A Protracted Developmental Trajectory for English-Learning Children's Detection of Consonant Mispronunciations in Newly Learned Words

*Carolyn Quam**

Department of Speech and Hearing Sciences, Portland State University, USA
Department of Psychology, University of Pennsylvania, USA

Daniel Swingley

Department of Psychology, University of Pennsylvania, USA

Abstract

Children are adept at learning their language's speech-sound categories, but just how these categories function in their developing lexicon has not been mapped out in detail. Here, we addressed whether, in a language-guided looking procedure, two-year-olds would respond to a mispronunciation of the voicing of the initial consonant of a newly learned word. First, to provide a baseline of mature native-speaker performance, adults were taught a new word under training conditions of low prosodic variability. In a second experiment, 24- and 30-month-olds were taught a new word under training conditions of high or low prosodic variability. Children and adults showed evidence of learning the taught word. Adults' target looking was reduced when the novel word was realized at test with a change in the voicing of the initial consonant, but children did not show any such decrement in target fixation. For both children and adults, most learners did not treat the phonologically distinct variant as a different word. Acoustic-phonetic variability during teaching did not have consistent effects. Thus, under conditions of intensive short-term training, 24- and 30-month-olds did not differentiate a newly learned word from a variant differing only in consonant voicing. High task complexity during training could explain why mispronunciation detection was weaker here than in some prior studies.

[210 words]

Keywords: word learning; phonology; processing; prosody

* Corresponding author. Portland State University Speech and Hearing Sciences, PO Box 751, Portland, OR 97207-0751, USA. 1-503-725-3558. cquam@pdx.edu.

Introduction

The present study investigates the degree to which two-year-olds' learning and recognition of novel words is guided by their knowledge of native-language phonological categories. The notion of phonological categories is central to psycholinguistic accounts of word recognition and word learning. A language's phonology, including its set of contrastive categories, provides standards that determine which phonetic sequences "count" as the same word, and which are distinct. These standards may be described in terms of phonological contrasts; thus, the English words "push" and "bush" count as different words because English contrasts the categories /p/ and /b/. Because different languages use different contrasting categories, the categories must be learned. Infants make substantial progress in learning these categories in the first year of life, as demonstrated by a reduction in discrimination for some non-native contrasts, and improvement in discrimination of native contrasts (Bosch & Sebastián-Gallés, 2003; Polka & Werker, 1994; Werker & Tees, 1984; see Swingley, 2022, for a review).

While infants demonstrate precocious learning of native sound contrasts, this knowledge does not always seem to be applied in early word learning (see Quam & Creel, 2015, for discussion). This conclusion comes from research that has addressed two questions: (1) does children's knowledge of words include enough phonetic information to differentiate words that are (or could be) phonologically distinct in their language?; and (2) under what conditions do children take phonologically relevant distinctions (and not other phonetic differences) as relevant to lexical contrast? As we review below, although children represent familiar words with sufficient phonetic detail for making phonological distinctions among words, they do not consistently interpret phonological variation as dictating lexical differentiation.

Toddlers' use of their phonological knowledge in the service of word learning has been probed using several different experimental tasks, all of which have revealed a mixture of successes and failures. These tasks, reviewed below in the section on *Task Difficulty or Complexity in Word-Learning Tasks,* include teaching two phonologically similar words and testing learning (e.g., Stager & Werker, 1997); teaching one or more words and probing a difference in recognition performance between the taught pronunciation and a deviant one (e.g., Ballem & Plunkett, 2005); and teaching a novel word that is phonologically similar to a familiar word (e.g., Swingley & Aslin, 2007). Based on the results of studies such as these, we cannot be confident that very young children "translate" the speech signal into language-specific categories (consonants, vowels, tones, etc.) and base their lexical categorization of novel words on a strict phonological comparison of familiar and unfamiliar sequences.

A gap between children's knowledge of native-language sound categories and their use of these same categories in word-learning tasks could indicate a failure to encode novel word representations in full phonological detail. Or, it could be an issue not of representation but of failure to demonstrate phonological knowledge under the specific demands of the task, either because the cognitive demands of the task prevent it, or because children have not yet grasped the relationship between phonetic variation and lexical variation (Werker & Curtin, 2005). Comparison of toddlers' learning and differentiation of similar-sounding words across experiments can help disentangle representation-level explanations from performance-level explanations. As we will see, the degree to which children's behavior aligns with predictions based on categorical phonological comparison of speech sounds depends on a number of factors: word familiarity, task difficulty or complexity, acoustic-phonetic variability during training, number of phonetic features mispronounced, and discourse context. Here, we briefly review these

factors, and then present a study evaluating the potential impact of acoustic-phonetic variability on children's sensitivity to consonantal variation in pronunciation.

Word Familiarity

Many prior studies of early phonological knowledge have employed highly frequent words as test stimuli, to maximize the likelihood that most infants will have had sufficient exposure to them. Highly frequent words may present a "best-case scenario," on the plausible hypothesis that fidelity in phonological representations is a function of exposure frequency. Infants' representations of at least some familiar words are phonologically intact by 11 months (Hallé and de Boysson-Bardies, 1996; but see Bergelson & Swingley, 2018; Segal, Keren-Portnoy, & Vihman, 2020; Swingley, 2005). The available evidence suggests that toddlers encode highly familiar, early acquired words like "ball" or "dog" with substantial fidelity to the canonical phonological form. Much of this evidence comes from studies in which children are presented with pictures on a display, one of which is picked out by an utterance like "Which one is the ball?" Children's eye movements to the named picture are monitored. Typically, from about 12 months onward, children look at the named object less when the target word ("ball") is spoken with a deviant pronunciation ("gall," "bool") than when it is spoken canonically. This pattern has been taken as evidence that by the second year, children encode at least some familiar words in a phonologically accurate way (e.g., Mani & Plunkett, 2007; Swingley, 2009; Swingley & Aslin, 2000; White & Morgan, 2008).

Words with which young children have less experience may be lacking in phonological specificity. To evaluate this, researchers have controlled exposure frequency by teaching children invented words with which they presumably have no prior experience. This part of the empirical literature is more complex, as studies have employed a diverse range of teaching and testing

methods at different ages. However, in general, children are less likely to demonstrate knowledge of the phonological detail of newly learned words—as investigated in the present study—than of highly familiar words.

Task Difficulty or Complexity in Word-Learning Tasks

Evaluating children's knowledge of newly learned words in a behavioral task necessarily requires an effective word-teaching method and a means for testing the learning that has taken place. When studies push the limits of children's capacities, it is not always predictable which task features will most effectively allow children to display their knowledge. With this in mind, Stager and Werker (1997) developed the "Switch" habituation method, intending to minimize extraneous demands. In this method's most common implementation, children are exposed to minimally distinct words (like "bin" and "din") and then tested on their reaction to switched word-object mappings. Detecting the switch (as revealed in longer looking times) requires that children encode the difference between the words. In the standard version of this procedure, it is not until 17 months that toddlers learn and differentiate novel minimal pairs differing in their consonants (Werker, Fennell, Corcoran, & Stager, 2002). Again, this contrasts with evidence of phonological specificity in at least some familiar words by 11 months (Hallé and de Boysson-Bardies, 1996). Reducing task demands or clarifying the referential nature of the task can enable successful learning at 14 months (Fennell & Waxman, 2010; Fennell, 2012; Thiessen, 2007). The latter set of studies have led to a more generous interpretation of 14-month-olds' skill in rapid learning of minimal pairs than was implied by the original Stager and Werker experiments. The success of 17-month-olds in the Switch procedure is consistent with work from Nazzi's lab showing, in 20-month-olds, an ability to explicitly group together two objects that have been named with the same

word, excluding an object that was named with a consonantally varied form of that word (e.g., Nazzi & New, 2007).

Compared with the Switch method, the language-guided looking (or "looking while listening") procedure, employed in the present study, has sometimes been argued to have lower task demands (e.g., see Ballem & Plunkett, 2005; Yoshida, Fennell, Swingley, & Werker, 2009). Rather than having to dishabituate to a single object, children look back and forth at a pair of pictures, one of which is named in a sentence. When familiar words are tested, children typically gaze at the named target for a greater proportion of time than the distracter. When the target word is pronounced in a phonologically noncanonical manner, this usually reduces gaze proportions to the target picture (e.g., Swingley & Aslin, 2000). Pronunciation changes generally only cause this gaze reduction when the change signals a different phonological category in the test language (e.g., Ramon-Casas, Swingley, Sebastian-Gallés, & Bosch, 2009).

Preferential-looking studies testing newly taught words have revealed less consistent effects of altered pronunciation on recognition than for familiar words. Mani and Plunkett (2008) showed that changes to the vowel of a newly taught monosyllabic word, like "mott" to "mitt," reduced target-object fixation in 14- and 18-month-old children. However, Ballem and Plunkett (2005) did not find significant effects of changing an initial consonant, like "vope" to "zope," in 14-month-olds, though there was a trend in the expected direction in one of the two testing blocks. Swingley (2007) taught 18- to 20-month-old Dutch learners a novel word and probed their sensitivity to a single-feature consonantal substitution (for one stimulus, "droekel" mispronounced as "troekel"), as well as a more substantial change (e.g., "droekel" as "toekel"). Toddlers who were given 14 exposures to the sound form of the word before being explicitly taught what the word meant then looked at the named object less upon hearing the one-feature mispronunciation than

the trained one; children who were given only 8 such exposures did not show this effect. Children in both training groups were sensitive to the larger phonological changes.

The fact that children recognize a deviation from the normal pronunciation of a word by, for example, looking less at the target object, does not imply that children therefore conceive of the altered form as a novel word to which a meaning should be attached. Swingley and Aslin (2007) attempted to teach 19-month-olds a phonological neighbor of a highly familiar word like *dog* (such as *tog*), as a name for a novel toy, and consistently failed. Swingley (2016) reported similar results in 2.5-year-olds. Eighteen-month-olds did succeed in this task if the novel object label resembled a word of a different syntactic class (like "tiv," which resembles "give"; Dautriche, Swingley, & Christophe, 2015), suggesting that by this age, toddlers' representations of phonological distinctions might be intact, under some conditions, but their willingness to interpret phonological distinctions lexically under conditions of lexical competition hinges on something like plausibility considerations (see Dautriche, Fibla, Fievet, & Christophe, 2018; Swingley, 2016).

Although no studies have parametrically explored many of the variations in the word-teaching methods employed in studies of this sort, it seems reasonable to imagine that several aspects of the training method matter. Ballem and Plunkett (2005) and Mani and Plunkett (2008) used a quite simple familiarization in which the object was shown alone on the screen and ostensively labeled, much like the typical Switch training phase but with a fixed number of repetitions. By contrast, Swingley (2007) and Quam and Swingley (2010) employed a more elaborate training sequence involving a simple story with multiple characters and a (thin) plot line. The Dautriche et al. (2015) study fell in between, teaching words using a video of a talking person handling and naming novel toys. These variations reflect researchers' interest in creating

procedures that, while brief, have a plausible connection to word-learning experience outside the lab, and yet that do not present more information than children can handle. Considerations of this sort provide a reason to not rely wholly on relatively artificial procedures like the Switch procedure in evaluating phonological aspects of word learning. It is possible that the successful behavior of longer looking on Switch trials could, in some circumstances, be a result of teaching children to make a phonological distinction that they otherwise would not have made based on their knowledge of the native language, as shown by Yeung, Chen, and Werker (2013). In defense of the habituation procedure, though, such training effects may be limited (e.g., Dietrich, Swingley, & Werker, 2007).

