screen). We focus on the trial-level data here, but more traditional t-test analyses over subject means, which were also included in our pre-registration, were largely consistent with these results and are reported in the Supplementary Material. Our analyses investigate two main questions: 1) how well did infants recognize words when they were pronounced correctly, and 2) how was infants' target looking affected by condition (correct pronunciations vs. mispronunciations)? If even very young infants have precise phonological representations of words, we would expect to see successful recognition of correct pronunciations, and decreased target looking in the mispronunciation trials. If infants have less precise, more holistic representations of words, however, we would expect them to accept loose matches (e.g., "nottle" for "bottle") with no significant decrement in target looking. These primary analyses are confirmatory rather than exploratory, testing in each of the three pre-registered age groups the hypothesis that infants would show increased looking to named targets, compared to looking to the same items when they appear as unnamed distracter images (i.e., that naming would have a significant positive effect on looking), and the hypothesis that target looking would be greater for correct pronunciations compared to mispronunciations (i.e., that condition, or correct pronunciation vs. mispronunciation, would have a significant positive effect on target looking).

Transparency and Openness

The pre-registration for this study, anonymized data, analytic code, and materials will all be publicly available on the Open Science Framework, <https://osf.io/ah2nv/>. We report above

how we determined our sample size, as well as all data exclusions, manipulations, and measures in the study. Data were analyzed using R, version 4.4.2 (R Core Team, 2022), and the tidyverse (Wickham et al., 2019), lme4 (Bates et al., 2015), and emmeans (Lenth, 2023) packages.

Results

Recognition

To assess how well infants recognized words when they were pronounced correctly, we tested how much infants looked to an image (e.g., the ball) when it was named (“Look at the ball!”), versus when the other image in the pair was named (“Look at the shoe!”). Specifically, to account for repeated measurements over subjects and items, we used a mixed effects logistic regression model over the trial-level data, with *proportion_x_looking* as the outcome and *x_named* (a binary variable indicating whether the image was the target or the distracter on the trial) as a fixed effect. This approach operationalizes successful recognition as significantly higher *proportion_x_looking* when *x_named* is true, as in previous work using t-tests over subjects means (e.g., Bergelson & Swingley, 2012), and allows for the inclusion of additional trial- and item-level variables such as *location_at_onset* (did the infant happen to be looking at the designated image at the moment the spoken target word began), *side* (was the image on the left or the right), and *salience* (how much did the infant prefer the image in the period before target-word onset across all trials using that image pair).

Using a mixed effects logistic regression model with these covariates (model formula: $\text{proportion_x_looking} \sim \text{x_named} * \text{age_group} + \text{location_at_onset} + \text{salience} + \text{side} + (\text{x_named} + 1 \mid \text{subject}) + (\text{x_named} + 1 \mid \text{item})$) revealed a significant effect of *x_named* on average across the sample ($\beta = 0.25$, $\text{SE} = 0.05$, $p < 0.001$, odds ratio = 1.64; Figure 2; Table 2). However, considering the estimated (simple) effect in each age group, only the 12- to 14-month-olds ($\beta = 0.91$, $\text{SE} = 0.16$, $p < 0.001$, odds ratio = 2.48) and the 9- to 11-month-olds ($\beta = 0.37$, $\text{SE} = 0.17$, $p = 0.028$, odds ratio = 1.45) looked significantly longer at the images when they were named, with the 6- to 8-month-olds showing no significant difference ($\beta = 0.20$, $\text{SE} = 0.17$, $p = 0.254$, odds ratio = 1.22). A pre-registered secondary analysis using the same model formula but considering looking over a shorter 367-to-2000-ms time window, which is commonly used for older toddlers (e.g., Swingley & Aslin, 2002), was consistent with all of these results (Table S1).

Mispronunciation Sensitivity

To measure infants’ sensitivity to mispronunciations and thus the phonological specificity of their lexical representations, we compared infants’ proportion target looking in the correct pronunciation and mispronunciation trials using another mixed effects logistic regression model, with a fixed effect for condition (correct pronunciation vs. mispronunciation) and the same covariates plus a *location_at_onset* interaction term, which was not significant in the recognition analysis reported above. Thus, the full model formula was $\text{proportion_target_looking} \sim \text{condition} * \text{age_group} + \text{location_at_onset} + \text{salience} + \text{side} + \text{condition} : \text{location_at_onset} + (\text{condition} + 1 \mid \text{subject}) + (\text{condition} + 1 \mid \text{item})$, with

age_group forward difference coded and location_at_onset deviation coded such that the condition estimate reflected the main effect of condition on average across levels of those variables.

The model selection process was the same for recognition (previous section) and mispronunciation sensitivity, and was guided foremost by subject matter considerations (Sonderegger, 2023). Per our research questions and pre-registration, x_named and condition were considered the ‘critical’ term in the recognition analysis and mispronunciation sensitivity analysis respectively while location_at_onset, salience and side were considered ‘control’ terms. Assessing the effect of the critical term in each age group required including a critical term * age_group interaction (and a possible main effect for age_group). Based on the goals of the research, the only other interactions considered for inclusion were the two-way interactions between the critical term and the control terms. Each of these was only included in the final model if it significantly improved model fit (only true of the location_at_onset interaction, and only in the mispronunciation model). Similarly, in selecting the random effects structure, we included random intercepts for subject and item to reflect the study design (repeated measurements over both subjects and items), and by-subject and by-item random slopes for only the critical term (x_named or condition). These slopes were retained in the models because there is good theoretical reason to believe that these effects could be a stronger or weaker for certain children or certain items, and we wanted to prevent overconfident coefficient estimates for the fixed effect of the critical term, though in fact the results do not change if the random slopes are dropped from the model.

The mixed effects model investigating mispronunciation sensitivity revealed a significant effect of condition (correct pronunciation vs. mispronunciation) on proportion target looking on average across the sample ($\beta = 0.15$, $SE = 0.04$, $p < 0.001$, odds ratio = 1.34; Figure 3; Table 3). As before, we assessed the effect in each age group by pairwise comparison of the estimated marginal means. In the 6- to 8-month-old group, there was no significant mispronunciation effect ($\beta = 0.14$, $SE = 0.16$, $p = 0.385$, odds ratio = 1.14). However, this result reveals little about these infants’ word form representations because although the pattern of the means was in the expected direction, the proportion of target looking was not significantly above 50% in either condition (correct pronunciations: $\hat{P} = 0.55$, 95% CI = [0.49, 0.60]; mispronunciations: $\hat{P} = 0.51$, 95% CI = [0.46, 0.57]), and we found no significant evidence of recognition in this age group (see Recognition). In the oldest age group, the 12- to 14-month-olds, on the other hand, we observed the expected result: target looking in response to correct pronunciations was significantly above 50% ($\hat{P} = 0.63$, 95% CI = [0.58, 0.67]), and significantly greater ($\beta = 0.47$, $SE = 0.14$, $p = 0.001$, odds ratio = 1.59) than the target looking elicited by mispronunciations ($\hat{P} = 0.51$, 95% CI = [0.47, 0.56]).

