2024

# Developmental Change in English-Learning Children's Interpretations of Salient Pitch Contours in Word Learning

Carolyn M. Quam
*Portland State University*, cquam@pdx.edu

Daniel Swingley
*University of Pennsylvania*

DOI: 10.1111/infa.12587

RESEARCH ARTICLE

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

# Developmental change in English-learning children's interpretations of salient pitch contours in word learning

**Carolyn Quam**[1,2] | **Daniel Swingley**[2]

[1]Department of Speech and Hearing Sciences, Portland State University, Portland, Oregon, USA

[2]Department of Psychology, University of Pennsylvania, Philadelphia, Pennsylvania, USA

**Correspondence**
Carolyn Quam, Portland State University, Department of Speech and Hearing Sciences, PO Box 751, Portland, OR, 97207-0751, USA.
Email: cquam@pdx.edu

**Funding information**
National Science Foundation SBE Office of Multidisciplinary Activities, Grant/Award Number: 1917608; National Institute of General Medical Sciences of the National Institutes of Health, Grant/Award Number: RL5GM118963; National Science Foundation Division of Behavioral and Cognitive Sciences, Grant/Award Number: HSD-0433567; Eunice Kennedy Shriver National Institute of Child Health and Human Development, Grant/Award Number: R01-HD049681

**Abstract**
To efficiently recognize words, children learning an intonational language like English should avoid interpreting pitch-contour variation as signaling lexical contrast, despite the relevance of pitch at other levels of structure. Thus far, the developmental time-course with which English-learning children rule out pitch as a contrastive feature has been incompletely characterized. Prior studies have tested diverse lexical contrasts and have not tested beyond 30 months. To specify the developmental trajectory over a broader age range, we extended a prior study (Quam & Swingley, 2010), in which 30-month-olds and adults disregarded pitch changes, but attended to vowel changes, in newly learned words. Using the same phonological contrasts, we tested 3- to 5-year-olds, 24-month-olds, and 18-month-olds. The older two groups were tested using the language-guided-looking method. The oldest group attended to vowels but not pitch. Surprisingly, 24-month-olds ignored not just pitch but sometimes vowels as well—conflicting with prior findings of phonological constraint at 24 months. The youngest

The editor of this article is Gavin Bremner.

group was tested using the Switch habituation method, half with additional phonetic variability in training. Eighteen-month-olds learned both pitch-contrasted and vowel-contrasted words, whether or not additional variability was present. Thus, native-language phonological constraint was not evidenced prior to 30 months (Quam & Swingley, 2010). We contextualize our findings within other recent work in this area.

## 1 | INTRODUCTION

To successfully recognize words across different talkers and utterances, children learning an intonational language like English should encode words in a format that abstracts away from pitch contour, which is not lexically contrastive but which is used for other purposes in the language. Thus, when hearing a new word for the first time, English learners should not assume that the pitch contour it was realized with is somehow intrinsic to the specification of the word; it is more likely to have been a feature of the utterance broadly (Quam & Swingley, 2010). However, learners should not simply ignore pitch features, as they play several vital roles in English, including conveying emotion, helping to specify lexical stress, and marking meaningful distinctions like indicating yes/no questions. Learning to optimize this kind of attribution takes time over development. Infants show early sensitivity to the attention-grabbing properties of infant-directed speech prosody (Fernald, 1985, 1992; Fernald & Kuhl, 1987; Katz et al., 1996) and to its corresponding pragmatic functions (Moore et al., 1997). Still, children show protracted developmental trajectories for interpretation of pitch cues to emotional prosody (e.g., Quam & Swingley, 2012) and lexical stress (Quam & Swingley, 2014). Likewise, while infants exhibit reduced discrimination for many non-native contrasts by the end of the first year of life (Polka & Werker, 1994; Werker & Tees, 1984), the linguistic interpretation of readily discriminable sound changes in words exhibits a more protracted developmental time-course (e.g., Dietrich et al., 2007; Quam & Swingley, 2023; Stager & Werker, 1997).

Some studies have investigated the developmental time-course of pitch processing for intonation-language learners by testing their discrimination of phonological tone patterns signaled by pitch in tonal languages. While some studies have shown declines in tone discrimination across the first year (e.g., Yeung et al., 2013), recent evidence suggests that tones are sometimes discriminable by non-tone-learning infants as late as 11–12 months (Chen & Kager, 2016; Singh, Fu, Tay, & Golinkoff, 2018), particularly for salient tonal contrasts (Shi et al., 2017; Tsao, 2017). While Liu and Kager (2014) found decreased sensitivity to lexical tones at 9 months, they found a rebound in discrimination around 18 months (see also Singh, Fu, Seet, et al., 2018). Discrimination trajectories for lexical tone parallel those for non-linguistic (musical) pitch (Chen et al., 2017), suggesting that tone discrimination in non-tone-language learners may reflect general pitch-discrimination ability. Thus, it is not clear whether processing of tones and intonation is narrowed by native-language input in the same manner as processing of consonants and vowels (see, e.g., Ramon-Casas et al., 2009).

A few experimental studies have investigated the developmental course over which English-learning children come to rule out pitch contour as lexically contrastive. Together, these studies point to the conclusion that children learning English disregard pitch as lexically contrastive sometime around 17 or 18 months of age. However, these studies have used diverse lexical contrasts and have not extended beyond 30 months, making it difficult to paint a continuous developmental picture of children's interpretation of pitch variation between early toddlerhood and adulthood. The goal of the present study is to help specify the developmental progression over a broader age range, by extending prior work with 30-month-olds and adults (Quam & Swingley, 2010) earlier—to 18- and 24-month-olds—as well as to an intermediate age group, preschoolers.

The present study's methods and stimuli are based on a study by Quam and Swingley (2010), which, using a language-guided looking method, found that English-speaking 30-month-olds and adults disregarded intonational changes but attended to vowel changes in newly learned words. Children and adults were taught a word that in training was always pronounced with a consistent, exaggerated pitch contour (a rise-fall or low-falling contour). Pitch contours were chosen to be salient and plausibly interpretable as English prosodic patterns (and not as tone contours from a different, tonal language). This was with the goal of seeing whether children would interpret pitch as a potentially important feature of specific words, as opposed to a feature of the utterance independent of its lexical items. If all a child's experience with a new word manifested a consistent, salient pitch contour, would children consider the contour part of the word, or would they follow their native phonology and treat pitch as lexically irrelevant? Quam and Swingley (2010) found that 30-month-olds, and adults, recognized novel words without hindrance when the words were given a new pitch contour they had not heard the words realized with before, whereas they were hindered by a phonological change in the novel word's vowel, consistent with English phonology.

Several additional papers have built on Quam and Swingley's (2010) findings, extending experimental investigations to younger populations. Singh et al. (2014) taught English-learning and Mandarin-learning 18-month-olds pairs of words that were realized with Mandarin tones (such as *leng*[tone2] and *beng*[tone2]), and evaluated whether children would find altered realizations harder to recognize, with the alterations either to the vowel (*leng*[2] realized as *ling* [2]) or to the tone (*leng*[2] realized as *leng*[4]). English learners' recognition of the words was hindered by tonal and vowel alterations at 18 months, but only by vowel alterations at 24 months. Mandarin learners' recognition was hindered by both sorts of mispronunciation at both 18 and 24 months. This result suggests an initial state in which salient realizations of tone distinctions are interpreted as important, followed by a more language-specific pattern of interpretation 6 months later.