Acoustic-Phonetic Training Variability

One training aspect that has been investigated in some detail is the degree of phonetic variability present in the speech stimuli used to teach a word. In this literature, variability is considered "irrelevant" if it does not affect the sequence of consonants, vowels, tones, etc. that make up the novel word. For example, single-talker vs. multiple-talker training has been considered, or training exemplars that are either consistent or inconsistent in their pitch pattern (in a non-tonal language). The latter type of variability is explored in the present study.

Acoustic-phonetic variability can inhibit or facilitate learning and processing of sounds and words, depending on factors like learners' perceptual skills and the complexity and nature of the task (see Quam & Creel, 2021, for a review). Toddlers' word learning can be facilitated by acoustic-phonetic training variability. Rost and McMurray (2009, 2010; replicated by Quam, Knight, & Gerken, 2017, and Höhle et al., 2020) taught 14-month-olds words using the Switch method. Children were habituated to 18 different male or female voices, rather than the single

female talker used previously (Stager & Werker, 1997). Children hearing multiple talkers successfully learned minimal pairs, whereas in the standard one-talker task 14-month-olds typically fail. A similar facilitation effect has been demonstrated for stimuli spoken by a single talker instructed to produce words with varying pitch patterns and durations (Galle, Apfelbaum, & McMurray, 2015).

Apfelbaum and McMurray (2011) proposed an associative model to account for facilitation from acoustic-phonetic variability for minimal-pair learning at 14 months. In the model, across exemplars, variability on particular acoustic dimensions reduces cue weights between those dimensions and visual objects, while stable, relevant dimensions of contrast build up stronger cue weights. A similar model, WRAPSA (Jusczyk, 1993), is also exemplar-based, and also incorporates cue weights. This model suggests that contrastive dimensions gain stronger cue weights as experience with the native language accumulates. Finally, facilitation from acoustic-phonetic variability is also expected under the PRIMIR framework (Werker & Curtin, 2005), which argues 14-month-olds do not yet process words phonemically, but instead process and store word forms as holistic exemplars.

This holistic-exemplar view of early lexical representation is consistent with a range of studies showing that infants are affected by changes to phonologically irrelevant differences between a trained form and a test form in recognition-based preferential-listening studies. At around 8 months, matching of a familiarized form and a test form can be disrupted by changes to affect (e.g., Singh, Morgan, & White, 2004) or pitch (Singh, White, & Morgan, 2008; see also Houston & Jusczyk, 2000). This makes sense if infants' word-form matching is not dominated by phonological sequences. The fact that infants in these procedures become more successful at

generalizing over phonetic variation by 10 months is also consistent with a developmental trend toward more adultlike phonological interpretation (e.g., Singh et al., 2008).

That said, the impact of non-criterial acoustic-phonetic variation never disappears entirely. For example, adults' word identifications show decrements when a talker's voice changes between familiarization and test (Goldinger, 1996; see also Goldinger, 1998). Several studies have indicated that adults' learning can be affected by non-criterial variability. As with children, studies with adults have shown both positive (e.g., Barcroft & Summers, 2005; Sadakata & McQueen, 2013) and negative (Mullenix & Pisoni, 1990) effects of variability, with inhibitory effects being more likely for learners with weaker perceptual skills (Antoniou & Wong, 2016; Perrachione et al., 2011; Sadakata & McQueen, 2014).

The picture that emerges, then, is that variation introduced in training sometimes helps infants isolate the criterial phonological features, leading to greater generalization over non-phonological variation (Singh, 2008; see also Houston & Jusczyk, 2003); whereas variation imposed between training and test may impair recognition. These considerations motivated us to compare lower- and higher-variability training conditions in the present study.

Phonetic Features and the Context of the Task

Intuitively, a phonological deviation might be expected to have behavioral consequences proportional to the degree of deviation. For example, target-picture looking in a language-guided looking procedure might decline by some proportion for a single-feature mispronunciation, and by some larger proportion for a two-feature mispronunciation. This expectation has been met in some studies, and not others. For example, Bailey and Plunkett (2002), Swingley and Aslin (2002), and Zesiger, Lozeron, Lévy, and Frauenfelder (2012) found little sign of an effect of featural distance

(e.g., "tog" for "dog" hindered recognition just as much as "mog" for "dog"). Similarly, Swingley (2003) found no difference between a change from one common consonant to another, and a change from that common consonant to a very rare one. On the other hand, several studies have found that learners are more likely to detect a mispronunciation if it mismatches the trained word form to a greater extent. White and Morgan (2008) found that 19-month-olds' sensitivity to mispronunciations that mismatched familiar words was graded by phonological distance. The effects of mispronunciations of words like *shoe* by one feature (place of articulation: "foo"), two features (place and voicing: "voo"), or three features (place, voicing, and manner: "goo") were larger for greater numbers of features changed. Even a one-feature mispronunciation of a familiar word reduced fixations to the target picture, but only a three-feature mispronunciation led to a (non-significant) tendency toward greater fixation of the distracter object than the target (but see Mani & Plunkett, 2011, who found gradient sensitivity to acoustic size of vowel mispronunciations—not to number of features—only by 24 months). Similar effects have been found in studies of 22-month-olds tested on familiar words spoken by a child talker (Bernier and White, 2019; Experiment 2); 30-month-olds tested on familiar words using eye-gaze and pupillometry (Tamási, McKean, Gafos, & Höhle, 2019; see also Tamási, McKean, Gafos, Fritzsche, & Höhle, 2017); and adults tested on newly learned words (White, Yee, Blumstein, & Morgan, 2013).

In some cases, differences in outcomes can be traced to differences in testing methods. White and colleagues have presented children with one familiar object and one novel object, perhaps facilitating the interpretation that the mispronounced word was in fact a label for the novel object, whereas Swingley and colleagues have presented children with two familiar objects, making such an interpretation less likely (indeed, restricting the likelihood of such an interpretation

was part of the motivation for Swingley and Aslin's, 2002, experimental design). A difficulty with this account is that toddlers are resistant to interpreting minimal pairs of familiar words as new words, as reviewed above (e.g., Swingley & Aslin, 2007). Still, the use of an unfamiliar vs. familiar distracter word might account for the effect or noneffect of featural difference counts. When a mispronunciation leads to reduced activation of the lexical item corresponding to the fixated target image, and children therefore look away, if the alternative image they land on is a familiar object, they know immediately that it is not a plausible candidate referent, and can speedily shift back to the target. If the alternative image is a novel object, they face greater uncertainty. Thus, they might linger on that object longer, perhaps in proportion to their confidence that the initially fixated picture was not a referent of the spoken word (based partly on the number of features mispronounced). The conclusions that children come to in a given instance may well depend on their developmental stage and on the particulars of the discourse context, though it is important to note that even adults are sometimes willing to accept a variant as a version of the original word, vs. treating it as a novel word form (e.g., White et al., 2013).

Although this account of the role of the novel distracter is speculative, in the present work our testing trials employed a novel distracter image rather than a familiar one, partly on the grounds that this procedure might be more sensitive to effects of stimulus variation. Indeed, children's apparent resistance to considering a variant of a familiar word as a new word entirely might be attenuated or eliminated when the "familiar" word has just been learned moments before, and thus may exert less of a pull in interpretation. Quam and Swingley (2010) found evidence of this in a study of 30-month-olds, described more fully below, in which many children hearing a deviant pronunciation of a newly taught word actually looked more at the distracter than the target, suggesting a novel-word interpretation of the phonologically distinct word.

The Present Study

The experiments presented here continue a line of experimentation exploring the distinction between lexically contrastive variation in words (like substitutions of consonants or vowels) and salient but non-contrastive variation. The first of these experiments (Quam & Swingley, 2010) taught 30-month-olds a novel word, always presented during teaching with a single prominent pitch contour, and then tested recognition of this word (displayed alongside a familiarized, but unnamed, distractor object) spoken with the familiar pronunciation, a variant pronunciation with a quite different pitch contour, or a variant pronunciation with a different vowel (/a/ rather than /i/). Children's recognition of the word spoken with the substituted vowel was significantly impaired, while their recognition of the word spoken with an alternative pitch contour was not impaired at all. This suggested that children had created a representation of the new word that abstracted away from some of its phonetic attributes (namely, those tied to pitch contour), while still being attentive to lexically significant phonological variation.

Given this result, here we employed the same word-teaching procedure to evaluate lexical representations in younger children using a different kind of contrast. We tested sensitivity to a single-feature, word-initial consonantal substitution in 19-, 24-, and 30-month-olds' recognition of a newly taught word. Rather than manipulate pitch contour as a potentially (ungrammatically) lexically contrastive feature, we manipulated the variability of the pitch contours with which the word was presented in teaching, to evaluate the possibility that sensitivity to consonantal changes might be affected by acoustic-phonetic variability.

We tested 30-month-olds to compare children's performance at this age when given the present consonantal contrast vs. a vowel contrast (Quam & Swingley, 2010). We also tested 19-

and 24-month-olds because, having expected that 30-month-olds would differentiate *deebo* and *teebo*, we wanted to examine possible developmental changes in this response. We originally predicted that children would learn words and detect one-feature mispronunciations as early as 19 months. However, 19-month-olds showed inconsistent word learning, suggesting that the narrated story we used for word teaching might have been too complex, given other procedures' success in word teaching using simpler teaching methods at 1.5 years (e.g., Ballem & Plunkett, 2005; Mani & Plunkett, 2008). Thus, here we focus on the work with 24- and 30-month-olds, presenting the 19-month-olds' results in the Supplemental Materials.

In Experiment 1, we tested 18 adults, to confirm the expected developmental endpoint. In Experiment 2, we tested 64 two-year-olds (at 24 and 30 months) in a similar task. Again, as in prior studies (Ballem & Plunkett, 2005; Mani & Plunkett, 2008; Quam & Swingley, 2010; Swingley, 2007), we expected that children's recognition of the newly taught object labels would be hindered by the mispronunciation. The discourse context we used—specifically, use of a novel distracter object—could boost attention to the mispronunciation, by offering a plausible potential referent for the variant pronunciation. Nevertheless, as in Swingley (2016), we expected that the majority of children would *not* treat the consonantal change as indicating another word, which would also be consistent with children's responses to a vowel change (Quam & Swingley, 2010).

Experiment 1

In Experiment 1, we tested adults with the same method used by Quam and Swingley (2010; Experiment 1). Adults were included to establish the developmental endpoint for interpretation of a subtle, one-feature consonant contrast in the particular teaching context used here, to which children's responses in Experiment 2 can be compared. Inclusion of adults also

enabled comparison with Experiment 1 of Quam and Swingley (2010). In that study, all adults detected a vowel mispronunciation, while 75% of them interpreted the divergent word form as a label for the distractor object. Here, we can compare adults' responses to consonantal changes to responses to vowel changes in the prior study.

Method

Participants

Eighteen adults (12 female, 6 male), all native English speakers, were included in the analysis. Participants were recruited at the University of Pennsylvania, in Philadelphia, Pennsylvania, USA, and most were undergraduate students. Trials were only included as usable if the participant fixated the pictures for at least 20 frames during the analysis window, out of a possible 55. For all 18 participants, the number of usable trials in each condition was at least half of the total number of trials (at least 3 of 5 in mispronunciation trials and at least 4 of 8 in correct-pronunciation trials), so no participants were excluded.