In the middle, 9- to 11-month-old group, correct-pronunciation target looking was significantly above 50% ($\hat{P} = 0.57$, 95% CI = [0.51, 0.62]) and mispronunciation target looking was not ($\hat{P} = 0.50$, 95% CI = [0.44, 0.55]), but the difference between conditions was non-significant ($\beta = 0.29$, $SE = 0.15$, $p = 0.057$, odds ratio = 1.33). Thus, this sample does not provide clear evidence of mispronunciation sensitivity among 9- to 11-month-olds (at least when p -values are calculated using lme4’s glmer() function and the Wald

z method¹). However, considering looking behavior by item, it seems possible that the weakness of this effect might be the result of certain salience-mismatched item pairs. For a few item-pairs (bottle-kitty, foot-window), instead of the average salience being close to equal, children strongly preferred to look at one image (e.g., kitty) over the other (e.g., bottle) at the beginning of the trial. Across the 9- to 14-month-olds (Figure 4), these salience mismatches led to a less consistent mispronunciation effect, and considering 9- to 11-month-olds individually, a similar picture emerges (Figure S1). Essentially, for these two salience-mismatched pairs, infants continued to look at the preferred picture for most of the analysis window regardless of what the target word was. For the words in the other six pairs, recognition was usually stronger for the correct pronunciations than the mispronunciations. In an exploratory analysis, using the same regression model but excluding data from bottle-kitty or foot-window trials, the condition difference at 9 to 11 months was significant ($\beta = 0.41$, $SE = 0.17$, $p = 0.015$, odds ratio = 1.51).

Figure 5 shows the time course of recognition for the 9- to 14-month-olds. Trials were divided into those on which children were already fixating the target at word onset (solid lines in the figure) and those on which children were not (dashed lines). The y-axis reveals, for each moment in time, the proportion of trials on which infants were fixating the other picture (the distracter if they started on target, or the target if they started off-target). In line with the statistical analyses reported above, infants' looking behavior depended on the speech they heard, with mispronunciations eliciting more switching to the incorrect, distracter image if the infant happened to start on the target image, and (eventually) less switching to the correct, target image if the infant happened to start off target, compared to correct pronunciations. Perfect performance would be shown by the target-onset lines remaining low (no defection from the target) and the distracter-onset lines rising to one (always rejecting the distracter). The figure illustrates the interaction between condition and location_at_onset: The condition effect (difference between correct pronunciations and mispronunciations) was significantly stronger in target-initial trials ($\beta = 0.09$, $SE = 0.04$, $p = 0.043$; Table 3). Indeed, in a secondary analysis considering looking over the shorter 367-to-2000-ms window, the condition effect was only significant on target-initial trials (Table S2).

Changes with Age and Vocabulary

In addition to our analyses of categorical age groups (6–8 months, 9–11 months, 12–14 months), we also explored how the recognition effect and mispronunciation effect changed with age considered continuously, and with receptive vocabulary size, for infants whose parents returned the CDI vocabulary measure. Reported receptive vocabulary ranged from 0 to 290 words, with larger scores for older infants (6–8 months: mean=21, $SD=27.1$; 9–11 months: mean=55, $SD=71.4$; 12–14 months: mean=134, $SD=72.5$).

¹We specified using a mixed effects model for proportion data in our pre-registration, but the best implementation of this analysis is debated. To be conservative in our estimates, we focus on results from the lme4 package (the currently standard R package for fitting mixed effects models in psychology) and the default p -values it provides for logistic models using the Wald z approximation, but it is worth noting that some related approaches (e.g., using bootstrap resampling (Humphrey, 2022) to calculate p -values from a lme4 glmer() model, or considering the 95% credible interval from a Bayesian mixed effects zero-one-inflated beta regression (Bürkner, 2017)) return a significant rather than marginally significant effect of correct vs. mispronunciation among the 9- to 11-month-olds.

Considering the recognition effect, using the same model as before but with age as a continuous rather than categorical variable, we found that the effect of x_{named} on $\text{proportion}_{\text{x_looking}}$ significantly interacted with age ($\beta = 0.05$, $SE = 0.02$, $p < 0.001$), with older children showing a bigger difference between named targets versus unnamed distracters (i.e., better recognition). Substituting vocabulary size for age in the model, we also found a significant interaction between x_{named} and vocabulary ($\beta = 0.003$, $SE = 0.001$, $p = 0.001$), indicating a larger effect of x_{named} for infants with larger vocabularies.

Considering the mispronunciation effect, again using the same model as before but with continuous rather than categorical age, there was a significant main effect of continuous age on proportion target looking ($\beta = 0.03$, $SE = 0.01$, $p = 0.021$), with older children demonstrating more target looking. There was not a significant interaction ($\beta = 0.02$, $SE = 0.01$, $p = 0.151$) between continuous age and condition (correct pronunciation vs. mispronunciation), although this could be because the relationship between age and proportion target looking may itself change with age (see Supplementary Material, Figure S2). Substituting vocabulary size for age in the model, we likewise found a significant main effect of vocabulary size on target looking ($\beta = 0.002$, $SE = 0.001$, $p = 0.002$) but no significant relationship between vocabulary size and the size of the condition effect ($\beta = 0.0003$, $SE = 0.0007$, $p = 0.699$), in line with some previous studies of older infants (e.g., Swingley & Aslin, 2002).

Discussion

The present work investigated 6- to 14-month-old infants' word recognition and sensitivity to deliberate mispronunciations in an online language-guided-looking task. On average across the sample, infants recognized correctly pronounced words and demonstrated a significant mispronunciation effect, looking less to the target image when the initial consonant in the target word was mispronounced. Considering the age groups individually, while the 6- to 8-month-olds in the sample demonstrated at-chance performance in both conditions, the mispronunciation effect was marginally significant among the 9- to 11-month-olds, and significant among the 12- to 14-month-olds. Thus, the results do not constrain theorizing about the youngest children (6–8 months), because they did not perform well in this implementation of the procedure. Although we also cannot make firm claims about the 9- to 11-month-olds, our interpretation is that 9- to 11-month-olds do represent the specific onset consonants of at least some object names, given that for most of the items not showing floor or ceiling effects, children performed significantly better when words were pronounced canonically.

Overall, this study provides new data about young infants' sensitivity to consonant mispronunciations during word comprehension. Using the standard test of infant word comprehension, which has been widely used to investigate mispronunciation sensitivity in infants over 12 months (Von Holzen & Bergmann, 2021), we took a standardized cross-sectional developmental approach and tested children at different points of development using exactly the same procedure and items. In line with Von Holzen and Bergmann's (2021) overall meta-analytic findings of a consistent sensitivity to consonant mispronunciations for

older infants, our results are suggestive of phonologically specific representations in word comprehension for infants as young as 9 months.

Because this study only investigated English-learning infants in a United States cultural context, the generalizability of these findings to other populations is not known. However, our results align with related research conducted with other language populations using listening preference experiments to investigate infants' sensitivity to mispronunciations in word form recognition. Results from this procedure have found that, for example, French-, British English-, and Dutch-learning 11-month-olds prefer listening to lists of correct pronunciations or other kinds of deviations over initial-consonant mispronunciations (Poltock & Nazzi, 2015; Ratnag et al., 2023; Swingley, 2005), offering further support for the idea that infants represent familiar words' onset consonants early in development (though see Bouchon et al., 2015; Delle Luche et al., 2017 for contrasting results in 5-month-olds' own-name recognition).