In a second study, Hay et al. (2015) used a different method, the Switch habituation procedure, to investigate English-learning 14-, 17-, and 19-month-olds' willingness or ability to learn two words differing only in their tonal pattern. Hay et al. also used Mandarin tone 2 (rising) versus 4 (falling), but both words (rising /kʊ/ and falling /kʊ/) were taught as referents for objects during habituation. Only the 14-month-olds detected mismatches of words and objects, whereas 17- and 19-month-olds apparently interpreted the two tones as equivalent (though 17-month-olds' responses were somewhat intermediate between 14- and 19-month-olds'). Hay et al. (2019) later reported that even 14-month-old English learners could only learn a Mandarin tonal contrast when one of the tones was rising (tone 2), suggesting infants were applying knowledge of the relevance of rising intonation in English for conveying questions, uncertainty, and so forth. However, unlike in Quam and Swingley (2010) and Singh et al. (2014), no segmental (e.g., vowel) baseline was included for comparison with detection of tonal contrasts in the Hay et al. studies, so it is possible that the children who failed to keep track of the tonal contrast would also have failed on segmental contrasts. Children of this age often succeed in the Switch procedure, but the typical implementation involves syllables exhibiting substantial pitch variation among instances of a given word. This variation, necessarily absent in Hay et al. (so that tones could be consistently realized), might assist children in staying on task during the Switch procedure and in focusing their attention on relevant features of the contrast (e.g., Galle et al., 2015).

Burnham et al. (2018) also used the Switch habituation procedure to teach English-learning and Mandarin-learning 17-month-olds lexical-tone contrasts from Mandarin and Thai. English learners were also tested on English intonational contrasts. As in the Hay et al. studies, no segmental baseline was included (which would rule out the possibility of general insensitivity to both phonological and non-phonological contrasts in the task). Mandarin-monolingual and Mandarin-English bilingual children were able to learn words differing in Mandarin tone when the tonal contrast was tone 1 (high) versus tone 2 (rising; though they did not learn words contrasting in Mandarin tone 2 vs. 4, falling, nor did they learn Thai tone contrasts). Monolingual English-learning children were unable to learn any tonal contrasts—nor did they learn English-intonation-contrasted words.

Our goals in the present study were related to those of these previous studies. Rather than tracing the development of interpretation of phonological pitch patterns from tone languages, we aimed to evaluate how children come to partition *English* pitch variation as lexically relevant or not. It is possible that what English-learning toddlers had learned between 18 and 24 months in the Singh et al. (2014) study was to ignore other-language tone contrasts, having learned that such pitch patterns are different from typical English prosody. Burnham et al. (2018) found English learners did not learn intonation-contrasted words at 17 months. However, this lack of sensitivity was not compared to a segmental baseline. Comparison to a segmental baseline is particularly useful in the language-guided looking procedure employed here in Experiments 1 and 2, given recent evidence of insensitivity to segmental (consonant) mispronunciations at 24 and 30 months in this same procedure (Quam & Swingley, 2023). We compared pitch-contrasted words to vowel-contrasted words in Experiment 3 as well, considering that low-variability training could suppress overall sensitivity to contrasts even in the Switch procedure, where 15-month-olds have been shown to detect only some vowel contrasts (Curtin et al., 2009). Our overall goal was to address how children interpret English intonational variation—compared with vowel variation—on novel lexical items: as a feature of the noun to be encoded into the lexical representation, or not?

While infants learn a tremendous amount about their native language's speech sounds and phonological patterns in the first year of life, there is evidence of developmental change in speech processing in preschool and even school age (Hazan & Barrett, 2000; Oh et al., 2011; see Creel & Quam, 2015, for discussion). Children are less adept than adults at flexibly shifting cue weights to capitalize on local regularities (Nittrouer et al., 2000; Quam & Swingley, 2014). The idea that children are still learning to attribute meaning to pitch variation through the preschool years is consistent with the finding that they cannot reliably interpret prosodic cues to emotions until age four (Quam & Swingley, 2012; see also Creel & Jimenez, 2012).

There is also substantial evidence of continued development in word-learning ability in preschool and school age. Vocabulary size in toddlerhood is predictive of vocabulary size in preschool, but vocabulary at both ages may be influenced by individual differences in lexical-processing speed (Mahr & Edwards, 2018; see also Law & Edwards, 2015; Law et al., 2017). While toddlers often succeed in lab-based fast-mapping tasks, retention of word-meaning mappings is still poor at 24 months (Horst & Samuelson, 2008). Encoding may be the main bottleneck limiting retention at 33 months (Munro et al., 2012).

Children's phonological and lexical knowledge are viewed as developing in tandem (Edwards et al., 2011; Jusczyk, 1992). Still, phonological and semantic aspects are subserved by distinct underlying mechanisms (Gray et al., 2020). Even at age 8, word learning is not instantaneous; children build up knowledge of a particular word gradually over time, for both phonological and semantic aspects (McGregor et al., 2007). Children's word learning improves across age along several dimensions. First, children are increasingly able to cope with irrelevant phonetic variability. Ryalls and Pisoni (1997) found that between ages 3 and 5, English-learning children's word recognition becomes more robust to multiple talkers, a form of lexically irrelevant variability. Singh and Chee (2016) tested Mandarin-learning toddlers and preschoolers on recognition of tone-bearing words in the presence of intonational variation, finding that only children over age 4 were able to recognize tones despite intonational variation. Second, preschool children are still developing the ability to interpret small changes in the sounds of words. Law and Edwards (2015) found that vocabulary size predicted English-learning 30- to 46-month-old children's ability to detect one-feature mispronunciations of familiar words. They also found that vocabulary size predicted the degree to which children interpreted novel words as names for previously unlabeled objects (see also Bion et al., 2013).

## 2 | THE PRESENT STUDY

Here, we tested 18-month-old, 24-month-old, and 3- to 5-year-old children, in 3 experiments. Our study aimed to extend Quam and Swingley's (2010) findings with 30-month-olds and adults to earlier and intermediate ages to characterize the developmental trajectory of interpretation of pitch and vowel mispronunciations of newly learned words. First, we tested 3- to 5-year-olds and 24-month-olds in the language-guided looking procedure used by Quam and Swingley (2010; see Table 1 for an overview of the present experiments). In Experiment 1, 3- to 5-year-old children were taught a single word in the language-guided looking procedure and then tested on word recognition in response to correct pronunciations, pitch mispronunciations, and vowel mispronunciations (cf. Quam & Swingley, 2010). In Experiment 2, 24-month-old children were tested using the same method, except that each child heard one or the other mispronunciation type, not both.

In Experiment 3, we tested 18-month-olds, using a different method. Work reported elsewhere (Quam & Swingley, 2023; *Supplemental Materials*) indicates that children under 2 years do not always learn words robustly in the training and testing procedure used by Quam and Swingley (2010). Thus, here we tested 18-month-olds in the Switch habituation procedure. In Experiment 3, 18-month-old toddlers were taught two pitch- or vowel-contrasting words in the two-object version of the Switch procedure (Hay et al., 2015). To

**TABLE 1** Experimental procedures (in rows) and ages tested (columns) in the three experiments reported in the main text and two pilot experiments reported in Supplemental Materials (*S1*, *S2*).

| | 3–5 years | 24 months | 18 months |
|---|---|---|---|
| Training phase + language-guided looking test | Expt. 1 | Expt. 2 | |
| Switch habituation procedure | | *Expt. S1* | Expt. 3 |
| Switch + language-guided looking test | | *Expt. S2* | |

address whether low phonetic variability might have contributed to children's lack of learning in other studies (Burnham et al., 2018; Hay et al., 2015), half of children were trained with phonetic variability introduced on an irrelevant dimension. The other half were trained without this variability.