Apparatus and Procedure

The method was nearly identical to the one used by Quam and Swingley (2010). A fuller account of the experimental procedure and the visual and auditory stimuli from the word-teaching phase are detailed in Quam and Swingley (2010; Figures 1-3). The task lasted approximately 20 minutes. Adults were taught a novel word, "deebo," in a narrated, animated story. The word was always pronounced with a consistent pitch contour: either a rise-fall contour or a low-falling contour. The word was taught first in a storybook-like narration in which a monkey tried to recruit playmates to play with two toys: a red knobby toy and a purple disk toy. One of the two toys was

labeled the "deebo" 10 times during the animation and 12 more times during an ostensive-labeling phase in which the object was presented alone on the screen. In both these phases, a second novel object was present equally often but was never labeled. All visual stimuli were identical to those used by Quam and Swingley (2010).

In the test phase, adults (unlike children in the subsequent experiment) were tested with two types of mispronunciations: a consonant change and a pitch change. In each test trial, the two novel objects from the story appeared on the screen, and participants heard a question (like "Where's the [target]?") containing the original word or a version with either the initial consonant or the pitch contour altered. Participants' eye movements in response to the question were measured. Adults saw, intermixed, 8 correct-pronunciation (CP) trials, 5 consonant-mispronunciation (consonant-MP) trials, and 5 pitch-MP trials. Interspersed across the ostensive-labeling and test phases, they also saw 69 filler (familiar-word) trials (only 8 of which were coded for eye gaze; the remainder were included to conceal the goals of the study from adults).

Because our primary focus here is on interpretation of consonant changes, a complete analysis of responses to pitch MPs is reported in supplemental materials (Experiment S1). Briefly, pitch MPs did not impact adults' responses. This result indicates that, to some extent, adults' representation of the word's sound forms was abstracted away from the phonetics of the experienced instances. In our interpretation, the pitch features were attributed to the utterances and not to the novel word (Quam & Swingley, 2010).

After the fixation trials were complete, participants were given a questionnaire asking about their recollections of the study and their interpretation of the novel word. The questionnaire assessed conscious awareness of the consonant and pitch MPs. It also asked participants whether

they had interpreted each variant pronunciation as a label for the distracter object or instead as merely a mispronunciation of "deebo."

Auditory Stimuli

Auditory stimuli for the word-teaching phase were identical to those used by Quam and Swingley (2010). The taught word was *deebo*. Correct-pronunciation test sentences were those employed in Quam and Swingley (2010). Consonant-MP versions of these sentences were informally matched in their acoustic properties to the CP versions and were recorded in the same recording sessions by the same speaker. The MP sentences were "Where's the *teebo*?" and "Which one is the *teebo*?" each recorded with rise-fall and low-fall pitch contours, as shown in **Figure 1**. The pitch pattern in the test phase was the same one each participant had heard in the training phase (rise-fall or low falling). **Table A1** (appendix) reports duration, maximum pitch, and mean pitch of each CP and consonant-MP word token (refer to rows labeled *Variability: Low*).

[INSERT FIGURE 1 HERE]

Data Preparation

Eye movements were coded offline, frame by frame, following the procedure reported by Quam and Swingley (2010), using the *SuperCoder* software program (Hollich, 2005), with 33-millisecond resolution. For statistical analyses, we averaged fixation proportions over the time window 200-2000 ms after noun onset (e.g., Swingley, 2009, Quam & Swingley, 2010). Over that time window, we calculated the proportion of target looking: on each trial, the number of frames the participant looked at the deebo object divided by total looking to either picture. Trials with

fewer than 20 usable frames (out of the 55 total frames between 200-2000 ms) were excluded from analysis. We also addressed the possibility that picture preferences might affect target looking by repeating the analyses using preference-corrected fixation proportions, subtracting the target-fixation proportion during the one second prior to noun onset from the target-fixation proportion during the main analysis window (200-2000 ms). While imperfect, this method has often been used in prior studies, and is repeated here for comparability with other studies.

Results and Discussion

For analysis, raw target-fixation proportions over trials were averaged by participant and trial type (CP, consonant-MP, pitch-MP). **Figure 2** displays raw *deebo*-fixation proportions in CP, consonant-MP, and pitch-MP trials, and **Table 1** reports means for CP and consonant-MP trials (means and analyses for pitch-MP trials are reported in Supplemental Materials, Experiment S1). In order to confirm that adults had learned the word, we first compared their target-fixation proportions to chance (50%) in correct-pronunciation (CP) trials, using a two-tailed, one-sample *t* test. Adults' *deebo* fixation in CP trials was significantly above chance (M = 91.4%, SD = 9.8%), $t(17) = 17.96$, $p < .001$. We next evaluated whether the consonant change significantly affected adults' fixation of the *deebo*. In response to the consonant change, adults' *deebo* fixation was not significantly different from chance (M = 61.8%, SD = 32.1%), $t(17) = 1.56$, $p = .14$. Preference-corrected difference scores showed the same patterns, being significantly above chance (0%) in CP trials (M = 36.9%, SD = 12.8%), $t(17) = 12.27$, $p < .001$, but not consonant-change trials (M = 3.0%, SD = 36.2%), $t(17) = 0.35$, $p = .73$.

[INSERT FIGURE 2 HERE]

[INSERT TABLE 1 HERE]

A repeated-measures ANOVA on raw target fixations, with Trial Type (CP, consonant-MP) as the within-subjects predictor, revealed a significant effect of Trial Type, $F(1,17) = 21.76$, $p < .001$, indicating that adults looked significantly less at the *deebo* object in response to the consonant MP than the CP (*mean decrease* = 29.7%). This decrement was shown (numerically) by 15/18 participants (83%), binomial $p = .008$. However, only 6/18 adults (33%) fixated the *deebo* less than 50% of the time in consonant-MP trials. This indicates that most adults did not use a mutual-exclusivity strategy to map the word "teebo" onto the distracter object (Markman & Wachtel, 1988; in contrast to Quam & Swingley, 2010).

We also conducted an analogous ANOVA on preference-corrected target fixations, which showed a similar effect of Trial Type, $F(1,17) = 15.00$, $p = .001$ (mean decrease in consonant-MP trials = 33.9%, again shown by 15/18 participants). In consonant-MP trials, only 7/18 adults (39%) fixated the target less during the analysis window than they had during the preview time window, confirming that the majority of adults did not interpret "teebo" as a label for the distracter object.

An additional ANOVA on raw target fixations evaluated the robustness of the effect of Trial Type to differences in the Trained Pitch Contour (rise-fall vs. low fall), which picture was used as the *Deebo* Object ("red knobs" or "purple disk"), or First MP to be presented in the test (consonant or pitch). The inclusion of these additional variables did not meaningfully change the main effect of Trial Type, $F(1,10) = 17.41$, $p = .002$, and there were no significant effects of or interactions with other variables.

In questionnaire responses, 16/18 adults (89%) spontaneously reported noticing the consonant change. The remaining 2 participants remembered it after prompting. In contrast to our prior study using the same method, in which 17/24 adults (71%) reported that they had learned

two words differing only in their vowel, here, only 5/18 participants (28%) reported having learned two words differing only in their consonant, while another 3 (13%) reported some confusion as to whether they had learned one word or two. The remaining 10 participants (56%) only reported learning one word ("deebo").

To summarize, gaze and questionnaire data converged to indicate that English-speaking adults showed robust word learning, and that most adults were affected by the consonant MP in their looking behavior and reported having noticed the consonant change. However, only 33% of adults mapped the word "teebo" onto the distracter object, in contrast to a previous experiment with adults (Quam & Swingley, 2010) in which 75% of adults were reported to do so for a vowel-changed word.

Experiment 2

Experiment 2 tested 24- and 30-month-olds in a similar experimental task, but with two child-friendly modifications to the test phase (described in *Apparatus and Procedure* below). These changes resulted in the experiment lasting less than 10 minutes. Roughly half of children were tested in a low-variability condition similar to that of Experiment 1. For the other children, increased acoustic-phonetic variability (in pitch) was introduced in the training phase. Given prior findings that increased acoustic-phonetic variability in training can aid in minimal-pair differentiation (Rost & McMurray, 2009, 2010) and in the formation of more robust and generalizable word-form categories (Singh, 2008), we predicted that introducing pitch variability in the training phase might lead to more detailed encoding of phonologically relevant dimensions of the target word (Apfelbaum & McMurray, 2011) and therefore better detection of subtle consonant mispronunciations.

As stated above, we initially recruited children at three ages, 19, 24, and 30 months, in the low-variability condition used with adults, to facilitate drawing a continuous developmental picture of consonant interpretation in newly learned words. However, 19-month-olds did not consistently show robust word learning. Only one of two groups of 19-month-olds trained with low variability showed above-chance recognition of the novel word, when correctly pronounced, in test. Thus, only 24- and 30-month-olds were recruited for the high-variability condition, and we report results with just these two ages here. Results from 19-month-olds (including a group tested with pitch mispronunciations) can be found in Supplemental Materials, Experiment S3.

Method

Participants

All caregivers reported that children were learning English as their native and dominant language. Sixty-four children were included in the study. A majority of children had no or negligible exposure to languages other than English. Seven of the sixty-four children (11%), while still strongly dominant in English, had moderate exposure to other languages: Spanish (3), Mandarin (1), Cantonese (1), both Bulgarian and German (1), and both Dutch and Bahasa Indonesian (1). Thirty-two children were included at 24 months: 15 in the low-variability condition (4 female, 11 male) and 17 in the high-variability condition (6 female, 11 male). They were between the ages of 22 months, 24 days and 26 months, 11 days (M = 24 months, 22 days, SD = 28 days). Their mean productive vocabulary was 334 words (SD = 148 words; vocabulary data not collected for 1 participant). Thirty-two children were included at 30 months: 16 in the low-variability condition (6 female, 10 male) and 16 in the high-variability condition (4 female, 12 male). They were between the ages of 28 months, 15 days and 33 months, 24 days (M = 30

months, 13 days, *SD* = 1 month, 5 days). Their mean productive vocabulary was 435 words (*SD* = 201 words; vocabulary data not collected for 1 participant).

Twenty-eight more children participated but were excluded (9 from the 24-month group, 19 from the 30-month group) for fussiness, inattentiveness, or not completing enough usable trials (15), equipment failure or experimenter error (9), parent-reported speech delay (2), age outside of range on the date of testing (1), and parental interference (1). Several additional children were screened from the sample for significant exposure to languages other than English. Trials were only included as usable if the child fixated the pictures for at least 20 frames during the analysis window, out of the 50 total frames between 367-2000 ms. As in Experiment 1, the number of usable trials in each condition was required to be at least half of the total number of trials (at least 4 of 8 trials in each condition).

The number of 30-month-olds excluded due to fussiness, inattentiveness, or having insufficient usable trials was over three times as large in the high-variability condition (n=7) as in the low-variability condition (n=2), while the number did not differ across variability conditions at 24 months (n=3 for each). A higher rate of exclusions due to fussiness in a higher-variability (or otherwise more complex) training condition has also been reported for 14-month-olds in a Switch word-learning task (Quam, Knight, & Gerken, 2017) and for 7.5-month-olds in a sound-discrimination task (Quam, Clough, Knight, & Gerken, 2020).

Apparatus and Procedure

For children in the low-variability condition, the experiment was nearly identical to the one used with adults in Experiment 1 (and the high-variability condition differed only in the auditory stimuli used in the training—see below). Two modifications were implemented to shorten the task

for children. First, each child was tested in only two test-trial conditions (CP and consonant-MP) rather than three, to maximize the number of trials presented in each condition. Second, children saw only 8 filler (familiar-word) trials, instead of the 69 presented to adults. In the test phase, children saw, intermixed, 8 filler (familiar-word) trials, 8 CP trials, and 8 consonant-MP trials.