If infants' representations of the consonants in words are well specified from as soon as they can successfully recognize words, this would have important implications for other aspects of language acquisition. For example, in addition to learning to recognize clear, canonical pronunciations of words, children also have to learn to recognize more ambiguous pronunciations that arise from accent and dialect variation (e.g., Best et al., 2009; Johnson & White, 2020; Schmale et al., 2010; Wagner et al., 2014) or everyday intraspeaker co-articulation and reduction processes (Beech & Swingley, 2023; Lahey & Ernestus, 2014; Warner, 2019). In principle, having broader, more inclusive representations of speech sounds early on might help the child collapse across some of this variability (e.g., Creel, 2022). Yet, the results of the present study do not support the idea that infants have only vague representations of the sounds in familiar words, certainly not after 12 months. This suggests that infants rely on an alternative strategy to overcome phonological variability in speech.

The mispronunciations in the present study were fairly large: They all involved multiple-feature changes to the initial consonant, rather than smaller changes to a single feature (voicing, manner of articulation, or place of articulation) only. Work with older children (e.g., Altvater-Mackensen et al., 2014; White & Morgan, 2008) has explored how changing single features of consonants affects word recognition, but for this first investigation of children under 12 months, we chose to test the kind of consonant mispronunciations we thought would have the greatest effect on recognition, if these infants do represent words' phonological forms with any precision. Future studies with infants of this age could examine whether more fine-grained mispronunciations still elicit a mispronunciation effect.

However, our results regarding these mispronunciations already demonstrate that infants at this age use more phonological detail than required by their lexicons. First, the words we tested have few phonological neighbors (words whose canonical pronunciations differ by a single phoneme) in infant-directed speech. Among a ten-thousand-word subset of the Brent and Siskind (2001) corpus, our stimulus words had only two neighbors of any kind (e.g., vowel substitution, consonant deletion) on average (median count). Moreover, considering each child's reported receptive vocabulary on the CDI, very few children knew minimal pairs exemplifying the contrasts in our stimuli (e.g., /k/ vs. /j/ in "kitty")

vs. “yitty”). The majority (63%) of children knew minimal pairs for 0 of the contrasts, and 94% knew minimal pairs for less than half of the contrasts. This suggests that infants adopt phonologically specified representations of word forms even in the absence of direct pressure from the lexicon.

We also found no relationship between vocabulary size and mispronunciation sensitivity. Children with larger vocabularies did show better word recognition (bigger effect of naming on how much they looked at a given picture considering the correct-pronunciation trials, and higher rates of raw target looking on average across conditions), but the strength of the advantage for correct pronunciations over mispronunciations was not related to vocabulary size. This is consistent with several previous studies investigating older toddlers (e.g., Mani & Plunkett, 2007; Swingley, 2009b). If the development of phonologically precise word representations were contingent on the pressures or learning opportunities provided by knowing more words, we would expect a significant positive relationship. Thus our results suggest instead that infants represent words’ phonological forms accurately, at least to the tested level of detail, regardless of how many words they know.

An alternative possibility which our results do not rule out is that infants’ representations of word forms are precise but not phonological in nature. For example, rather than representing speech in terms of categorical phonological segments, infants’ representations might be probabilistic, e.g., probability distributions over phonological segments (e.g., Caplan et al., 2021); or categorical but not phonological, e.g., fine-grained acoustic categories acquired via unsupervised distributional learning (Schatz et al., 2021); or neither categorical nor phonological, as in perceptual space-learning models (Feldman et al., 2021). In principle, any of these kinds of representations could explain young infants’ sensitivity to single-phoneme mispronunciations, although accounts that propose that adults but not infants use phonological categories must eventually explain how learners advance from one stage to the other. Our results do not adjudicate between these different representational accounts, but they do rule out the possibility that infants’ knowledge is extremely vague or underdetermined, at least by 12 months.

While not the focus of the investigation, this work also provides further evidence that infants’ word recognition can be tested remotely over the internet. Testing infants in the home may have yielded a higher participant exclusion rate compared to in-lab studies (due to more distractions and equipment failures), but the infants retained in the final sample, at least in the two older age groups, engaged with the task online and recognized words at significantly above-chance rates. Infants in this study all heard the same materials pre-recorded by a female research assistant, rather than produced live by their parent (e.g., Bergelson & Swingley, 2012), yet both the 9- to 11-month-olds and 12- to 14-month-olds successfully generalized to the unfamiliar talker’s speech.

Unexpectedly, the 6- to 8-month-old group we tested failed to demonstrate significant recognition of the correct pronunciations. Proportion target looking in this condition-age combination was not significantly above chance (50%) and which image was named had no significant effect on infants’ proportion image looking, in contrast with previous in-lab research using some of the same words and words with similar reported age of acquisition

(Bergelson & Swingley, 2012). Performance at this age is variable (e.g., Bergelson & Swingley, 2018), thus we might have failed to find an effect by chance. However, another possibility is that these 6- to 8-month-olds performed less well due to our testing method: testing infants at home on parents' personal computers over the internet, as opposed to traditional in-lab testing. If so, then our results could have important practical implications for future online versus in-lab research with infants at these ages. Online data collection poses less of a burden on parents' and lab staff's time and has the potential to make participating in research accessible to a wider diversity of families (although we did not see evidence of this here), but the noise (e.g., home distractions, missing data due to glitches in webcam recording) that online data collection introduces may make it less practical for very young infants.

This study builds on a body of previous work using language-guided looking to investigate infants' word recognition and sensitivity to deliberate mispronunciations. While several recent studies have demonstrated that infants recognize words far earlier in development than previously thought (Bergelson & Swingley, 2012, 2015; Nomikou et al., 2019; Parise & Csibra, 2012; Syrnyk & Meints, 2017), little research has explored how mispronunciations affect word recognition at these early ages. Our results here are consistent with the idea that infants are sensitive to one-phoneme changes to words' initial consonants from as soon as they are shown empirically to know those words (9 months in this sample). It is possible that, as with older toddlers, smaller mispronunciations would elicit a smaller effect (e.g., White & Morgan, 2008), but the present study begins to set lower bounds on the precision of infants' word-form representations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements:

This work received financial support from NSF grant 1917608 to Daniel Swingley, from NIH (NIDCD) grant T32 DC016903, on which Caroline Beech was a trainee, and from NIH (NICHD) grant F31 HD113235 to Caroline Beech. We would also like to thank former lab manager Anna Runova and the Penn undergraduate RAs who assisted with this project.

References

- Altwater-Mackensen N, van der Feest SVH, & Fikkert P (2014). Asymmetries in early word recognition: The case of stops and fricatives. *Language Learning and Development*, 10(2), 149–178. 10.1080/15475441.2013.808954
- Bacon D, Weaver H, & Saffran J (2021). A framework for online experimenter-moderated looking-time studies assessing infants' linguistic knowledge. *Frontiers in Psychology*, 12. 10.3389/fpsyg.2021.703839
- Bates D, Maechler M, Bolker B, & Walker S (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. 10.18637/jss.v067.i01
- Beech C, & Swingley D (2023). Consequences of phonological variation for algorithmic word segmentation. *Cognition*, 235, 105401. 10.1016/j.cognition.2023.105401 [PubMed: 36787685]
- Beech C, & Swingley D (2026, May 31). Consonants in English-learning infants' recognition of familiar words. Retrieved from osf.io/ah2nv
- Bergelson E (2017). Bergelson SEEDLingS HomeBank corpus. 10.21415/T5PK6D