Two additional experiments with 24-month-olds are reported in the Supplemental Materials. Both of these used the same Switch habituation training as the low-variability condition of Experiment 3, but the two experiments differed from each other in their test procedures. Experiment S1 used a Switch test procedure identical to Experiment 3, whereas Experiment S2 used a potentially more sensitive language-guided looking test phase (Yoshida et al., 2009). In both Experiments S1 and S2, 24-month-olds showed no evidence of learning either vowel- or pitch-contrasted words. While this finding may speak to 24-month-olds' difficulty applying phonological knowledge to word learning, it also might indicate that the Switch procedure is not well calibrated to the developmental level of 24-month-old learners, who are on the older end for the procedure (though see Singh & Tan, 2021, for evidence of 24-month-olds' successful learning in Switch). Thus, these null findings at 24 months are reported in Supplemental Materials.

## 2.1 | Experiment 1

In Experiment 1, we tested 3- to 5-year-olds in the same language-guided looking procedure in which adults previously detected vowel changes but not pitch changes in newly learned words (Quam & Swingley, 2010). In testing preschoolers, our goals were to paint a more continuous developmental picture of pitch and vowel interpretation in newly learned words from 18 months through adulthood, and to add to the knowledge-base about word learning beyond toddlerhood (e.g., McGregor et al., 2007). We taught children a word with a consistent pitch pattern and then tested their responses to the original word versus versions with the pitch pattern or the vowel mispronounced (each child was tested with all pronunciations).

### 2.1.1 | Method

The method was nearly identical to the one previously used with adults by Quam and Swingley (2010). Details of the experimental procedure and the visual and auditory stimuli are given there (see in particular that paper's figs. 1–3, which are reprinted for reference in the Supplemental Materials for this article). The task lasted approximately 20 min. Children 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, both of which extended through the second (unstressed) syllable of the word. 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 (while the other toy was present but not labeled) and again 12 more times during an ostensive-labeling phase in which the object was presented alone on the screen. In both of these training phases, the other novel object was shown equally often but was never labeled.

In the test phase, children saw, intermixed, 8 familiar-word filler trials, 8 correct-pronunciation (CP) trials, five pitch-change trials, and 5 vowel-change trials. In CP and change trials, the two objects appeared on the screen, and children heard a question containing either the original word or a version of that word with either the pitch contour or the vowel altered ("mispronounced"). Like the children and adults tested by Quam and Swingley (2010), at the end of the experiment participants were asked to point to and name the pictures. Pointing trials offered another measure of children's interpretation of the pronunciation variants, and were also included in the eye-gaze analyses.

*Participants*

All procedures involving human participants were approved by the Institutional Review Board at the University of Pennsylvania and conducted in accordance with the Declaration of Helsinki. Written parental permission was collected from each child's parent or guardian prior to the testing session. Sixty-two children were recruited into the study and tested. The final sample consisted of 47 children (26 girls and 21 boys) between the ages of 3 years, 14 days and 5 years, 8 months, and 28 days. (We had intended to test to 48, but exclusions for insufficient useable trials—see below—could not be confirmed until after we had closed recruitment for the study.) All caregivers reported that children were learning English as their native and dominant language, and no children were reported to have over 10% exposure to a tonal language. Their mean age was 4 years, 4 months, and 6 days ($SD$ = 10 months, 19 days). Fifteen children participated but were excluded for having fewer than 3 useable trials (including the point trial) in one or more trial types (5), fussiness and/or not completing the experiment (4), equipment failure (4), or parent report of language or developmental delays (2). Trials were included as useable if the child fixated the picture for at least 20 frames during the analysis window, out of a possible 50 (i.e., was on task in at least 40% of the frames; see Swingley & van der Feest, 2019, for a similar criterion). Vocabulary information was not collected for 3- to 5-year-olds.

### 2.1.2 | Results and discussion

We analyzed children's looking times to the pictures as well as children's pointing and naming of the pictures. Gaze responses are a more implicit and gradient measure of word learning and sensitivity to mispronunciations, while pointing and naming are discrete and explicit measures of learning. Target-fixation proportions were averaged over all trials with each pronunciation (correct pronunciation, or CP; vowel mispronunciation, or MP; tone MP), including point trials. Figure 1 (left) displays *deebo*-fixation proportions in CP and MP trials. In order to determine whether children had learned the word, in a series of preliminary analyses we compared their target fixation to chance (50%) in correct-pronunciation (CP) trials, using a two-tailed, one-sample $t$ test. Children's *deebo* fixation in CP trials was significantly above chance ($M = 68.2\%$, $SD = 15.4\%$), $t(46) = 8.09$, $p < 0.001$. We next evaluated whether children's responses revealed significant fixation of the *deebo* given either the pitch or vowel change. Children's *deebo* fixation was also significantly above chance in pitch-MP trials ($M = 63.0\%$, $SD = 18.6\%$), $t(46) = 4.80$, $p < 0.001$. In response to the vowel change, children actually fixated the *deebo* object significantly *below* chance levels ($M = 42.1\%$, $SD = 20.3\%$), $t(46) = -2.69$, $p = 0.010$.

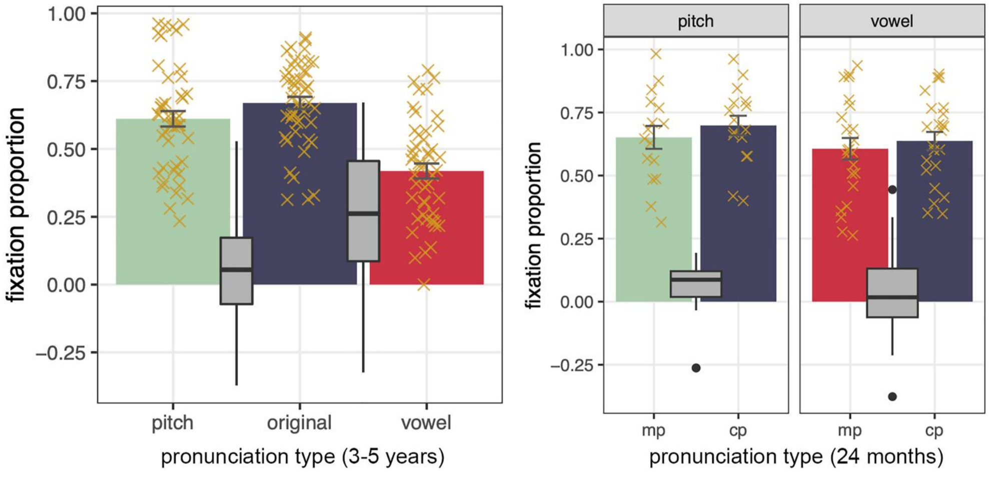


**FIGURE 1** Looking patterns in the language-guided looking method. Three- to five-year-olds (left) were tested in all 3 conditions. Twenty-four-month-olds (right) were tested with either pitch or vowel mispronunciations. Box plots indicate within-subject difference scores between correct/original-pronunciation and mispronunciation trials.