Questionnaires were not administered to children. In three out of four groups (30-month-olds tested with both low and high variability, and 24-month-olds tested with high variability) children were asked to point to and name objects at the end of the experiment (as in Quam & Swingley, 2010). Where available, pointing and naming data are reported in Supplemental Materials, Experiment S2.

Auditory Stimuli

Auditory stimuli for the test phase were identical to those of Experiment 1 (other than the two modifications described above). The pitch pattern in the test phase (rise-fall or low falling) was counterbalanced across participants. For children in the low-variability condition, auditory stimuli in the training phase were identical to those of Experiment 1.

For the training phase of the high-variability condition, a new set of recordings was produced by the same speaker, in the same recording environment, about four years after recording the original, low-variability recordings from Quam and Swingley (2010). The speaker listened to the original stimuli immediately prior to the recording sessions and imitated the speech rate, mean pitch of the carrier phrases, and other features of the original recordings as closely as possible. The same sentence frames were used as in Experiment 1, but, across the training, the word *deebo* was pronounced with four different intonation contours. Examples of each contour, taken from the ostensive-labeling portion of the training, are depicted in **Figure 3**. Two of these were the rise-

fall and low-falling contours used in Experiment 1 (where each participant was trained with one or the other contour). The other two were a high-falling contour and a rising contour. Each of these contours was presented 5-6 times throughout the training phase. Because rising contours have a fairly restricted intonational meaning in English, typically conveying questions or uncertainty, the rising contour was presented only in felicitous pragmatic contexts (e.g., "I don't want to play with that. A deebo? No way."). **Table A1** reports duration, maximum pitch, and mean pitch of each high-variability training token (refer to rows labeled *Variability: High*; the grand mean for each acoustic measurement across tokens of all four intonation contours is shown in ***bold, italicized*** font in row 7).

[INSERT FIGURE 3 HERE]

Results and Discussion

Target-fixation proportions were calculated over the time window 367-2000 ms after noun onset. The time window typically used with toddlers begins slightly later than the time window typically used with adults, to compensate for children's slower response times (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998; Swingley & Aslin, 2000; Quam & Swingley, 2010). Target-fixation proportions were averaged over all trials with each pronunciation (CP or consonant-MP). We also addressed the possibility that children's picture preferences might influence their target looking by repeating the analyses using preference-corrected fixation proportions, subtracting the target-fixation proportion during the one second prior to noun onset from the target-fixation proportion during the main analysis window (367-2000 ms).

Overall, children recognized the target word quite well when it was correctly pronounced (M = 66.1%, SD = 16.9%)—in fact, not significantly worse than they recognized familiar filler items (M = 71.2%, SD = 9.9%), paired $t(63) = 1.98$, $p = .052$. To determine whether children of

each age had learned the word, we first compared their target fixation to chance (50%) in CP trials, using a two-tailed, one-sample *t* test. **Figure 4** displays *deebo*-fixation proportions in CP and MP trials. For 24-month-olds across both variability conditions, children's *deebo* fixation in CP trials was significantly above chance (M = 65.5%, SD = 16.2%), $t(31) = 5.39$, $p < .001$, as was their *deebo* fixation in consonant-MP trials (M = 69.9%, SD = 14.3%), $t(31) = 7.89$, $p < .001$. Preference-corrected difference scores showed the same patterns, and were significantly above chance (0%) in CP trials (M = 12.8%, SD = 15.3%), $t(31) = 4.70$, $p < .001$, as well as consonant-MP trials (M = 15.8%, SD = 15.4%), $t(31) = 5.80$, $p < .001$.

[INSERT FIGURE 4 HERE]

For 30-month-olds, children's *deebo* fixation in CP trials was significantly above chance (M = 66.8%, SD = 17.7%), $t(31) = 5.36$, $p < .001$, as was their *deebo* fixation in consonant-MP trials (M = 69.3%, SD = 17.2%) $t(31) = 6.35$, $p < .001$. Preference-corrected difference scores showed the same patterns, being significantly above chance (0%) in CP trials (M = 13.2%, SD = 19.3%), $t(31) = 3.87$, $p < .001$, as well as consonant-MP trials (M = 15.6%, SD = 19.2%), $t(31) = 4.58$, $p < .001$.

As 7 of the 64 children in the sample (11%), while strongly dominant in English, had moderate exposure to other languages, we confirmed that these patterns held when these children were temporarily removed from the sample. Across both age groups, children's *deebo* fixation was again significantly above chance in both CP trials (M = 65.6%, SD = 17.5%), $t(56) = 6.73$, $p < .001$, and consonant-MP trials (M = 69.4%, SD = 16.2%), $t(56) = 9.03$, $p < .001$. The 7 children with moderate other-language exposure also showed above-chance target fixation in CP trials (M

= 70.5%, *SD* = 10.4%, *t*(6) = 5.23, *p* = .002) and MP trials (*M* = 71.7%, *SD* = 11.5%, *t*(6) = 4.99, *p* = .002).

A repeated-measures ANOVA with Trial Type (CP, MP) as the within-subjects predictor and between-subjects predictors Variability Condition (low variability, high variability) and Age (24 months, 30 months) revealed no significant effects. Children showed no differences in *deebo* fixation between CP and MP trials, and there were no effects of age or of variability condition (*F*'s < 3, *p*'s > .1) . Only 27/64 children (42%) looked less at the *deebo* object when the consonant of the word was mispronounced than when it was correctly pronounced. Only 7/64 children (11%) fixated the *deebo* less than 50% of the time when the consonant was mispronounced, suggesting children generally did not use a mutual-exclusivity strategy to map the variant word onto the distracter object (Markman & Wachtel, 1988; Quam & Swingley, 2010).

We conducted an analogous ANOVA on preference-corrected target fixations, which also showed no significant effects. The effect of variability condition again did not reach the threshold for statistical significance, *F*(1,60) = 3.20, *p* = .08. There was a numerical trend for higher *overall* preference-corrected target fixations in the high-variability condition (*M* = 17.2%, *SD* = 13.4%) than in the low-variability condition (*M* = 11.3%, *SD* = 12.3%), but this was not modulated by trial type (CP vs. MP). In consonant-MP trials, only 12/64 children (19%) fixated the target less during the analysis window than they had during the preview time window, confirming that, like adults, the majority of children did not interpret "teebo" as a label for the distracter object.

Further analysis of raw target fixations revealed a number of interactions having to do with the pitch contour used in the test phase, or which specific object was the referent of the novel word. None of these revealed conditions under which children exhibited lower target fixation upon hearing a mispronunciation. First, an additional ANOVA checked for potential effects of Pitch in

Test (rise-fall vs. low fall) or *Deebo* Object ("red knobs" or "purple disk"). The main effect of Variability Condition again did not reach the threshold for statistical significance, $F(1,48) = 3.60$, $p = .064$, despite a numerical trend for higher *overall* target fixation in the high-variability condition ($M = 70.5\%$, $SD = 15.1\%$) than the low-variability condition ($M = 65.1\%$, $SD = 12.4\%$), which was not modulated by trial type (CP vs. MP). There was a significant effect of *Deebo* Object, $F(1,48) = 7.67$, $p = .008$, where children taught that the *deebo* was the "red knobs" object showed overall higher target fixation ($M = 71.9\%$, $SD = 11.6\%$) than those taught the "purple disk" ($M = 63.3\%$, $SD = 15.2\%$).

There was a significant three-way interaction of Age by Variability Condition by Pitch in Test, $F(1,48) = 8.63$, $p = .005$. To investigate the interaction, we conducted t tests for each combination of Age and Pitch in Test separately, Bonferroni correcting for the four comparisons. For 30-month-olds tested with the rise-fall contour, there was a significant *overall* advantage (not modulated by trial type) for the high-variability training ($M = 79.9\%$, $SD = 14.0\%$) over the low-variability training ($M = 61.8\%$, $SD = 10.8\%$), $t(14) = 2.88$, $p = .012$ (which met the Bonferroni-corrected p-value threshold of .0125). None of the other three groups differed, all $t < 2$, all $p > .1$.

In sum, in the language-guided looking procedure in which 30-month-olds had previously been shown to robustly learn a word, 24- and 30-month-olds again learned the novel word. However, unlike 30-month-olds who previously attended to vowel changes (Quam & Swingley, 2010), here, 24- and 30-month-olds showed less phonologically constrained responses, showing no evidence of impaired recognition performance given an altered consonant. Only 42% of children looked less at the *deebo* object in response to the consonant MP, in contrast to 83% reported to do so in response to a vowel MP (Quam & Swingley, 2010). Only 11% of children

seemed to interpret the variant word form as an entirely new word, compared with 46% reported to do so when the word varied in its vowel (Quam & Swingley, 2010).

General Discussion

Learning the phonology of a language requires developing intuitions about how to handle phonetic variation. A word realized in a phonetically deviant manner that nonetheless respects the word's phonological requirements should give rise to a different set of hypotheses than a word realized in a deviant manner that fails to meet that word's phonological commitments. Here, we investigated toddlers' and adults' interpretations of phonological variation via a teaching procedure incorporating 22 presentations of a novel word across a simple story and ostensive labeling. Phonetically, the word was always realized in a hyperarticulated way, usually with prominent prosodic highlighting. The word was produced with either low prosodic variability or, for roughly half of children, high variability.

Recognition of the word was tested immediately after training. Toddlers learned the novel word robustly. However, their recognition of the novel word was not measurably impaired by a change to the initial consonant's voicing, whether the word was taught with high or low intonational variability. Evidence for developmental change from 24 to 30 months was scant. Only adults learned the word robustly *and* showed phonologically constrained responses, treating consonant changes as relevant. Only a third of adults treated the word form with the deviant consonant as a novel word, even though an unnamed novel object was available as a potential referent. This response was still rarer for two-year-olds.

While this study represents just one point in a space of training situations (characterized by intensive, short-term exposure and immediate test), it may nevertheless provide information about

the conditions under which toddlers can apply their phonological knowledge to novel-word learning. Teaching and testing materials were presented in a stereotypically infant-directed, hyperarticulated style. Toddlers in many studies have shown decrements in picture fixation when familiar words were mispronounced in this register (sometimes with this very same change, from /d/ to /t/). Considering these prior findings, it is unlikely that the absence of an effect of the voicing change here indicates a failure of immediate perception, in training or test, of the phonological distinction itself, but, rather, the challenge of applying the distinction at the word level.

These results differ from the findings of Quam and Swingley (2010), who tested children and adults using teaching stimuli identical to those used in the present experiment. In that study, adults were significantly more inclined to interpret a vocalic change in pronunciation as a novel word (18/24; 75%) than for the consonant change here (6/18; 33%), Chi-sq. 5.69, df=1, $p = 0.017$. Children also showed less sensitivity to the consonant change. In Quam and Swingley (2010), 20/24 children (83%) responded to the vowel mispronunciation, fixating the taught object less upon hearing *dahbo* than *deebo*, while 11/24 children (46%) showed a potential mutual-exclusivity response, fixating the taught object less than 50% upon hearing the vocalic change. Here, the analogous proportions were only 42% fixating the taught object less in response to the consonant mispronunciation and 11% showing a potential mutual-exclusivity response.