- Bergelson E, & Swingley D (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, 109(9), 3253–3258. 10.1073/pnas.1113380109 [PubMed: 22331874]
- Bergelson E, & Swingley D (2013). The acquisition of abstract words by young infants. *Cognition*, 127(3), 391–397. 10.1016/j.cognition.2013.02.011 [PubMed: 23542412]
- Bergelson E, & Swingley D (2015). Early word comprehension in infants: Replication and extension. *Language Learning and Development*, 11(4), 369–380. 10.1080/15475441.2014.979387 [PubMed: 26664329]
- Bergelson E, & Swingley D (2018). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*, 89(5), 1567–1576. 10.1111/cdev.12888 [PubMed: 28639708]
- Bernier DE, & White KS (2019). Toddlers' sensitivity to phonetic detail in child speech. *Journal of Experimental Child Psychology*, 185, 128–147. 10.1016/j.jecp.2019.04.021 [PubMed: 31132672]
- Best CT, Tyler MD, Gooding TN, Orlando CB, & Quann CA (2009). Development of phonological constancy: Toddlers' perception of native- and Jamaican-accented words. *Psychological Science*, 20(5), 539–542. 10.1111/j.1467-9280.2009.02327.x [PubMed: 19368700]
- Bouchon C, Floccia C, Fux T, Adda-Decker M, & Nazzi T (2015). Call me Alix, not Elix: vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, 18(4), 587–598. 10.1111/desc.12242 [PubMed: 25294431]
- Brent MR, & Siskind JM (2001). The role of exposure to isolated words in early vocabulary development. *Cognition*, 81(2), B33–B44. 10.1016/S0010-0277(01)00122-6 [PubMed: 11376642]
- Bulgarelli F (2024). Infants' comprehension of speech produced by children [Poster presentation]. International Congress of Infant Studies 2024, Glasgow, Scotland. https://ubwp.buffalo.edu/buffaloearlylearning/wp-content/uploads/sites/194/2024/07/Bulgarelli_ICIS2024_Poster_.pdf
- Bulgarelli F, & Bergelson E (2022). Talker variability shapes early word representations in English-learning 8-month-olds. *Infancy*, 27(2), 341–368. 10.1111/infa.12452 [PubMed: 34997679]
- Bürkner P-C (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80, 1–28. 10.18637/jss.v080.i01
- Caplan S, Hafri A, & Trueswell JC (2021). Now you hear me, later you don't: The immediacy of linguistic computation and the representation of speech. *Psychological Science*, 32(3), 410–423. 10.1177/0956797620968787 [PubMed: 33617735]
- Carbajal MJ, Peperkamp S, & Tsuji S (2021). A meta-analysis of infants' word-form recognition. *Infancy*, 26(3), 369–387. 10.1111/infa.12391 [PubMed: 33620781]
- Charles-Luce J, & Luce PA (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language*, 22(3), 727–735. 10.1017/S0305000900010023 [PubMed: 8789521]
- Chuey A, Boyce V, Cao A, & Frank MC (2024). Conducting developmental research online vs. in-person: A meta-analysis. *Open Mind*, 8, 795–808. 10.1162/opmi_a_00147 [PubMed: 38957506]
- Cole RA (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics*, 13(1), 153–156. 10.3758/BF03207252
- Cooper A, Paquette-Smith M, Bordignon C, & Johnson EK (2023). The influence of accent distance on perceptual adaptation in toddlers and adults. *Language Learning and Development*, 19(1), 74–94. 10.1080/15475441.2022.2047688
- Creel S (2022). Learning words amidst speech sound variability. In Papafragou A, Trueswell JC, & Gleitman LR (Eds.), *The Oxford Handbook of the Mental Lexicon* (pp. 291–310). Oxford University Press. 10.1093/oxfordhb/9780198845003.013.7
- Cui A (2020). The emergence of phonological categories (Publication no. 27835855) [Doctoral dissertation, University of Pennsylvania]. ProQuest Dissertations Publishing.
- Curtis PR, Estabrook R, Roberts MY, & Weisleder A (2023). Specificity of phonological representations in U.S. English-speaking late talkers and typical talkers. *Infancy*, 28(4), 771–792. 10.1111/infa.12536 [PubMed: 36939533]
- Cychosz M, Mahr T, Munson B, Newman R, & Edwards JR (2023). Preschoolers rely on rich speech representations to process variable speech. *Child Development*, 94(4), e197–e214. 10.1111/cdev.13922 [PubMed: 37036081]

- Delle Luche C, Floccia C, Granjon L, & Nazzi T (2017). Infants' first words are not phonetically specified: Own name recognition in British English-learning 5-month-olds. *Infancy*, 22(3), 362–388. 10.1111/inf.12151 [PubMed: 33158358]
- Durrant S, Delle Luche C, Cattani A, & Floccia C (2015). Monodialectal and multidialectal infants' representation of familiar words. *Journal of Child Language*, 42(2), 447–465. 10.1017/S0305000914000063 [PubMed: 24655564]
- Eimas PD, & Miller JL (1980). Contextual effects in infant speech perception. *Science*, 209(4461), 1140–1141. 10.1126/SCIENCE.7403875 [PubMed: 7403875]
- Eimas PD, Siqueland ER, Jusczyk P, & Vigorito J (1971). Speech perception in infants. *Science*, 171(3968), 303–306. 10.1126/SCIENCE.171.3968.303 [PubMed: 5538846]
- ELAN (Version 6.4) [Computer software]. (2022). Max Planck Institute for Psycholinguistics, The Language Archive. <https://archive.mpi.nl/tla/elan>
- Feldman NH, Goldwater S, Dupoux E, & Schatz T (2021). Do infants really learn phonetic categories? *Open Mind*, 5, 113–131. 10.1162/opmi_a_00046 [PubMed: 35024527]
- Feldman NH, Griffiths TL, Goldwater S, & Morgan JL (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review*, 120(4), 751–778. 10.1037/a0034245 [PubMed: 24219848]
- Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ, Tomasello M, Mervis CB, & Stiles J (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), 1–185. 10.2307/1166093
- Fernald A, Pinto JP, Swingley D, Weinberg A, & McRoberts GW (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9(3), 228–231. 10.1111/1467-9280.00044
- Frank MC, Braginsky M, Yurovsky D, & Marchman VA (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, 44(3), 677–694. 10.1017/S0305000916000209 [PubMed: 27189114]
- Hallé PA, & de Boysson-Bardies B (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, 19(4), 463–481. 10.1016/S0163-6383(96)90007-7
- Höhle B, Van De Vijver R, & Weissenborn J (2006). Word processing at 19 months and its relation to language performance at 30 months: A retrospective analysis of data from German learning children. *International Journal of Speech-Language Pathology*, 8(4), 356–363. 10.1080/14417040600970614
- Højten A, Madsen TO, & Bleses D (2023). Danish 20-month-olds' recognition of familiar words with and without consonant and vowel mispronunciations. *Phonetica*, 80(5), 309–328. 10.1515/phon-2023-2001 [PubMed: 37533184]
- Humphrey C (2022). glmmboot: Bootstrap resampling for mixed effects and plain models (0.6.0). <https://github.com/ColmanHumphrey/glmmboot>
- Johnson EK, & White KS (2020). Developmental sociolinguistics: Children's acquisition of language variation. *Wiley Interdisciplinary Reviews: Cognitive Science*, 11(1), 1–15. 10.1002/wcs.1515
- Kuhl PK, Stevens E, Hayashi A, Deguchi T, Kiritani S, & Iverson P (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, 9(2), F13–F21. 10.1111/j.1467-7687.2006.00468.x [PubMed: 16472309]
- Lahey M, & Ernestus M (2014). Pronunciation variation in infant-directed speech: Phonetic reduction of two highly frequent words. *Language Learning and Development*, 10(4), 308–327. 10.1080/15475441.2013.860813
- Law F, & Edwards JR (2015). Effects of vocabulary size on online lexical processing by preschoolers. *Language Learning and Development*, 11(4), 331–355. 10.1080/15475441.2014.961066 [PubMed: 26508903]
- Lenth RV (2023). emmeans: Estimated marginal means, aka least-squares means (1.8.4–1). <https://cran.r-project.org/package=emmeans>
- Li D, Zhong S, & Schuler K (2023). Adapting infant looking time paradigms for the web. *Proceedings of the 47th Annual Boston University Conference on Language Development (BUCLD 47)*, 499–504. <https://www.lingref.com/bucl/47/BUCLD47-41.pdf>