To evaluate whether children showed different patterns of looking across experimental groups and conditions, we fit multilevel binomial regression models that predicted trial-by-trial target fixations.[1] A binomial regression model has two primary advantages over the more traditional approach of conducting analyses of variance (ANOVAs) on subject by condition means. First, by analyzing trial-by-trial data, it enables inclusion of variables that vary by trial (or in some studies, by item). Second, multilevel models allow for explicit modeling of subject variation. Binomial models in particular are appropriate for looking-time proportions because these proportions are bounded by zero and one, and are often not normally distributed when considered at the trial level (making ordinary linear regression inappropriate). Standard error estimates for non-binary outcomes generated via standard logistic models can be overly conservative, so we estimated standard errors via a bootstrapping technique (following Humphrey & Swingley, 2018; Swingley & van der Feest, 2019, using the R package glmmTMB; Brooks et al., 2017) based on 25,000 runs.

The dependent variable was the target-fixation proportion in each trial over the time window 367–2000 ms. after noun onset. In addition to our predictor variable of interest, Trial Type (CP, pitch-MP, vowel-MP), we considered three additional variables that have sometimes been shown to influence toddlers' gaze patterns in prior work. First, we included Age, and the interaction of Age with Trial Type, to probe for development within the 3- to 5-year age range in this task. Second, as in prior work (Quam & Swingley, 2010, 2023) we checked for effects of Trained Pitch Contour (rise-fall or low fall) to determine the consistency of effects across non-phonological variation in the training items. Finally, we checked for effects of Gaze Location at Noun Onset. In language-guided looking procedures, children who happen to begin the trial fixating the target picture are more likely to continue fixating the target than children who

[1]More traditional analyses of variance (ANOVAs) are included in Supplemental Materials.

happen to begin on the distracter are to shift over to the target. The inclusion of this nuisance variable helps account statistically for this asymmetry across trials.

As the variable of primary interest was Trial Type, we followed a common procedure in multilevel modeling of checking for effects of additional variables in initial models and then removing them from the final model if they showed no significant effects or interactions and the effects of interest were not meaningfully changed by their inclusion/exclusion. Age (and the Age X Trial Type interaction), Trained Pitch Contour, and the interaction between Gaze Location at Noun Onset and Trial Type were dropped from the final model following this procedure. The predictors in the final model were Trial Type, Gaze Location at Noun Onset, and Subject (a random effect). Fixed effects were treatment coded with CP as the Trial-Type baseline and distracter as the Gaze-Location baseline. A follow-up analysis set the reference condition to pitch-MP to evaluate the significance of the response difference between pitch-MP and vowel-MP.

As shown in Table 2, the multilevel regression analysis of trial-by-trial target looking proportions revealed a significant difference in target fixation when the vowel was mispronounced (see "Trial Type—Vowel"). The nuisance variable Gaze Location at Noun Onset was also a significant predictor of target looking. The exponentiated coefficients give an estimate of the multiplicative change in the odds of fixating the target relative to the distracter given a change of one unit of the measured variable. For example, the 0.36 exponentiated coefficient for the vowel-mispronunciation trials ("Trial Type—Vowel") means that if no other effects were present, we would expect that a child looking at the target 75% of the time window in CP trials (0.75/(1–0.75) = odds of 3) would look at the target (0.36 X 3 = odds of 1.08) or (1.08/(1 + 1.08)) = 51.9% of the time in vowel-MP trials. This reduction in target looking given vowel MPs was large (mean, 26.1 percentage points) and shown by 38/47 participants (81%; binomial $p < 0.001$). By contrast, the reduction in target looking given pitch changes was smaller (mean, 5.2%) and shown by only 28/47 participants (60%; binomial $p$ ns). To compare the two MP conditions, we re-ran the same model with pitch-MP as the reference level. There was a significant difference between target fixation on vowel-MP trials and pitch-MP trials ($\beta = -0.822$, $\exp(\beta) = 0.440$, s.e. 0.183, $p = 0.0002$).

We also queried the degree to which children mapped each pronunciation variant onto the previously unlabeled object (Law & Edwards, 2015). A substantial proportion, 33/47 children (70%), fixated the *deebo* less than 50% of the time in vowel-MP trials, binomial $p = 0.008$, suggesting they used a mutual-exclusivity strategy to map the word "dahbo" onto the distracter object (Markman & Wachtel, 1988; Quam & Swingley, 2010). By contrast, only 14/47 children (30%) did so in the pitch condition, binomial $p$ n.s. These two proportions differed significantly, Pearson's chi-squared statistic = 13.79, df = 1, $p$(2-tailed) < 0.001.

**TABLE 2** Regression coefficients and other statistics for preschoolers' word learning in Experiment 1.

| *Predictor* | *Coef* | Exp*(coef)* | *Std.error* | *p(boot)* |
|---|---|---|---|---|
| (Intercept) | 0.503 | 1.654 | 0.125 | <0.001 |
| Trial type - pitch | −0.210 | 0.810 | 0.173 | 0.124 |
| Trial type - vowel | −1.032 | 0.356 | 0.170 | <0.001 |
| Gaze onset - target | 0.489 | 1.631 | 0.147 | <0.001 |

*Note*: "Coef." is the beta coefficient estimated by the regression analysis; "Exp(coef)" is its exponent, showing the multiplicative change in odds ratio; "Std.error" is the standard error of this estimate as given by the binomial regression; "p(boot)" is the probability of the effect under chance assumptions, as estimated using bootstrapping (see text).

Children's pointing and naming responses can provide another lens on their interpretations of pitch and vowel mispronunciations. Table 3 reports pointing and naming responses for the 3- to 5-year-olds tested here compared with the 30-month-olds and adults tested by Quam and Swingley (2010). We first report pointing data. Only children who responded in all three point trials are included (Quam & Swingley, 2010). When asked to "Point to the deebo" pronounced with the trained pitch pattern, 92% of consistent responders ($n = 37$) pointed to the deebo object and 8% pointed to the distracter object. When asked to "Point to the dahbo," only 14% pointed to the deebo object; 86% pointed to the distracter object. The deebo-pointing proportion in response to the vowel change was significantly lower than to both the trained pronunciation, Pearson's chi-squared statistic = 42.503, df = 1, $p$(2-tailed) < 0.001, and the pitch change, Pearson's chi-squared statistic = 28.784, df = 1, $p$(2-tailed) < 0.001. Unlike 30-month-olds and adults (Quam & Swingley, 2010), preschoolers reduced their *deebo* pointing when the pitch was mispronounced—78% pointed to the *deebo* and 22% pointed to the distracter—but not to a significant degree compared with the trained pronunciation, Pearson's chi-squared statistic = 1.709, df = 1, $p$(2-tailed) = 0.191.

We turn next to naming data. As in prior work (Quam & Swingley, 2010), we scored productions for whether the first syllable contained /i/ or /a/. When asked to label the *deebo* object, children produced more /i/ vowels (33) than /a/ vowels (0). One child said "dahbo...deebo" so we coded the second (presumably corrected) pronunciation. When asked to label the distracter object, children were a bit more reluctant to produce a label, but those who did produced slightly more /a/ vowels (10) than /i/ vowels (7). Two of the children who produced /i/ vowels did so in questions ("deebo?" "Is it a ball? Is it a deebo?").

To summarize, the gaze, pointing, and naming data converged to indicate that English-learning children ages 3 through 5, regardless of age, showed robust word learning and treated a vowel MP—but not a pitch MP—as relevant in deciding whether the taught object and the variant pronunciation nevertheless went together. Gaze data were fairly convergent with prior findings with older and younger learners (Quam & Swingley, 2010). Table 4 reports target-fixation proportions in CP, pitch-MP, and vowel-MP trials across age. As with younger and older learners, preschoolers did not show particular sensitivity to the pitch change (other than a non-significant decrease in pointing to the *deebo* object).