Continued orientation to the familiar object in the face of a subtle mispronunciation is consistent with prior findings with children (Swingley & Aslin, 2000, 2002; White & Morgan, 2008). Still, the contrast between the present results and the greater sensitivity to a vowel mispronunciation (Quam & Swingley, 2010) is perhaps surprising. Across the two studies, the participants were sampled from the same population and the materials and procedures were the same except for the nature of the mispronunciation. Thus, the contrast between the two studies

presents a counterexample to the more typical result, found in studies of children under two years, in which consonants are treated as more significant in determining lexical identity than vowels are (Nazzi & Cutler, 2019; though the stronger role for consonants is not as consistently found in English-learning toddlers as in adults). It is possible that this is due to the fact that the spoken words were substantially hyperarticulated, with long, drawn-out vowels. It is easier to emphasize a vowel in this way than to emphasize a consonant. On the other hand, prior studies that have tested both vowel and consonant alterations have not found a difference (Swingley & Aslin, 2000; see von Holzen & Bergmann, 2021, for a review).

Another relevant factor may be that the consonant mispronunciation involved a change to only one feature (voicing), whereas the vowel mispronunciation changed two features (tongue height and frontness). Although we would not expect phonetic feature counts to predict interpretation exactly, *a priori* one would consider [d] to [t] to be a more minimal phonological change than [i] to [a]. As noted above, prior research confronting children with a familiar object and a novel object and mispronouncing the name of the familiar object has revealed larger decrements to target fixation for more extreme phonological deviations. We may therefore be observing additive effects of using a relatively subtle phonological change (relative to the [i–a] contrast), and testing a novel word (rather than the familiar words tested more commonly).

The insensitivity toddlers showed to a one-feature consonant mispronunciation in the present study must also be reconciled with prior findings of sensitivity to similar consonant mispronunciations in Dutch-learning 19-month-olds (Swingley, 2007), as well as a trend in English-learning 14-month-olds (Ballem & Plunkett, 2005). In the Swingley (2007) study, children were taught a word (*tiebie*, /tibi/, or *droekel*, /drukəl/) and tested on small mispronunciations (/kibi/, /trukəl/) or larger ones (e.g., /kribi/). Children who had heard the word pre-exposed prior

to its being mapped to meaning were sensitive to both large and small mispronunciations, whereas children not given the pre-exposure only detected large mispronunciations. The total number of exposures to the target word was equivalent in the present study (22 total: 10 in the story, 12 in the labeling phase) and in the preexposure condition of the prior study (also 22 total: 14 in the story, 8 in the labeling phase) where 19-month-olds successfully detected comparable mispronunciations. Swingley (2007) did not pair word forms with visual referents during the story phase, instead waiting until the labeling phase to do so. It is possible that preexposure to the word form, before the introduction of meaning, reduced the task difficulty and allowed children to focus on the sounds of the word and encode them in more detail (though an opposite prediction could potentially have been made, given evidence that pairing word forms with objects can help infants differentiate minimal pairs; Yeung & Werker, 2009).

Ballem and Plunkett (2005) used a substantially simpler training method than the one used here, and found that 14-month-olds in the second of two training blocks, but not the first, learned words, performing above chance in fixating the named target when it was pronounced as it had been trained. In that second block, children did not perform above chance upon hearing a mispronunciation, although the difference between CP and MP performance was not itself statistically significant. Differences in task complexity and in number of exposures might account for the discrepancy in results. Our novel distracter object was frequently presented in the training phase (but never labeled), and it is possible that the inclusion of this second object during training also increased the task difficulty.

Our results with adults have a parallel in a study by White, Yee, Blumstein, and Morgan (2013), who also included single-feature voicing mispronunciations of newly learned words (in Experiment 1). White et al. found adults' sensitivity to mispronunciations was modulated by both

the number of exposures to words and the number of features mispronounced (1 vs. 2). The effect of number of features is similar to our finding that, while all adults detected mispronunciations, their likelihood of mapping the mispronounced form onto the distracter was lower for a one-feature consonant mispronunciation than for a two-feature vowel mispronunciation (Quam & Swingley, 2010). White et al. argued that, while adults have mature knowledge of the phonological content of words, the application of this knowledge during recognition of newly learned words can be obscured by competition between similar-sounding words (see also Magnuson, Tanenhaus, Aslin, & Dahan, 2003).

Prior work has indicated that consistent prosodic content, as in our low-variability condition, can mask infants' detection of consonant changes (Singh, 2008), and toddlers' ability to differentiate consonant-differentiated minimal pairs (Rost & McMurray, 2009). Nevertheless, we found only minimal effects of incorporating intonational variability into training stimuli. The introduction of variability marginally increased *overall* looking times, but it did not result in better detection of consonant mispronunciations. While we found null effects of variability and of its interaction with trial type, this does not necessarily mean variability has no impact on encoding of details of novel words at these ages. Given findings that incorporating acoustic-phonetic variability into familiarization aids word recognition at 7.5 months (e.g., Singh, 2008), and that 14-month-olds differentiate similar-sounding words better when habituated with acoustic-phonetic variability (e.g., Rost & McMurray, 2009), we anticipated that we might find more robust learning with greater variability. However, not all prior studies have shown facilitation from training variability (see Quam & Creel, 2021, for an overview). For example, Quam and Swingley (2021, in prep.) found that 18-month-olds' word learning in the Switch task was not affected by the

introduction of irrelevant acoustic-phonetic variability (vowel for pitch-contrasted words, or pitch for vowel-contrasted words).

Models that predict (or are consistent with) facilitation from acoustic-phonetic variability, such as PRIMIR (Werker & Curtin, 2005), WRAPSA (Jusczyk, 1993), and Apfelbaum and McMurray's (2011) associative model, all conceptualize infants and younger toddlers (e.g., 14-month-olds) as relatively more unconstrained by native phonology than our 24- and 30-month-olds. In WRAPSA and in Apfelbaum and McMurray's associative model, increasing experience with the native language leads to heavier weighting of contrastive dimensions, while in PRIMIR, by 17 months, children are argued to process words phonemically. Thus, it could be that by 24 and 30 months, children are less likely to benefit from facilitation from acoustic-phonetic variability, though this is not to say that such effects ever disappear entirely, as they sometimes appear in adult native speakers (e.g., Barcroft & Sommers, 2005).

An essential skill for word learning is the ability to recognize a word across changes in the speaker's voice, the intonation pattern, duration, sentence position, and even mildly deviant pronunciations, if they are caused by inadvertent misspeakings or dialect differences (see Quam & Creel, 2021, for discussion). The language-guided looking method we used here is sensitive enough to detect a hindrance in word recognition when the spoken word fails to match the listener's phonological representation, even when the word is interpreted as "close enough" to indicate the familiar lexical item (Ramon-Casas et al., 2009; Swingley, 2016; Swingley & Aslin, 2000; White & Morgan, 2008). The present results indicate that well into the second year, children do not always respond to phonologically relevant changes in newly learned words. Children's developing vocabularies are composed of some words with which children have massive long-term experience, some words just barely making their way into the vocabulary, and many words in

between. Studies of children's "best" words, and of words children have just been taught (possibly their "worst" words), show a range in the quality of children's phonological representations, indicated by the reliability with which children detect phonologically relevant mispronunciations. A challenge for future work is to develop a means for evaluating children's knowledge of the words in the middle.

Acknowledgments

We are tremendously grateful to the parents, children, and adult participants who participated in this study. We thank members of the Infant Language Center at the University of Pennsylvania who assisted with tasks such as participant scheduling and testing, including Sara Clopton, Jane Park, Alba Tuninetti, Kristin Vindler Michaelson, and Rebecca McCue, or manuscript preparation, including Sophia Heiser and Anna Runova. Additional students from the Child Language Learning Center at Portland State University assisted with manuscript preparation, including Genesis Ocegueda Enciso, Josie Johnson, Katharine Ross, and Helena Sai. Funding was provided by NSF Graduate Research Fellowship and NSF IGERT Trainee Fellowship grants to C.Q., the National Institute of General Medical Sciences of the National Institutes of Health Award Number RL5GM118963 (which supported student research assistants working with C.Q.), NSF grant HSD-0433567 to Delphine Dahan and D.S., and NIH grant R01-HD049681 and NSF grant 1917608 to D.S. Research reported in this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

Declaration of Interest Statement

The authors have no relevant financial or non-financial competing interests to report.

Data Availability Statement

The data that support the findings of this study are available for download at https://doi.org/10.15760/sphr-data.01 (DOI: 10.15760/sphr-data.01).

References

Antoniou, M., & Wong, P. C. M. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *The Journal of the Acoustical Society of America, 139*, 271–278.

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*(6), 1105–1138. https://doi.org/10.1111/j.1551-6709.2011.01181.x

Bailey, T. M. & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development 17*, 1265–82.

Ballem, K. D., & Plunkett, K. (2005). Phonological specificity in children at 1;2. *Journal of Child Language, 32*(1), 159-173.

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition, 27,* 387–414.

Bergelson, E., & Swingley, D. (2018). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*, *89*(5), 1567–1576.

Bernier, D. E., & White, K. S. (2019). Toddlers' sensitivity to phonetic detail in child speech. *Journal of Experimental Child Psychology, 185,* 128-147.

Bosch, L., & Sebastián-Gallés, N. (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. *Language and Speech*, *46*(2–3), 217–243. https://doi.org/10.1177/00238309030460020801

Creel, S. C., & Quam, C. (2015). Apples and oranges: Developmental discontinuities in spoken-language processing? *Trends in Cognitive Sciences*, *19*(12), 713–716. https://doi.org/10.1016/j.tics.2015.09.006

Dautriche, I., Fibla, L., Fievet, A.-C., Christophe, A. (2018). Learning homophones in context: Easy cases are favored in the lexicon of natural languages. *Cognitive Psychology, 104,* 83-105.

Dautriche, I., Swingley, D., & Christophe, A. (2015). Learning novel phonological neighbors: Syntactic category matters. *Cognition*, *143*, 77–86.

Dietrich, C., Swingley, D., & Werker, J. F. (2007). Native language governs interpretation of salient speech sound differences at 18 months. *Proceedings of the National Academy of Sciences*, *104*(41), 16027–16031.

Fennell, C. T. (2012). Object familiarity enhances infants' use of phonetic detail in novel words. *Infancy*, *17*(3), 339–353.

Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, *81*(5), 1376–1383. https://doi.org/10.1111/j.1467-8624.2010.01479.x

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, *9*(3), 228–231.

Galle, M. E., Apfelbaum, K. S., & McMurray, B. (2015). The role of single talker acoustic variation in early word learning. *Language Learning and Development*, *11*(1), 66–79.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*(2), 251-279.

Hallé, P. A., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior & Development*, *19*(4), 463–481.

Höhle, B., Fritzsche, T., Meß, K., Philipp, M., & Gafos, A. (2020). Only the right noise? Effects of phonetic and visual input variability on 14-month-olds' minimal pair word learning. *Developmental Science, 23*(5), e12950.

Hollich, G. (2005). *Supercoder: A program for coding preferential looking (Version 1.5) [computer software]*. West Lafayette: Purdue University.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(5), 1570–1582. https://doi.org/10.1037/0096-1523.26.5.1570

Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6), 1143–1154. https://doi.org/10.1037/0096-1523.29.6.1143

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. *Journal of Experimental Psychology: General, 132*(2), 202–227. https://doi.org/10.1037/0096-3445.132.2.202

Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, *57*(2), 252–272. https://doi.org/10.1016/j.jml.2007.03.005

Mani, N., & Plunkett, K. (2008). Fourteen-month-olds pay attention to vowels in novel words. *Developmental Science, 11*(1), 53–59.