- Lovcevic I, Burnham D, & Kalashnikova M (2024). Infants' lexical processing: Independent contributions of attentional and clarity cues. *Language Learning and Development*, 20(1), 1–18. 10.1080/15475441.2022.2149402
- Mani N, & Plunkett K (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, 57(2), 252–272. 10.1016/j.jml.2007.03.005
- Mani N, & Plunkett K (2010). Twelve-month-olds know their cups from their keps and tups. *Infancy*, 15(5), 445–470. 10.1111/J.1532-7078.2009.00027.X [PubMed: 32693507]
- Metsala JL, & Walley AC (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In Metsala JL & Ehri LC (Eds.), *Word recognition in beginning literacy* (1st ed., pp. 89–120). Routledge. <http://ebookcentral.proquest.com/lib/upenn-ebooks/detail.action?docID=474573>
- Moore C, & Bergelson E (2022). Examining the roles of regularity and lexical class in 18 – 26-month-olds' representations of how words sound. *Journal of Memory and Language*, 126, 104337. 10.1016/j.jml.2022.104337
- Nazzi T, & Cutler A (2019). How consonants and vowels shape spoken-language recognition. *Annual Review of Linguistics*, 5, 25–47. 10.1146/annurev-linguistics-011718-011919
- Nguyen DKL, Fitzpatrick N, & Floccia C (2024). Adapting language development research paradigms to online testing: Data from preferential looking, word learning and vocabulary assessment in toddlers. *Journal of Child Language*, 1–33. 10.1017/S0305000924000035 [PubMed: 36353801]
- Nittrouer S (2021). Speech perception by children: The structural refinement and differentiation model. In Pardo JS, Nygaard LC, Remez RE, & Pisoni DB (Eds.), *The Handbook of Speech Perception* (pp. 485–516). Wiley.
- Nomikou I, Rohlfing KJ, Cimiano P, & Mandler JM (2019). Evidence for early comprehension of action verbs. *Language Learning and Development*, 15(1), 64–74. 10.1080/15475441.2018.1520639
- Parise E, & Csibra G (2012). Electrophysiological evidence for the understanding of maternal speech by 9-Month-old infants. *Psychological Science*, 23(7), 728–733. 10.1177/0956797612438734 [PubMed: 22692337]
- Poltrock S, & Nazzi T (2015). Consonant/vowel asymmetry in early word form recognition. *Journal of Experimental Child Psychology*, 131, 135–148. 10.1016/j.jecp.2014.11.011 [PubMed: 25544396]
- Quam C, & Swingley D (2010). Phonological knowledge guides 2-year-olds' and adults' interpretation of salient pitch contours in word learning. *Journal of Memory and Language*, 62(2), 135–150. 10.1016/j.jml.2009.09.003 [PubMed: 20161601]
- Quam C, & Swingley D (2014). Processing of lexical stress cues by young children. *Journal of Experimental Child Psychology*, 123(1), 73–89. 10.1016/j.jecp.2014.01.010 [PubMed: 24705094]
- Quam C, & Swingley D (2023). A protracted developmental trajectory for English-learning children's detection of consonant mispronunciations in newly learned words. *Language Acquisition*, 30(3–4), 256–276. 10.1080/10489223.2022.2069026 [PubMed: 37377488]
- R Core Team. (2022). R: A language and environment for statistical computing (4.2.2). R Foundation for Statistical Computing. <https://www.r-project.org/>
- Ramon-Casas M, Swingley D, Sebastián-Gallés N, & Bosch L (2009). Vowel categorization during word recognition in bilingual toddlers. *Cognitive Psychology*, 59(1), 96–121. 10.1016/j.cogpsych.2009.02.002 [PubMed: 19338984]
- Ratnage P, Nazzi T, & Floccia C (2023). Vowels and consonants matter equally to British English-learning 11-month-olds' familiar word form recognition. *Journal of Child Language*, 1–24. 10.1017/S0305000923000223
- Ren J, Cohen Priva U, & Morgan JL (2019). Underspecification in toddlers' and adults' lexical representations. *Cognition*, 193, 103991. 10.1016/j.cognition.2019.06.003 [PubMed: 31525643]
- Schatz T, Feldman NH, Goldwater S, Cao XN, & Dupoux E (2021). Early phonetic learning without phonetic categories: Insights from large-scale simulations on realistic input. *Proceedings of the National Academy of Sciences of the United States of America*, 118(7). 10.1073/pnas.2001844118
- Schmale R, Cristià A, Seidl A, & Johnson EK (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy*, 15(6), 650–662. 10.1111/J.1532-7078.2010.00032.X [PubMed: 32693460]