**TABLE 3** Percentage of children pointing to the *Deebo* object in CP, Pitch-MP, and Vowel-MP trials; and fraction (with percentage) using /i/ vowel when naming the *Deebo* versus the distracter object.

| | | Pointing | | Naming | |
|---|---|---|---|---|---|
| | | *CP* | *MP* | *Deebo* | *Distracter* |
| 30 months | *Pitch* | 75% | 83% | 100% (15/15) | 100% (2/2) |
| | *Vowel* | 100% | 55% | 93% (14/15) | 17% (1/6) |
| 3–5 years | *Pitch* | 92% (34/37) | 78% (29/37) | 100% (33/33) | 41% (7/17) |
| | *Vowel* | | 14% (5/37) | | |
| Adults | *Pitch* | 100% | 100% | 100% (22/22) | 6% (1/16) |
| | *Vowel* | | 21% | | |

*Note*: Included are 30-month-olds from Quam and Swingley (2010), 3–5-year-olds from Experiment 1, and adults from Quam and Swingley (2010). For naming data, percentages are calculated as the number using the /i/ vowel divided by total number using /i/ or /a/; raw numbers of children are also provided, as few children responded in some conditions. Raw numbers of children are also included for pointing responses at 3–5 years (the age tested here).

**TABLE 4** Mean Target-Fixation Proportions in CP, Pitch-MP, and Vowel-MP Trials.

| | | CP | MP | % showing MP effect | % using ME strategy |
|---|---|---|---|---|---|
| 24 months | *Pitch* | 69.1% | 65.1% | 75% | 25% |
| | *Vowel* | 63.4% | 60.7% | 61% | 26% |
| 30 months | *Pitch* | 66.3% | 72.9% | 33% | 8% |
| | *Vowel* | 67.4% | 52.8% | 83% | 46% |
| 3–5 years | *Pitch* | 68.2% | 63.0% | 60% | 30% |
| | *Vowel* | | 42.1% | 81% | 70% |
| Adults | *Pitch* | 91.8% | 89.3% | 50% | 0% |
| | *Vowel* | | 39.7% | 100% | 75% |

*Note*: Included are 24-month-olds from Experiment 2, 30-month-olds from Quam and Swingley (2010), 3- to 5-year-olds from Experiment 1, and adults from Quam and Swingley (2010). The rightmost 2 columns list the percentage of children looking less to the *deebo* on MP trials than CP trials (showing an MP effect) and percentage looking less than 50% of the time in MP trials (using a mutual exclusivity, ME, strategy).

## 2.2 | Experiment 2

In Experiment 2, we tested 24-month-olds in a language-guided looking procedure very similar to that used in Experiment 1. Two modifications were made to the test phase, as described below.

### 2.2.1 | Method

The word-teaching phase was identical to that of Experiment 1. In the test phase, due to 24-month-olds' more limited attention spans, each child was presented with either the pitch change or the vowel change (as was done with 30-month-olds in Quam & Swingley, 2010). Thus, children saw, intermixed, 8 familiar-word filler trials, 8 correct-pronunciation (CP) trials, and *either* eight pitch-change trials or 8 vowel-change trials. Unlike in Experiment 1, children were not asked to point to and name the pictures at the end of the experiment, due to the lower likelihood that 24-month-olds would be able to provide explicit responses (as observed by Quam & Swingley, 2023; see *Supplemental Materials*).

*Participants*

Thirty-nine children (17 girls and 22 boys) between the ages of 22 months, 26 days and 25 months, 19 days were included in the analysis. Sixteen children (5 girls and 11 boys) were included in the *pitch-change* condition (mean age = 2 years, 14 days; *SD* = 31 days). Twenty-three children (12 girls and 11 boys) were included in the *vowel-change* condition (mean age = 1 year, 11 months, 26 days; *SD* = 28 days). Their mean productive vocabulary was 313 words (*SD* = 194 words; vocabulary data not collected for one participant). Inclusion criteria matched Experiment 1. Seventeen children participated but were excluded for having fewer than three useable trials in any one trial type (9); fussiness (6); lost data (1); and language background (1). Again, trials were only included as useable if the child fixated the picture for at least 20 frames during the analysis window, out of a possible 50.

### 2.2.2 | Results and discussion

As in Experiment 1, target-fixation proportions were averaged over all trials with each pronunciation (correct pronunciation, or CP; vowel mispronunciation, or MP; pitch MP). Figure 1 (right) displays *deebo*-fixation proportions in CP and MP trials. In order to determine whether children had learned the word, we first compared their target fixation to chance (50%) in correct-pronunciation (CP) trials, using a two-tailed, one-sample *t*-test. Children's *deebo* fixations in CP trials were significantly above chance (vowel-change group: mean, 63.4%, $SD = 17.2\%$; $t(22) = 3.74$; $p = 0.001$; pitch-change group: mean, 69.1%, $SD = 14.1\%$; $t(15) = 5.45$; $p < 0.001$). Children's *deebo* fixation was also significantly above chance in both vowel-MP trials ($M = 60.7\%$, $SD = 20.4\%$); $t(22) = 2.50$, $p = 0.02$, and pitch-MP trials ($M = 65.1\%$, $SD = 18.2\%$); $t(15) = 3.33$, $p = 0.005$.

As in Experiment 1, to evaluate whether children showed different patterns of looking across experimental groups and conditions, we conducted binomial regression models that predicted trial-by-trial target fixations. The dependent variable was the target-fixation proportion in each trial over the time window 367–2000 ms. after noun onset. The predictors were Trial Type (CP vs. MP); Mispronunciation Type (pitch vs. vowel); Trained Pitch (rise-fall vs. low fall); Gaze Location at Noun Onset (target vs. distracter); and Subject (a random effect). All two- and three-way interactions were included between Trial Type, Mispronunciation Type, and Trained Pitch. Fixed effects were treatment coded with CP as the Trial-Type baseline, pitch as the Mispronunciation-Type baseline, rise-fall as the Trained-Pitch baseline, and distracter as the Gaze-Location baseline. We again followed the procedure of running an initial model with nuisance variables included and then dropping nuisance variables or their interactions from the final model if they showed no significant effects or interactions and the effects of interest were not meaningfully changed by their inclusion/exclusion. Following this procedure, the interaction between Mispronunciation Type and Gaze Location at Noun Onset was excluded from the final model.

As shown in Table 5, the multilevel regression analysis of trial-by-trial target-looking proportions revealed a significant three-way interaction of Trial Type, MP Type, and Trained Pitch. The nuisance variable Gaze Location at Noun Onset significantly predicted target looking, similar to its effect in Experiment 1.

The 1.86 odds ratio for the effect of Gaze Location at Noun Onset means that if none of the other effects were present, we would expect that a child who looks at the target 75% of the time for trials on which she starts at the distracter picture (75% = odds of 3) would instead have odds of 5.58 of looking at the target (1.86 X 3 = 5.58) if she were to start the trial fixating the target picture. This corresponds to $(3.72/(1 + 3.72)) = 84.8\%$ estimated target-fixation proportion in target-initial trials. Such an increase is typical: children who happen to start at the target have an advantage on that trial. The three-way interaction of Trial Type, MP Type, and Trained Pitch indicates that children in the vowel-MP condition who were trained with the low-fall contour had target fixations in MP trials that were significantly higher than would be expected based on performance in the other conditions (e.g., their relatively low target fixations in CP trials; see Figure 2, lower-right quadrant).