Mani, N., & Plunkett, K. (2011). Does size matter? Subsegmental cues to vowel mispronunciation detection. *Journal of Child Language, 38,* 606-627.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.

Nazzi, T., & Cutler, A. (2019). How consonants and vowels shape spoken-language recognition. *Annual Review of Linguistics*, *5,* 25-47. https://doi.org/10.1146/annurev-linguistics-011718-011919

Nazzi, T., & New, B. (2007). Beyond stop consonants: Consonantal specificity in early lexical acquisition. *Cognitive Development, 22*(2), 271-279.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America, 130*, 461-472.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(2), 421-435.

Quam, C., Clough, L., Knight, S., & Gerken, L. (2020). Infants' discrimination of consonant contrasts in the presence and absence of talker variability. *Infancy*, *26*(1), 84–103.

Quam, C., & Creel, S. C. (2021). Impacts of acoustic-phonetic variability on perceptual development for spoken language: A review. *WIREs Cognitive Science,* e1558.

Quam, C., Knight, S., & Gerken, L. (2017). The distribution of talker variability impacts infants' word learning. *Laboratory Phonology*, *8*(1), 1-17. https://doi.org/10.5334/labphon.25

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. https://doi.org/10.1016/j.jml.2009.09.003

Quam, C., & Swingley, D. (2021, March). *English-learning children's processing of salient phonetic distinctions varying in phonological relevance for word identity* [Oral presentation]. 34th Annual CUNY Conference on Human Sentence Processing.

Quam, C., & Swingley, D. (in preparation). Developmental change in English-learning children's interpretations of salient pitch contours in word learning.

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.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*(2), 339–349. https://doi.org/10.1111/j.1467-7687.2008.00786.x

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*(6), 608–635. https://doi.org/10.1111/j.1532-7078.2010.00033.x

Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America, 134,* 1324-1335.

Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology, 5*, 1318.

Segal, O., Keren-Portnoy, T., & Vihman, M. (2020). Robust effects of stress on early lexical representation. *Infancy*, *25*(4), 500–521.

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*(2), 833–870. https://doi.org/10.1016/j.cognition.2007.05.002

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, *51*(2), 173–189. https://doi.org/10.1016/j.jml.2004.04.004

Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development, 4*, 157-178.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*(6640), 381–382.

Swingley, D., (2003). Phonetic detail in the developing lexicon. *Language and Speech, 46*(2-3), 265-294.

Swingley, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, *8*(5), 432–443.

Swingley, D. (2007). Lexical exposure and word-form encoding in 1.5-year-olds. *Developmental Psychology*, *43*(2), 454–464. https://doi.org/10.1037/0012-1649.43.2.454

Swingley, D. (2009). Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3617–3632. https://doi.org/10.1098/rstb.2009.0107

Swingley, D. (2016). Two-year-olds interpret novel phonological neighbors as familiar words. *Developmental Psychology*, *52*(7), 1011–1023.

Swingley, D. (2022). Infants' learning of speech sounds and word forms. *Papafragou, A.; Trueswell, J; and Gleitman, l. (Eds.), Oxford Handbook of the Mental Lexicon, Oxford*, 24.

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*(2), 147–166. 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.

Swingley, D., & Aslin, R. N. (2007). Lexical competition in young children's word learning. *Cognitive Psychology*, *54*(2), 99–132.

Tamási, K., McKean, C., Gafos, A., & Höhle, B. (2019). Children's gradient sensitivity to phonological mismatch: Considering the dynamics of looking behavior and pupil dilation. *Journal of Child Language, 46*(1), 1-23.

Tamási, K., McKean, C., Gafos, A., Fritzsche, T., & Höhle, B. (2017). Pupillometry registers toddlers' sensitivity to degrees of mispronunciation. *Journal of Experimental Child Psychology, 153,* 140-148.

Thiessen, E. D. (2007). The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*, *56*(1), 16–34.

Von Holzen, K., & Bergmann, C. (2021). The development of infants' responses to mispronunciations: A meta-analysis. *Developmental Psychology, 57*(1), 1-18.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A Developmental Framework of Infant Speech Processing. *Language Learning and Development*, *1*(2), 197–234. 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. https://doi.org/10.1207/S15327078IN0301_1

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

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

White, K. S., Yee, E., Blumstein, S. E., & Morgan, J. L. (2013). Adults show less sensitivity to phonetic detail in unfamiliar words, too. *Journal of Memory and Language*, *68*(4), 362–378. https://doi.org/10.1016/j.jml.2013.01.003

Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition*, *113*(2), 234–243. https://doi.org/10.1016/j.cognition.2009.08.010

Yeung, H. H., Chen, L. M., & Werker, J. F. (2013). Referential labeling can facilitate phonetic learning in infancy. *Child Development, 85*(3), 1036-1049.

Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science*, *12*(3), 412–418.

Zesiger, P., Lozeron, E. D., Lévy, A., & Frauenfelder, U. H. (2012). Phonological specificity in 12- and 17-month-old French-speaking infants. *Infancy, 17*(6), 591-609.

Appendix

Table A1: Acoustics of the Teaching and Test Words. Means (and standard deviations) for duration in seconds, pitch maximum (max) in Hz, and pitch mean in Hz, are given for word tokens with each pitch contour from the low-variability and high-variability teaching conditions and the test phase (always low variability). Row 7 (in ***bold and italics***) reports the grand mean across all high-variability teaching tokens.

Variability	Phase	Word	Pitch	Duration (SD)	Pitch max (SD)	Pitch mean (SD)
Low	Teaching	Deebo	Rise-fall	1.245 (0.076)	587.7 (56.2)	284.8 (15.5)
Low	Teaching	Deebo	Low fall	1.370 (0.121)	264.1 (11.7)	215.1 (6.8)
High	Teaching	Deebo	Rise-fall	1.257 (0.133)	601.0 (51.8)	273.0 (15.0)
High	Teaching	Deebo	Low fall	1.358 (0.149)	261.4 (16.0)	210.9 (10.9)
High	Teaching	Deebo	High fall	1.376 (0.103)	676.0 (32.7)	381.0 (18.5)
High	Teaching	Deebo	Rising	1.271 (0.039)	458.4 (18.3)	289.6 (15.3)
High	***Teaching***	***Deebo***	***Variable***	***1.318 (0.120)***	***501.3 (169.0)***	***288.6 (65.6)***
Low	Test	Deebo	Rise-fall	1.321 (0.038)	673.4 (26.3)	300.1 (2.7)
Low	Test	Deebo	Low fall	1.292 (0.077)	283.9 (2.9)	232.7 (9.1)
Low	Test	Teebo	Rise-fall	1.284 (0.048)	647.9 (47.0)	294.7 (12.8)
Low	Test	Teebo	Low fall	1.379 (0.032)	435.2 (24.0)	237.6 (2.0)

Tables with Captions

Table 1: Mean Target-fixation Proportions (with Standard Deviations) in CP and Consonant-MP Trials. Included are 24-month-olds, 30-month-olds, and adults (with the grand mean for children overall in row 3 ***in bold, italicized font***). The rightmost 2 columns list the percentage of participants looking less to the *deebo* in MP trials than CP trials (showing an MP effect) and the percentage looking less than 50% of the time in MP trials (using a mutual exclusivity, ME, strategy).

	Correct pronunciation	Consonant MP	% Showing MP Effect	% Using ME Strategy
24 months	65.5% (16.2%)	69.9% (14.3%)	43.8% (14/32)	6.3% (2/32)
30 months	66.8% (17.7%)	69.3% (17.2%)	40.6% (13/32)	15.6% (5/32)
Children overall	***66.1% (16.9%)***	***69.6% (15.7%)***	***42.2% (27/64)***	***10.9% (7/64)***
Adults	91.4% (9.8%)	61.8% (32.1%)	83.3% (15/18)	33.3% (6/18)

Figure Captions

Figure 1: Waveforms and Spectrograms with Overlaid Pitch Tracks for the Consonant-Mispronunciation Test Sentences. The sentence depicted is "Where's the teebo?" with a rise-fall contour (A) and low-fall contour (B). Vertical lines depict word boundaries.

Figure 2: Adults' Fixation of the *Deebo* Object in Each Trial Type. The horizontal line indicates chance fixation, or 50%. Adults' fixation of the *deebo* object was impacted by the consonant mispronunciation ("MP_consonant"), indicated by *deebo* looking proportions that were not significantly above chance. Fixations were not impacted by the pitch-contour mispronunciation ("MP_contour"). Box plots indicate within-subject difference scores between CP and MP trials for each MP type.

Figure 3: Waveforms and Spectrograms with Overlaid Pitch Tracks for the Four Intonation Contours Used in the High-Variability Training. All training sentences were pronounced with the correct consonant. The sentences depicted are "Look at the deebo" with a rise-fall contour (A), low-fall contour (B), and high-fall contour (C); and "See that? The deebo?" with a rising contour (D). Vertical lines depict word boundaries.

Figure 4: Children's Fixation of the *Deebo* Object in Each Trial Type and Variability Condition. Top: 24-month-old participants' fixation of the target object (the *deebo*) in response to the correct pronunciation ("CP") and the consonant mispronunciation ("MP"), after high-variability training (left) or low-variability training (right). Bottom: 30-month-olds. The horizontal line indicates chance fixation, or 50%. Box plots indicate within-subject difference scores between CP and MP trials.

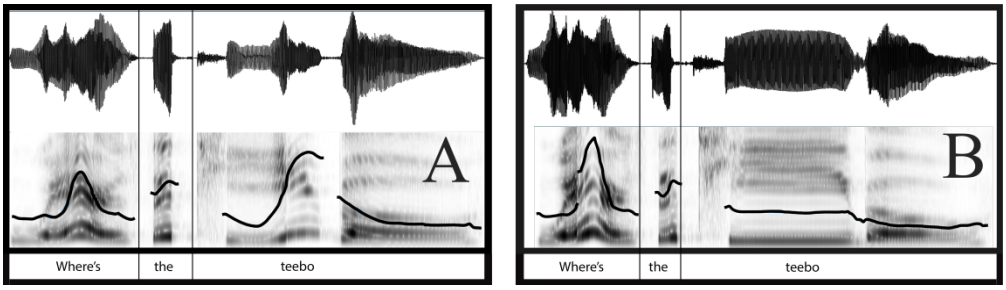

Figure 1: Waveforms and Spectrograms with Overlaid Pitch Tracks for the Consonant-Mispronunciation Test Sentences. The sentence depicted is "Where's the teebo?" with a rise-fall contour (A) and low-fall contour (B). Vertical lines depict word boundaries.