- Scott K, Chu J, & Schulz L (2017). Lookit (Part 2): Assessing the viability of online developmental research, results from three case studies. *Open Mind*, 1(1), 15–29. 10.1162/opmi_a_00001
- Singh L, & Cheng Q (2023). Words in context: Compensation for phonological assimilation in monolingual and bilingual toddlers. *First Language*, 43(4), 407–430. 10.1177/01427237231161135
- Skoruppa K, Mani N, Plunkett K, Cabrol D, & Peperkamp S (2013). Early word recognition in sentence context: French and English 24-month-olds' sensitivity to sentence-medial mispronunciations and assimilations. *Infancy*, 18(6), 1007–1029. 10.1111/infa.12020
- Sonderegger M (2023). *Regression modeling for linguistic data*. MIT Press.
- Steffan A, Zimmer L, Arias-Trejo N, Bohn M, Dal Ben R, Flores-Coronado MA, Franchin L, Garbisch I, Grosse Wiesmann C, Hamlin JK, Havron N, Hay JF, Hermansen TK, Jakobsen KV, Kalinke S, Ko ES, Kulke L, Mayor J, Meristo M, ... Schuwerk T (2024). Validation of an open source, remote web-based eye-tracking method (WebGazer) for research in early childhood. *Infancy*, 29(1), 31–55. 10.1111/infa.12564 [PubMed: 37850726]
- Storkel HL (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language*, 29(2), 251–274. 10.1017/S0305000902005032 [PubMed: 12109371]
- Swingley D (2003). Phonetic detail in the developing lexicon. *Language and Speech*, 46(2–3), 265–294. 10.1177/00238309030460021001 [PubMed: 14748447]
- Swingley D (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, 8(5), 432–443. 10.1111/j.1467-7687.2005.00432.x [PubMed: 16048516]
- Swingley D (2009a). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1536), 3617–3632. 10.1098/rstb.2009.0107
- Swingley D (2009b). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, 60(2), 252–269. 10.1016/j.jml.2008.11.003 [PubMed: 20126290]
- Swingley D (2012). The looking-while-listening procedure. In Hoff E (Ed.), *Research Methods in Child Language* (pp. 29–42). Wiley-Blackwell. 10.1002/9781444344035.ch3
- Swingley D (2022). Infants' learning of speech sounds and word forms. In Papafragou A, Trueswell JC, & Gleitman LR (Eds.), *The Oxford Handbook of the Mental Lexicon* (pp. 267–291). Oxford University Press. 10.1093/oxfordhb/9780198845003.013.6
- Swingley D, & Aslin RN (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, 76(2), 147–166. 10.1016/S0010-0277(00)00081-0 [PubMed: 10856741]
- Swingley D, & Aslin RN (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, 13(5), 480–484. 10.1111/1467-9280.00485 [PubMed: 12219818]
- Syrnyk C, & Meints K (2017). Bye-bye mummy – Word comprehension in 9-month-old infants. *British Journal of Developmental Psychology*, 35(2), 202–217. 10.1111/bjdp.12157 [PubMed: 27621053]
- Tao Y, & Qinmei X (2013). Phonological specificity of lexical tones in 12-month-old Chinese-speaking infants. *Acta Psychologica Sinica*, 45(10), 1111–1118. 10.3724/sp.j.1041.2013.01111
- Tincoff R, & Jusczyk PW (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, 17(4), 432–444. 10.1111/j.1532-7078.2011.00084.x [PubMed: 32693484]
- Tsuji S, Fikkert P, Yamane N, & Mazuka R (2016). Language-general biases and language-specific experience contribute to phonological detail in toddlers' word representations. *Developmental Psychology*, 52(3), 379–390. 10.1037/dev0000093 [PubMed: 26891017]
- van der Feest SVH, & Johnson EK (2016). Input-driven differences in toddlers' perception of a disappearing phonological contrast. *Language Acquisition*, 23(2), 89–111. 10.1080/10489223.2015.1047096
- Vihman M (2022). The developmental origins of phonological memory. *Psychological Review*, 129(6), 1495–1508. 10.1037/rev0000354 [PubMed: 35175093]
- Vihman M, & Croft W (2007). Phonological development: Toward a “radical” templatic phonology. *Linguistics*, 45(4), 683–725. 10.1515/LING.2007.021

- Vihman M, & Majorano M (2017). The role of geminates in infants' early word production and word-form recognition. *Journal of Child Language*, 44(1), 158–184. 10.1017/S0305000915000793 [PubMed: 26767502]
- Vihman M, Nakai S, DePaolis RA, & Hallé P (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, 50(3), 336–353. 10.1016/j.jml.2003.11.004
- Von Holzen K, & Bergmann C (2021). The development of infants' responses to mispronunciations: A meta-analysis. *Developmental Psychology*, 57(1), 1–18. 10.1037/dev0001141 [PubMed: 33382325]
- Wagner L, Clopper CG, & Pate JK (2014). Children's perception of dialect variation. *Journal of Child Language*, 41(5), 1062–1084. 10.1017/S0305000913000330 [PubMed: 23985300]
- Warner N (2019). Reduced speech: All is variability. *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(4), e1496. 10.1002/wcs.1496 [PubMed: 30811095]
- Weaver H, Zettersten M, & Saffran JR (2024). Becoming word meaning experts: Infants' processing of familiar words in the context of typical and atypical exemplars. *Child Development*, 352–372. 10.1111/cdev.14120
- Werk JF, & Curtin S (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, 1(2), 197–234. 10.1080/15475441.2005.9684216
- Werk JF, & Tees RC (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 7(1), 49–63. 10.1016/S0163-6383(84)80022-3
- Wewalaarachchi TD, Wong LH, & Singh L (2017). Vowels, consonants, and lexical tones: Sensitivity to phonological variation in monolingual Mandarin and bilingual English–Mandarin toddlers. *Journal of Experimental Child Psychology*, 159, 16–33. 10.1016/j.jecp.2017.01.009 [PubMed: 28266332]
- White KS, & Aslin RN (2011). Adaptation to novel accents by toddlers. *Developmental Science*, 14(2), 372–384. 10.1111/j.1467-7687.2010.00986.x [PubMed: 21479106]
- White KS, & Morgan JL (2008). Sub-segmental detail in early lexical representations. *Journal of Memory and Language*, 59(1), 114–132. 10.1016/J.JML.2008.03.001
- Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Golemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, ... Yutani H (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, 4(43), 1686. 10.21105/joss.01686
- Zehr J, & Schwarz F (2018). PennController for Internet Based Experiments (IBEX). 10.17605/OSF.IO/MD832
- Zesiger P, Lozeron ED, Lévy A, & Frauenfelder UH (2012). Phonological specificity in 12- and 17-month-old French-speaking infants. *Infancy*, 17(6), 591–609. 10.1111/j.1532-7078.2011.00111.x [PubMed: 32693490]

Public Significance:

This study suggests that around 9 to 14 months of age, when infants are just starting to understand words, their idea of how a familiar word sounds already includes the specific consonant at the start of the word.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



Figure 1. Example trial

Note. Accompanying this visual display, infants would hear a sentence naming one of the pictures using either a correct pronunciation (e.g., “Find the ball!”) or initial-consonant mispronunciation (e.g., “Find the chall!”).

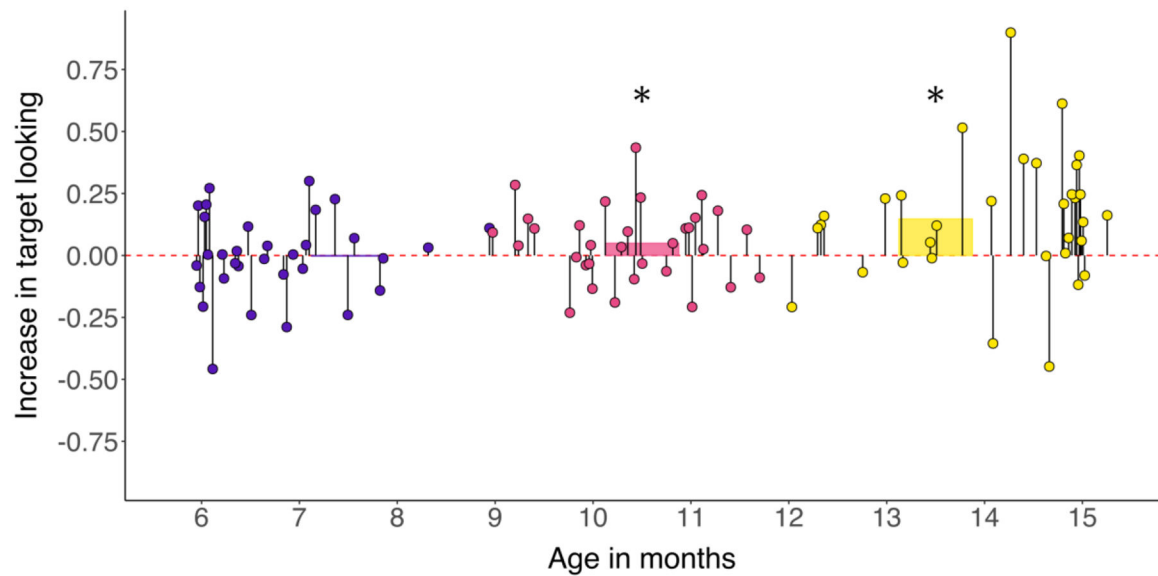


Figure 2. Increase in target looking by age for correct pronunciation stimuli only

Note. The y-axis indicates infants' mean difference scores (proportion looking to a particular image when it was named, minus proportion looking to that image when the other image in the pair was named) in the 367-to-3500-ms window. Points represent individual participants, colored according to age group, and bars represent the mean for each age group (starred if performance was significantly above chance). As age increases, there are more points above zero, or infants who looked more to images when they were named, demonstrating successful recognition of the correct pronunciations.