**TABLE 5** Regression coefficients and other statistics for 24-month-olds' word learning in Experiment 2.

| *Predictor* | *Coef* | *Exp(coef)* | *Std.error* | *p(boot)* |
|---|---|---|---|---|
| (Intercept) | 0.468 | 1.597 | 0.352 | 0.069 |
| Trial type - MP | 0.026 | 1.026 | 0.426 | 0.904 |
| MP type - vowel | 0.130 | 1.138 | 0.426 | 0.715 |
| Trained pitch - low fall | 0.073 | 1.075 | 0.495 | 0.815 |
| Gaze onset location- target | 0.623 | 1.864 | 0.197 | <0.001 |
| Trial type x MP type | −0.488 | 0.614 | 0.543 | 0.205 |
| Trial type x Trained pitch | −0.414 | 0.661 | 0.633 | 0.231 |
| MP type x Trained pitch | −0.652 | 0.521 | 0.625 | 0.092 |
| Trial type x MP type x Trained pitch | 0.996 | 2.708 | 0.798 | 0.040 |

*Note*: "Coef." is the beta coefficient estimated by the regression analysis; "Exp(coef)" is its exponent, showing the multiplicative change in odds ratio; "Std.error" is the standard error of this estimate as given by the binomial regression; "p(boot)" is the probability of the effect under chance assumptions, as estimated using bootstrapping (see text).

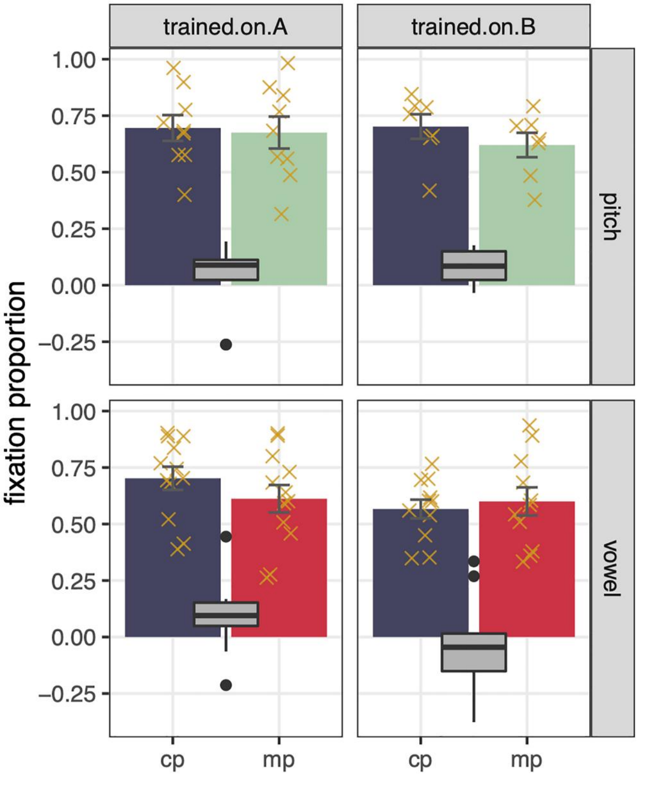


**FIGURE 2** Looking patterns for 24-month-olds in Experiment 2, split by trained pitch contour. "cp" = correct-pronunciation trials; "mp" = mispronunciation trials. "trained.on.A" = target word pronounced with rise-fall contour in training; "trained.on.B" = target word pronounced with low-fall contour in training. Children in the vowel-MP condition who were trained with the low-fall contour (bottom-right quadrant, red/lighter bar) had target fixations in MP trials that were significantly higher than would be expected based on performance in the other conditions.

To further investigate the three-way interaction, we conducted separate regressions for each MP Type separately (pitch vs. vowel). The results are summarized in Table 6. In the Pitch-MP condition, there was again only a similar effect of Gaze Location at Noun Onset, with an advantage for trials in which children were already fixating the target picture. This pattern is what would be expected if children intuited that pitch contour is not relevant to word identity. Results in the Vowel-MP condition were more complex. First, there was a significant effect of Trial Type, indicating that for children trained with the rise-fall contour (the baseline Trained Pitch; Figure 2, bottom-left quadrant), target fixations were lower in MP trials than CP trials, as expected. Next, there was a significant effect of Trained Pitch, indicating that in CP trials (the baseline Trial Type), target fixations were lower for children trained with the low-fall contour (Figure 2, bottom-right quadrant, dark-blue bar) than for children trained with the rise-fall contour. Finally, there was a significant interaction of Trial Type and Trained Pitch, indicating that, as indicated by the initial regression model, children trained with the low-fall contour had significantly higher target fixations in MP trials than would be expected based on performance in the other conditions.

The regression models taken together indicate that the responses of children trained with the low-fall contour in vowel-MP trials (Figure 2, bottom-right quadrant) differ from the other groups. The other 3 conditions (other three quadrants of Figure 2) show numerical trends toward higher target fixations in CP trials than MP trials—suggesting potential mispronunciation detection—but this trend is reversed in the low-fall pitch condition for vowel-MPs. Visual inspection of means suggests this is likely driven by lower-than-expected target fixations in CP trials.

In summary, in the language-guided looking procedure in which 30-month-olds and adults had previously attended to vowel changes but not pitch changes (Quam & Swingley, 2010), 24-month-olds showed less phonologically constrained responses. They robustly learned the

**TABLE 6** Regression coefficients and other statistics for 24-month-olds' responses to pitch versus vowel mispronunciations in Experiment 2.

| ***Pitch MP Condition*: *Predictor*** | ***Coef*** | **Exp(*coef*)** | ***Std.error*** | ***p(boot)*** |
|---|---|---|---|---|
| (Intercept) | 0.197 | 1.218 | 0.367 | 0.571 |
| Trial type - MP | 0.089 | 1.093 | 0.429 | 0.735 |
| Trained pitch—low fall | 0.129 | 1.137 | 0.466 | 0.686 |
| Gaze onset - target | 1.031 | 2.805 | 0.320 | <0.001 |
| Trial type x trained pitch | −0.463 | 0.630 | 0.637 | 0.215 |
| ***Vowel MP Condition*: *Predictor*** | ***Coef*** | **Exp(*coef*)** | ***Std.error*** | ***p(boot)*** |
| (Intercept) | 0.706 | 2.027 | 0.301 | 0.0125 |
| Trial type - MP | −0.463 | 0.630 | 0.337 | 0.0255 |
| Trained pitch—low fall | −0.613 | 0.542 | 0.399 | 0.0327 |
| Gaze onset - target | 0.406 | 1.501 | 0.249 | 0.0615 |
| Trial type x trained pitch | 0.607 | 1.835 | 0.488 | 0.0441 |

*Note*: "Coef." is the beta coefficient estimated by the regression analysis; "Exp(coef)" is its exponent, showing the multiplicative change in odds ratio; "Std.error" is the standard error of this estimate as given by the binomial regression; "p(boot)" is the probability of the effect under chance assumptions, as estimated using bootstrapping (see text).

words, but did not consistently detect vowel mispronunciations, with one group (trained with a rise-fall pitch contour) showing evidence of mispronunciation detection and the other (trained with a low fall) not detecting vowel mispronunciations. Pitch mispronunciations did not significantly hinder word recognition, suggesting that children had abstracted pitch contour away from their representation of the learned words; however, our effort to calibrate this theoretically interesting non-effect against a vowel mispronunciation led to inconsistent results.