609x171mm (300 x 300 DPI)

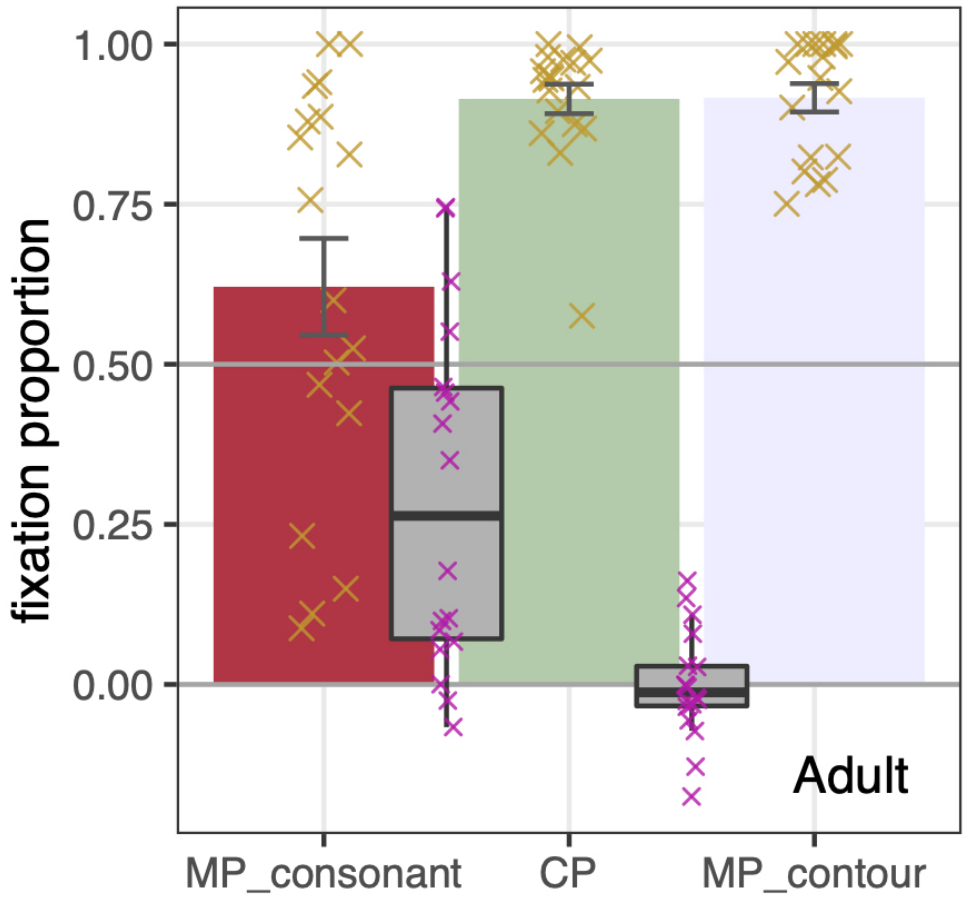

Figure 2: Adults' Fixation of the Deebo Object in Each Trial Type. The horizontal line indicates chance fixation, or 50%. Adults' fixation of the deebo object was impacted by the consonant mispronunciation ("MP_consonant"), indicated by deebo looking proportions that were not significantly above chance. Fixations were not impacted by the pitch-contour mispronunciation ("MP_contour"). Box plots indicate within-subject difference scores between CP and MP trials for each MP type.

76x76mm (300 x 300 DPI)

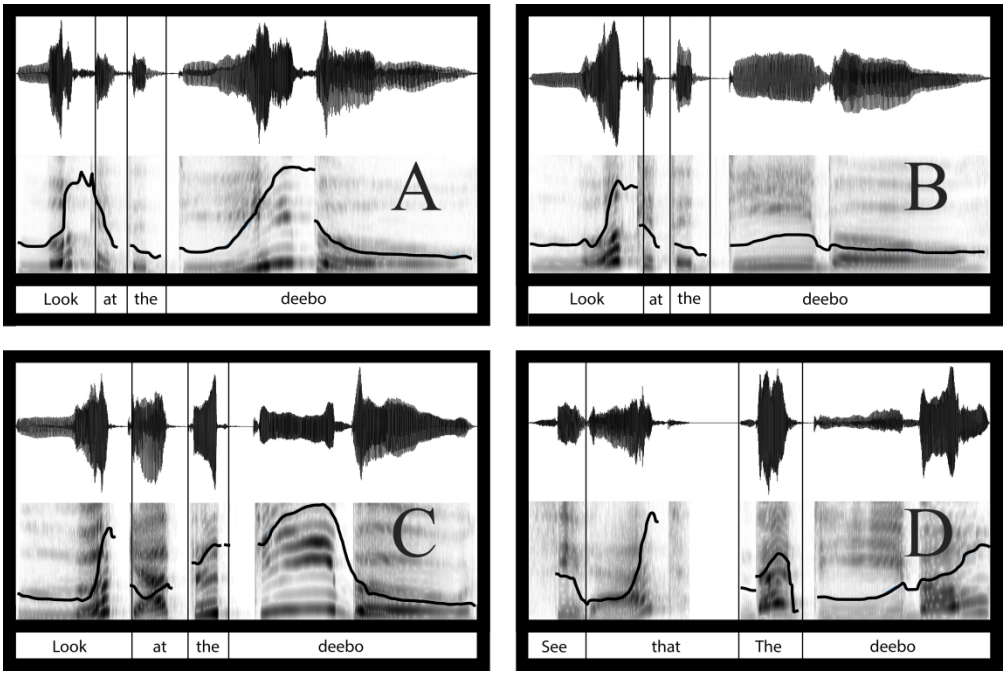

Figure 3: Waveforms and Spectrograms with Overlaid Pitch Tracks for the Four Intonation Contours Used in the High-Variability Training. All training sentences were pronounced with the correct consonant. The sentences depicted are "Look at the deebo" with a rise-fall contour (A), low-fall contour (B), and high-fall contour (C); and "See that? The deebo?" with a rising contour (D). Vertical lines depict word boundaries.

495x330mm (300 x 300 DPI)

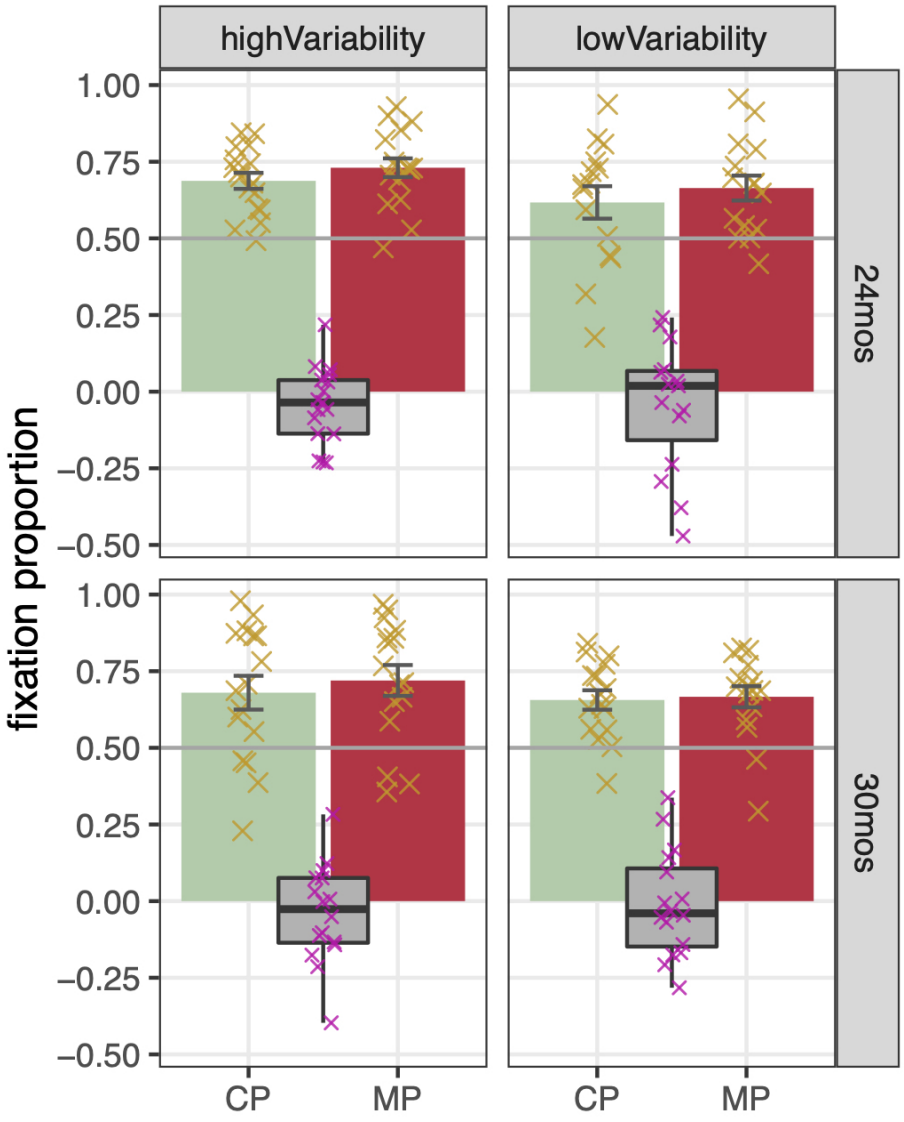

Figure 4: Children's Fixation of the Deebo Object in Each Trial Type and Variability Condition. Top: 24-month-old participants' fixation of the target object (the deebo) in response to the correct pronunciation ("CP") and the consonant mispronunciation ("MP"), after high-variability training (left) or low-variability training (right). Bottom: 30-month-olds. The horizontal line indicates chance fixation, or 50%. Box plots indicate within-subject difference scores between CP and MP trials.

88x114mm (300 x 300 DPI)

SUPPLEMENTAL MATERIALS

Experiment S1

Results and Discussion

Responses to Pitch Mispronunciations

The same 18 adults whose responses to consonant mispronunciations are reported in Experiment 1 of the main text also responded to pitch-contour mispronunciations (MPs). Here, we report means (see **Table S1**) and statistical analyses for target fixations in response to pitch MPs, comparing the results to a very similar published study (Quam & Swingley, 2010). We first evaluated whether the pitch MP significantly affected adults' fixation of the *deebo*. In response to the pitch-contour change, adults' *deebo* fixation remained significantly above chance ($M = 91.6\%$, $SD = 9.4\%$), $t(17) = 18.82$, $p < .001$ (see **Figure 2**, main text). We next conducted a repeated-measures ANOVA with Trial Type (correct pronunciation—CP—pitch MP, and consonant MP) as the within-subjects predictor, which revealed a significant effect of Trial Type, $F(2,34) = 20.56$, $p < .001$. Planned comparisons to investigate differences between trial types revealed that adults did not look less at the *deebo* object in response to the pitch MP than the CP (*mean difference* = -0.2%, n.s.). Only 6/18 participants (33%) looked less at the *deebo* in response to the pitch MP (compared with 12/24, 50%, reported by Quam & Swingley, 2010). Adults looked significantly less at the *deebo* in the consonant-MP condition than in the pitch-MP condition (*mean decrease* = 29.9%), $t(17) = 4.62$, $p < .001$. No adults fixated the *deebo* less than 50% of the time in pitch MP trials (as reported by Quam & Swingley, 2010).

Table S1: Mean Target-Fixation Proportions (with Standard Deviations) in CP, Pitch-MP, and Consonant-MP Trials. Included are adults, for pitch MPs (consonant MPs are reported in Table 1 in the main text), and 19-month-olds, tested between subjects with pitch or consonant MPs. The right-most 2 columns list the percentage of participants looking less to the *deebo* in MP trials than CP trials (showing an MP effect) and the percentage looking less than 50% of the time in MP trials (using a mutual exclusivity, ME, strategy).