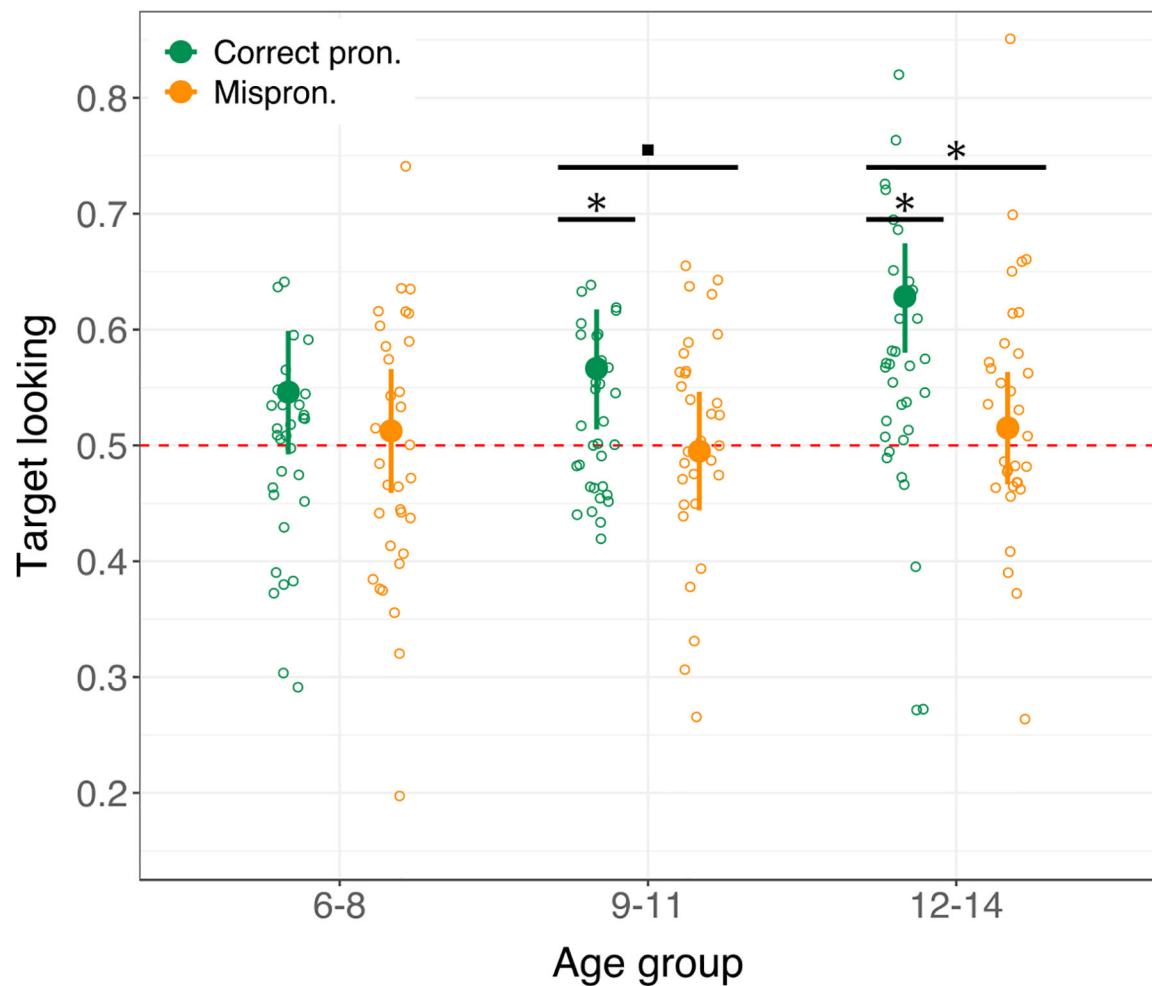


Figure 3. Proportion target looking (out of looking to either the target or the distracter image during the 367-to-3500-ms window) elicited by correct pronunciation stimuli (green) versus mispronunciation stimuli (orange) for each age group

Note. Open points show the subject means (computed over all items for which that subject had data for both conditions). Filled points and segments show the model estimates.

Annotations indicate ages at which there was an effect of condition with a p -value of <0.05 (*) or <0.1 (.), and age-condition combinations where target looking was significantly ($p < 0.05$, *) above chance according to a mixed effects logistic regression model.

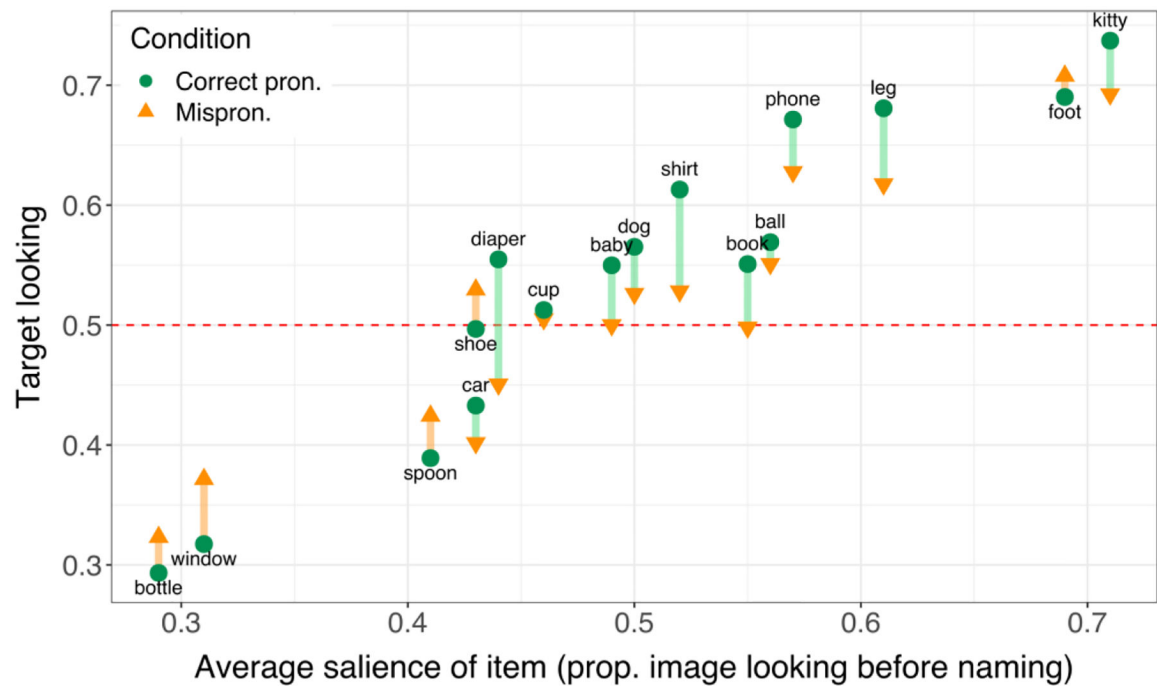


Figure 4. Item-by-item target looking in 9- to 14-month-olds

Note. Target items are arrayed according to average picture salience, with the least-favored pictures at left (bottle, window) and the most-preferred pictures at right (foot, kitty). Green circles indicate mean target fixation on correct-pronunciation (CP) trials; orange triangles indicate looking on mispronunciation (MP) trials. These points are connected with a line: green when CP looking was greater than MP looking, orange for the reverse.