These results in hand, we shifted to testing 18-month-olds in the Switch habituation method, in an effort to help establish the developmental timeline of children's lexical interpretation of pitch contours. Prior work showed that English learning 18-month-olds used vowel *identity* to differentiate words, but—unlike Dutch learners—they would not use vowel *duration* to do so (Dietrich et al., 2007). Given inconsistent prior findings on pitch interpretation around this age (e.g., detecting tonal MPs in Singh et al., 2014; but not in Hay et al., 2015), we tested whether toddlers would be sensitive to nonphonological differences in pitch, baselined against differences in vowel identity.

There is an unavoidable methodological discontinuity between Experiment 3, which uses the Switch method at 18 months, and Experiments 1 and 2, which used language-guided looking with older children. The switch to Switch at 18 months was necessary given prior findings of 19-month-olds' inconsistent word learning in the language-guided-looking method used in Experiments 1 and 2 (Quam & Swingley, 2023; *Supplemental Materials*). Nevertheless, the two methods differ on several dimensions. The language-guided looking procedure teaches a single word and then tests sensitivity to mispronunciations, leaving the relevance of pitch (and vowel) variation open for learners' interpretation. The two-object Switch method is more 'heavy handed' in conveying the relevance of contrastive variation, as it teaches minimal pairs. In language-guided looking, learners evidence mispronunciation detection by looking less to the target picture (and potentially gravitating toward the distracter picture) when the target word is mispronounced. In Switch, they look *longer* at the target picture when it is given a different label, indicating surprise (recovery from habituation).

We attempted to address this methodological discontinuity by testing 24-month-olds in Switch as well. We conducted two additional pilot experiments, using a "Switch" habituation procedure that has usually been employed with slightly younger children (Stager & Werker, 1997). As reported in the Supplemental Materials, these experiments did not provide evidence of word learning. While 24-month-olds' word learning has been successfully evaluated via the Switch procedure (e.g., Singh & Tan, 2021), the Switch procedure is more commonly used in children under 2 years of age. It is possible the procedure is less consistently reliable at 24 months than at earlier ages.

## 2.3 | Experiment 3

A Switch habituation experiment with 18-month-old infants was conducted to determine whether children would show phonologically constrained responses to mispronunciations of newly learned words, differentiating novel words by vowel identity but not pitch contour. Half of children were presented with words that included variability on an irrelevant dimension. We were interested in whether the introduction of irrelevant variability might help children to focus on phonologically relevant changes (vowels) and tune out non-phonological changes (pitch), similarly to how talker variability has been shown to boost differentiation of consonant-contrasted words (Quam et al., 2017; Rost &

McMurray, 2009, 2010). One group of 18-month-olds learned vowel-contrasted words (a phonological contrast) in the presence of (nonphonological) pitch variability. Another group learned the converse—pitch-contrasted words (a nonphonological contrast) in the presence of (phonological) vowel variability. Across Experiments 1 through 3, the overarching focus was on how children zero in on phonologically relevant dimensions of contrast and learn to listen through phonologically irrelevant ones. Given prior evidence that 18-month-olds are more likely to treat non-phonological (including intonational) variation as relevant to word learning (e.g., Hay et al., 2015; Singh et al., 2014), the high-variability conditions of Experiment 3 were intended to probe the limits of this 'open mindedness.' In particular, the condition teaching a pitch contrast in the presence of vowel variability was intended to push the limits of 18-month-olds' willingness to learn a non-native contrast in Switch. Would toddlers be willing to learn a non-native pitch contrast even when a *phonological* dimension (vowels) was varying?

### 2.3.1 | Method

*Participants*

Sixty-four English-learning children between the ages of 17 months, 2 days and 20 months, 8 days were included in the experiment. Inclusion criteria matched the previous experiments. Thirty-two children (15 girls and 17 boys) were included in the low-variability condition (mean age 18 months, 3 days, *SD* = 25 days; mean productive vocabulary = 209, *SD* = 120). Thirty-two more children (17 girls and 15 boys) were included in the high-variability condition (mean age 18 months, 13 days, *SD* = 25 days; mean productive vocabulary = 238, *SD* = 169; vocabulary not reported for 3 participants). Within each variability condition, half of children were habituated to pitch-contrasted words and half to vowel-contrasted words. Thirty-nine more children participated but were excluded (23 from the low-variability condition, 16 from the high-variability condition) for fussiness and/or failure to finish the experiment (29), equipment or recording failure (6), other-language exposure (2), the mother requesting to end the session (1), or parental influence on the child's responses (1). In the low-variability condition, one child did not complete a post-test trial (details below), but was retained in analyses because the post-test trial was not used as an exclusion criterion in the study overall.

*Apparatus and procedure*

Children were tested in the two-object version of the Switch habituation procedure (e.g., Stager & Werker, 1997, Experiment 1). In this procedure, children are habituated to repeated presentations of one word paired with an unfamiliar object, and another word paired with another object. Following habituation, children see more of the same word-object pairings (in two "same" trials), or the reversed word-object pairing (in two "switch" trials). If children have learned the word-object connections, and if they consider the words distinct, they are expected to look at the screen longer on "switch" trials than on "same" trials.

Audiovisual stimuli were presented using the software program Habit (Cohen et al., 2004; for which a newer version is now available; Oakes et al., 2019). Each trial began with an attention-getting stimulus that drew the child's gaze to the screen. This was a blue, rotating flower-like object accompanied by a pleasant, non-linguistic tone. After the infant had oriented to the attention getter, the experimenter pressed a button to initiate the trial. On each trial, one of two unfamiliar objects was presented visually on the screen accompanied by repeated tokens of one of two words. The two spoken words that were paired with the two objects differed in either their pitch contour or their vowel. Each child was randomly assigned to one of four pseudo-randomized trial orders for the habituation and test phases.

Within each trial, the experimenter pressed a button to mark start and end times of each of the child's looks to the screen. The total looking time for each trial was the sum of all looks within the trial. The Habit program calculated a baseline looking time, which was the sum of all looking in the first 3 habituation trials. It then calculated the sum of looking times across each of the following consecutive 3-trial sequences as they occurred, using a moving window. When the cumulative looking time across 3 consecutive trials reached 50% or less of the baseline looking time, the child was considered to have habituated (Oakes, 2010) and the Habit program presented the test trials. If children did not habituate by the 24$^{th}$ habituation trial, the habituation phase ended and they proceeded to the test phase. Of the 64 participants, 11 children did not habituate: 9/32 in the low-variability condition (28%) and 2/32 (6%) in the high-variability condition. These two proportions differed significantly, Pearson's chi-squared statistic = 3.95, df = 1, $p$(2-tailed) = 0.047. In consideration of the different rates of habituation across variability conditions, analyses are conducted over all participants (for maximum generalizability; Oakes, 2010) and again over only those who habituated, to rule out the possibility that non-habituators could add noise by showing familiarity preferences (Oakes, 2010).

The habituation phase began with a pre-test containing 3 familiar-word trials (ball, car, and shoe) intended to help get across to children the referential nature of the task (Fennell & Waxman, 2010). Then, we presented a pseudo-randomized sequence of the two novel word-object pairings. Once children had reached the habituation criterion (or the 24$^{th}$ habituation trial) they proceeded to the test phase. In the test phase, children saw two "Same" trials, in which the original word-object pairs were presented once each, and two "Switch" trials, in which the assignment of words to objects was switched from familiarization. In the final, post-test trial, the ball from the pre-test was presented again. The purpose of the post-test trial was to check whether, as a group, children would detect a very different word-object pair, indicating that they were still attending to the task.