	Correct pronunciation	MP	% Showing MP Effect	% Using ME Strategy
Adults-Pitch MP	91.4% (9.8%)	91.6% (9.4%)	33.3% (6/18)	0.0% (0/18)
19 mo.-Pitch MP	66.7% (17.1%)	66.7% (20.6%)	31.6% (6/19)	26.3% (5/19)
19 mo.-Consonant MP	56.9% (23.2%)	62.6% (21.5%)	36.8% (7/19)	21.1% (4/19)

An additional ANOVA evaluated the robustness of the effect of Trial Type to differences in the Trained Pitch Contour (rise-fall vs. low fall), which picture was used as the *Deebo* Object ("red knobs" or "purple disk"), or First MP to be presented in the test (consonant or pitch). The inclusion of these additional variables did not meaningfully change the main effect of Trial Type, $F(2,20) = 17.09$, $p < .001$, and there were no significant effects of or interactions with other variables.

In the questionnaire, 8/18 adults (44%) spontaneously reported noticing the pitch change (12/24, 50%, did so in Quam & Swingley, 2010), compared with 89% who reported noticing the consonant change. Eight more participants (44%) remembered the pitch change after prompting, but two participants (11%) had no memory of the pitch change. As found previously (Quam & Swingley, 2010), no participants reported having learned two words differing only in their pitch pattern (compared with 28% who reported having learned two words differing only in their consonant).

Experiment S2

Results

At the end of the experiment, 30-month-old children were asked to point in response to each pronunciation ("Point to the [deebo/teebo]") and to label each of the two pictures ("Tell Elmo what that is!"). Twenty-four-month-olds tested in the high-variability condition also completed pointing and naming trials (those tested in the low-variability condition were tested prior to the addition of this latter portion of the experiment). Children's pointing and naming responses were designed to provide another lens on their interpretations of pitch and vowel mispronunciations.

Pointing Data

Only children who pointed in both pointing trials were included in the analysis (Quam & Swingley, 2010). Across the low- and high-variability conditions, 14 of 32 30-month-olds (44%) responded in both pointing trials. When asked to "Point to the deebo," 10/14 children (71%) pointed to the *deebo* object; 2 (14%) pointed to the distracter object, and 2 (14%) responded ambiguously. When asked to "Point to the teebo," again, 10/14 children (71%) pointed to the *deebo* object; 2 (14%) pointed to the distracter object, and 2 (14%) responded ambiguously. Thus, responses to both pronunciations were comparable.

At 24 months, only children in the high-variability condition provided pointing and naming responses. Only two of seventeen children (12%) responded to both pointing trials at 24 months. When asked to "Point to the deebo," one pointed to the *deebo* object, and one pointed to the distracter object. When asked to "Point to the teebo," again, one (the same one) pointed to the *deebo* object, and one pointed to the distracter object.

Naming Data

We scored productions for whether the onset of the first syllable was /d/ or /t/ (Quam & Swingley, 2010). When asked to label the *deebo* object, 30-month-olds across both variability conditions produced more /d/ consonants (11) than /t/ consonants (2). (Note that one child who we coded as producing /d/ first said "teebu" with a pacifier impeding their articulation and then removed the pacifier and said "deebo," which we assumed was the intended pronunciation.) When asked to label the distracter object, children were more reluctant to produce a label: only 2 children labeled the distracter, and both used /t/ (e.g., "A teebo. Is that a teebo?"). Other children implied the object did not have a label (e.g., "I dunno"; "It's a toy that monkey put there for Elmo to play with"; "I wanna play with that toy") or gave it their own label ("stop sign").

At 24 months, only children in the high-variability condition were prompted to provide naming responses. As with pointing, few 24-month-olds produced interpretable names for the objects. When asked to label the *deebo* object, children actually produced more /t/ onsets (2) than /d/ onsets (0). This was also the case for the distracter object, where one child produced a /t/ onset and no children produced a /d/.

Experiment S3

We tested 19-month-olds in the low-variability condition of the same experiment used in Experiment 2 of the main text. However, half of 19-month-olds were tested with consonant mispronunciations, and the other half with pitch mispronunciations (as used with adults, analyzed in Experiment S1, above; and Quam & Swingley, 2010).

Method

Participants

Nineteen children (10 female and 9 male) were included in the pitch-MP condition. They were between the ages of 17 months, 27 days and 20 months, 18 days (mean age 19 months, 1 day, *SD* = 19 days; mean productive vocabulary 136 words; vocabulary data not collected for 1 participant). Nineteen children (9 female and 10 male) were included in the consonant-MP condition. They were between the ages of 17 months, 24 days and 20 months, 21 days (*M* = 19 months, 9 days, *SD* = 27 days; mean productive vocabulary 140 words; vocabulary data not collected for 2 participants). Twenty more children participated but were excluded for fussiness or not having a sufficient number of trials. Additional children were screened from the sample for significant exposure to languages other than English. Children were required to have at least 4 usable trials in each of the trial types. Trials were only included if the child fixated the pictures for at least 20 frames during the analysis window, out of the 50 total frames between 367-2000 ms.

Results and Discussion

We first evaluated whether children learned the word and whether either MP significantly affected their fixation of the *deebo*. **Figure S1** displays participants' responses in each trial type. For children in the consonant-MP condition, target fixation in CP trials was not significantly different from chance (*M* = 56.9%, *SD* = 23.2%; means are summarized in **Table S1**), $t(18) = 1.29$, $p = .214$. However, children's *deebo* fixation in consonant-MP trials was significantly above chance (*M* = 62.6%, *SD* = 21.5%), $t(18) = 2.55$, $p = .020$. For the pitch-MP group, children's *deebo* fixation in CP trials was significantly above chance (*M* = 66.7%, *SD* = 17.1%), $t(18) = 4.25$; $p <$

.001. Children's *deebo* fixation was also significantly above chance in pitch-MP trials ($M = 66.7\%$, $SD = 20.6\%$), $t(18) = 3.53$, $p = .001$.

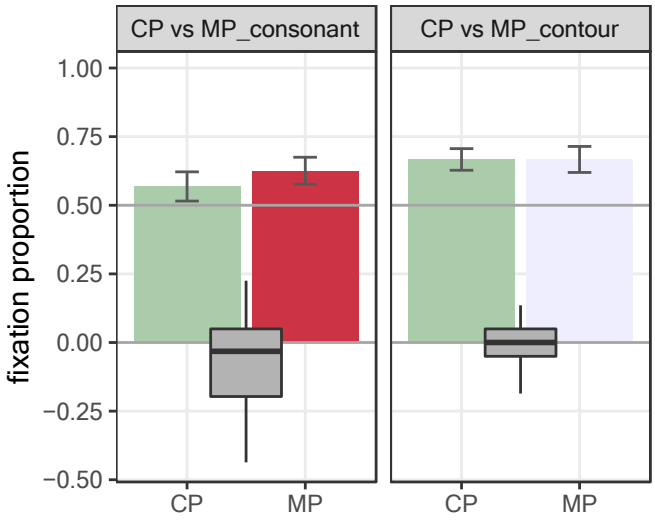

Figure S1: Nineteen-Month-Old Children's Fixation of the *Deebo* Object in Each Trial Type. Left: participants tested in the consonant-mispronunciation ("MP_consonant") condition. Right: participants tested in the pitch-contour-mispronunciation ("MP_contour") condition.

A repeated-measures ANOVA with Trial Type (CP or MP) as the within-subjects predictor and MP Type (pitch MP vs. consonant MP) as a between-subjects factor revealed no significant effects. Children showed no differences in *deebo* fixation between CP and MP trials and there were no differences between mispronunciation types. Only 6/19 in the pitch-MP condition (32%) and 7/19 (37%) in the consonant-MP condition looked less at the *deebo* object when the word was mispronounced than when it was correctly pronounced. Not surprisingly, only 5/19 children in the pitch-MP condition (26%) and 4/19 in the consonant-MP condition (21%) fixated the *deebo* less than 50% of the time when the word was mispronounced, suggesting children generally did not use a mutual-exclusivity strategy to map either variant onto the distracter object (Markman & Wachtel, 1988; Quam & Swingley, 2010).

An additional ANOVA checked for potential effects of Trained Pitch Contour (rise-fall vs. low fall) or *Deebo* Object ("red knobs" or "purple disk"). This analysis did not meaningfully change the variables of primary interest, but there was a significant main effect of *Deebo* Object, $F(1,30) = 4.58$, $p = .041$, indicating higher overall target fixations by children for whom the "red knobs" object was the target ($M = 72.1\%$, $SD = 11.6\%$) than by those for whom the "purple disk" object was the target ($M = 56.8\%$, $SD = 21.5\%$). A main effect of trained object has not previously emerged with older age groups in this method (Quam & Swingley, 2010), though it did emerge with 24- and 30-month-olds in the present study (Experiment 2, main text) and was aso found in one similar study with 24-month-olds (Quam & Swingley, 2021, in prep.). It likely reflects the fact that younger children's fixations are more driven by visual salience of objects. While we attempted to equate visual salience, it could be, e.g., that the red object was brighter.

Given that the main effect of Object likely indicated that children had a visual preference for the "red knobs" object that was impacting target-fixation proportions, we conducted the ANOVA again with preference-corrected looking times as the dependent variable. The main effect of *Deebo* Object went away, but there was a new main effect of Trial Type, $F(1,30) = 6.09$, $p = .02$, reflecting overall higher preference-corrected looking in MP ($M = 9.9\%$, $SD = 19.2\%$) than CP trials ($M = 3.0\%$, $SD = 17.0\%$). There was also a significant four-way interaction of Trial Type, MP Type, Trained Pitch, and Object, $F(1,30) = 4.43$, $p = .044$, but as the study design did not allow the statistical power to investigate a 4-way interaction (sample sizes in some of the subgroups were as small as n=3), and it was not of strong theoretical interest, we declined to further investigate it.

Nineteen-month-olds showed inconsistent word learning in the paradigm—only the group tested with pitch-MPs showed above-chance word learning—and also did not detect mispronunciations of the initial consonant, nor of the pitch contour. It is not clear why we did not

find robust word learning in the consonant-MP group. The word was taught in a fairly complex narrated story-book context, which, along with the presence of the unlabeled distracter object, could have increased the task difficulty. Lack of robust word learning for the consonant-MP group means detection of the consonant MP could not be fairly evaluated.

It is important to note that, while children in the consonant-MP condition did not show robustly above-chance word learning, children in the pitch-MP condition did. Thus, we can fairly evaluate whether 19-month-olds detected the pitch change. Results indicated that children did not treat the pitch change as relevant, looking at the *deebo* object no differently when the pitch was correctly pronounced versus mispronounced. This result is consistent with the findings of Hay, Graf Estes, Wang, and Saffran (2015). Hay et al. tested 14-, 17-, and 19-month-olds' willingness to learn two words differing only in their tonal pattern in the Switch habituation procedure. Words contained Mandarin tone 2 (rising) vs. 4 (falling) and both words (rising /kʊ/ and falling /kʊ/) were taught during habituation. Only the 14-month-olds detected mismatches of words and objects, i.e., tonal mispronunciations of words (and even 14-month-olds only seem to do so when one tone in the pair is rising; Hay, Cannistraci, & Zhao, 2019).

While 19-month-olds' insensitivity to pitch changes in newly learned words here is consistent with Hay et al.'s (2015) findings, it contrasts with the findings of two other studies. First, Singh, Hui, Chan, and Golinkoff (2014), like Hay et al., taught words containing rising vs. falling Mandarin tones. They used a method more similar to the one used here, but taught the words only via ostensive labeling. They also taught two similar-sounding words, so that both objects they presented in the test phase had been previously labeled. Singh et al. found that 18-month-olds were willing to treat both tone mispronunciations (e.g., leng2 changing to leng4; numbers are standardly used to refer to the four tone contours of Mandarin: 2=rising, 4=falling)