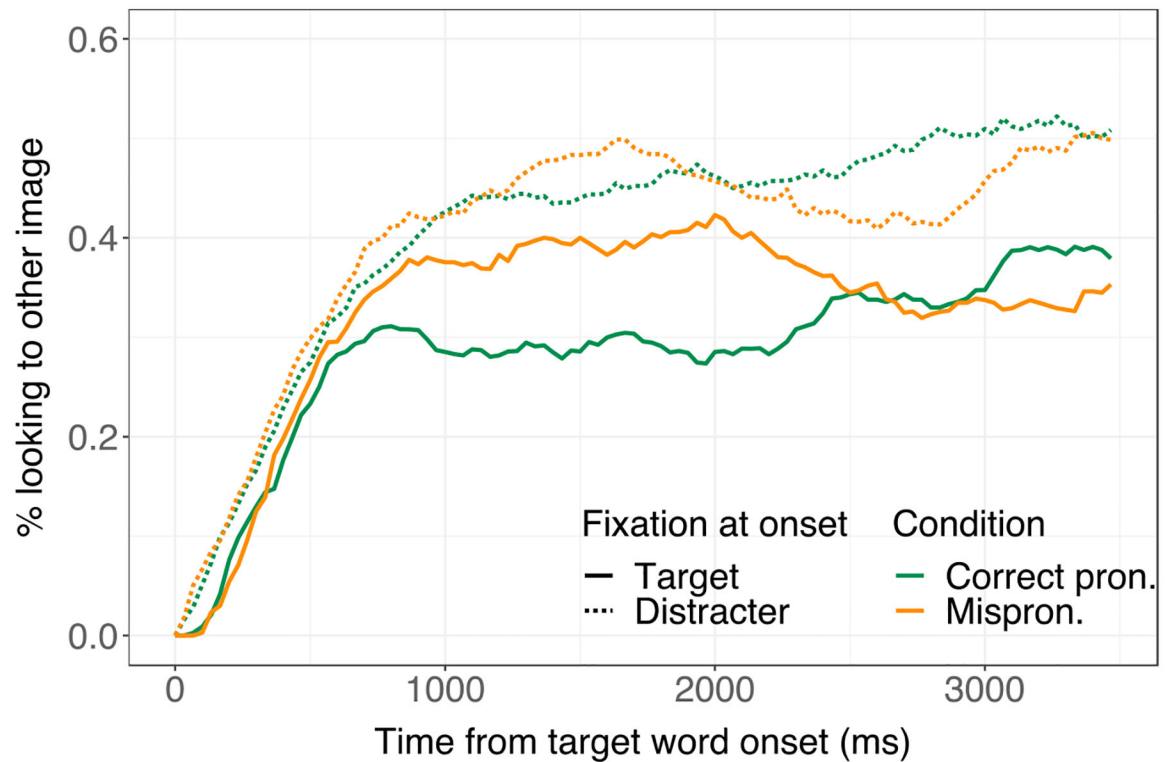


Figure 5. Time-course of fixations, considering 9- to 14-month-olds

Note. Trials are divided according to condition (correct pronunciation (green) vs. mispronunciation (orange)) and according to where children happened to be looking when the target word began (target (solid line) vs. distracter/off-target (dashed line)). The x-axis shows time, starting from the acoustic onset of the spoken target word. The y-axis shows the proportion of trials on which infants were, at that moment, looking away from their initial gaze location.

Table 1

Target word stimuli (MPs varied from CPs only in their initial consonants)

Correct pronunciation (CP)	Mispronunciation (MP)	Paired Word
baby	raby	dog
ball	chall	shoe
book	sook	diaper
bottle	nottle	kitty
car	lar	phone
kitty	yitty	bottle
cup	wup	shirt
diaper	myper	book
dog	mog	baby
foot	yoot	window
leg	feg	spoon
phone	roan	car
shirt	lirt	cup
shoe	woo	ball
spoon	croon	leg
window	kindow	foot

Table 2
Recognition model results, for correct pronunciation trials only

Coefficient	Beta	SE	Z	p
(Intercept)	0.00	0.04	−0.11	0.911
X named (true vs. on average)	0.25	0.05	5.07	<0.001 *
Age group (6–8 months vs. 9–11 months)	0.01	0.11	0.14	0.889
Age group (9–11 months vs. 12–14 months)	0.01	0.10	0.07	0.942
Location at onset (already fixating x vs. on average)	0.54	0.06	9.71	<0.001 *
Salience	3.49	0.23	14.89	<0.001 *
Side (right vs. on average)	−0.01	0.04	−0.13	0.898
X named : Age group (6–8 months vs. 9–11 months)	−0.09	0.12	−0.73	0.465
X named : Age group (9–11 months vs. 12–14 months)	−0.27	0.12	−2.30	0.022 *

Number of observations: 2654. Groups: item (16), subject (97).

* $p < 0.05$.

Note. Results of a mixed effects logistic regression model predicting trial-level looking to a particular image (proportion_x_looking) from whether that image or the other image in the pair was named (x_named). The model was estimated using by-item and by-subject random intercepts and random slopes for x_named, and p-values were calculated via Wald z-tests. Comparisons in this table described as “A vs. B” denote the effect of A (positive if $A > B$, holding all other variables constant).

Table 3

Mispronunciation sensitivity model results

Coefficient	Beta	SE	Z	p
(Intercept)	0.18	0.04	4.08	<0.001 *
Condition (correct pronunciation vs. on average)	0.15	0.04	3.33	0.001 *
Age group (6–8 months vs. 9–11 months)	−0.01	0.11	−0.06	0.954
Age group (9–11 months vs. 12–14 months)	−0.17	0.10	−1.65	0.099 .
Location at onset (already fixating target vs. on average)	0.32	0.05	7.20	<0.001 *
Salience	3.67	0.23	16.04	<0.001 *
Side (right vs. on average)	0.04	0.04	0.96	0.337
Condition : Age group (6–8 months vs. 9–11 months)	−0.08	0.11	−0.71	0.475
Condition : Age group (9–11 months vs. 12–14 months)	−0.09	0.10	−0.87	0.385
Location at onset : Condition	0.09	0.04	2.03	0.043 *

Number of observations: 2656. Groups: item (16), subject (97).

.p < 0.1.

*p < 0.05.

Note. Results of a mixed effects logistic regression model predicting trial-level looking to the target image (proportion_target_looking) from condition (correct pronunciation vs. mispronunciation). The model was estimated using by-item and by-subject random intercepts and random slopes for condition, and p-values were calculated via Wald z-tests. Comparisons in this table described as “A vs. B” denote the effect of A (positive if A>B, holding all other variables constant).