The point when each infant reached the habituation criterion was determined via the experimenter's real-time coding of the infant's looking to the screen. However, all analyses were based on offline coding of infants' looking responses. Trained observers coded all participant videos frame-by-frame (with 33-ms resolution) in the SuperCoder program (Hollich, 2005). Alignment of the timing of eye-movement events with auditory and visual stimulus events was ensured using a custom hardware unit that placed visible signals into the recorded video stream of the participant's face.

### *Visual stimuli*

Visual stimuli were displayed on a rectangular plasma video screen measuring 94 by 53 cm. Visual stimuli were photographs of objects on gray backgrounds, displayed one at a time on the screen. All photos were edited to balance their salience by roughly equating brightness and size. The two novel toys, depicted in Figure 3, were a green and orange, triangular, tree-shaped toy and a yellow, plastic-looking, porcupine-like toy. Both toys were animate, with faces. Equating animacy across the two pictures is important, as young children show a strong bias for animate objects (e.g., Hofrichter et al., 2021). The pairing of words and objects was counterbalanced across participants.

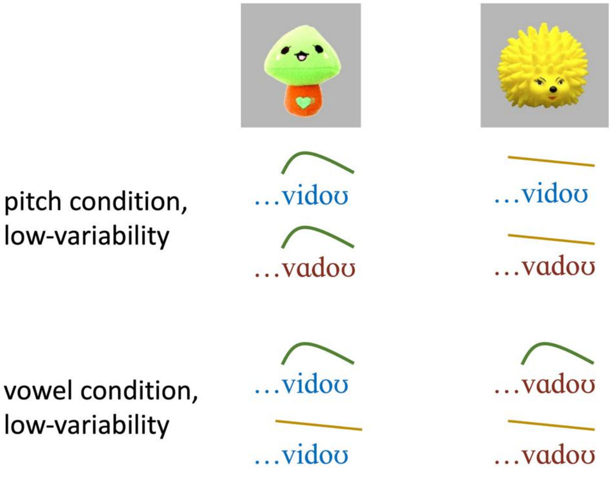


**FIGURE 3** The two visual objects used in Experiment 3, and a schematic of the auditory stimuli used in the low-variability conditions. Each row of auditory stimuli depicts one of the two word pairs to which an 18-month-old child might have been habituated in each condition.

### *Auditory Stimuli*

Auditory stimuli were recorded by a native English speaker (the first author) in an infant-directed register, with exaggerated prosody. Words were similar to those used by Quam and Swingley (2010, 2023). As in the prior studies, the phonological contrasts used were a rise-fall versus low-falling pitch contrast (see Quam & Swingley, 2010, for details) and the vowel contrast was /i/ versus /a/. Two of the eight tokens were presented in isolation, and six were presented in different carrier phrases (This is a ___; Look at this ____; Such a pretty ____; A nice ____; Wow, a _____), to promote word-referent mapping (Fennell & Waxman, 2010). Four different possible orderings of each of the eight tokens of each word type were generated, to form the audio sequences for each trial. For the low-variability condition, four different word types were recorded: the word /vidoʊ/ with a rise-fall pitch contour, /vidoʊ/ with a low-falling contour, /vɑdoʊ/ with a rise-fall, and /vɑdoʊ/ with a low-fall (Figure 4 shows pitch tracks, waveforms, and spectrograms for the four word type + pitch-contour combinations and Table A1 (appendix) reports descriptive statistics across several acoustic measures). Each child was habituated to two of the words, which composed one of the four possible minimal pairs: either pitch-contrasting (/vidoʊ/-risefall & /vidoʊ/-low fall; /vɑdoʊ/-risefall & /vɑdoʊ/-low fall); or vowel-contrasting (/vidoʊ/-risefall & /vɑdoʊ/-risefall; /vidoʊ/-low fall & /vɑdoʊ/-low fall).

The auditory stimuli for the high-variability condition were designed similarly, but pitch variability was incorporated into the vowel-contrast stimuli, while vowel variability was incorporated into the pitch-contrast stimuli. Note that because Switch experiments about word differentiation typically vary pitch contour within categories, and a segmental contrast between categories, the high-variability vowel-contrast experiment resembled most Switch studies. The pitch variability consisted of the original two pitch contours used in the original experiments (rise-fall and low fall) and two additional contours, a high fall (which began at a high pitch target rather than a low pitch target, as in the low fall) and a rise. Figure 5 depicts the four contours and Table A1 reports acoustic measurements. Within each

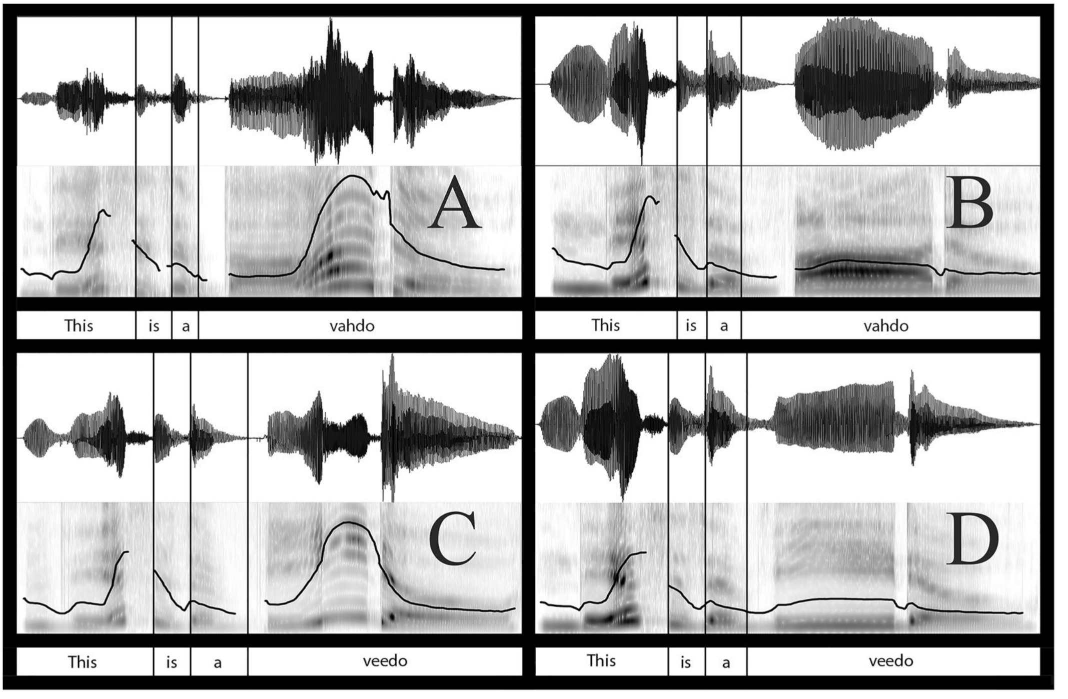


**FIGURE 4** Pitch contours in the low-variability condition of Experiment 3. Waveform, spectrogram, and pitch contour for "This is a vahdo" with a rise-fall contour (a) and low fall contour (b); and "This is a veedo" with a rise-fall contour (c) and low fall contour (d).

habituation and test trial in the vowel-contrast condition, all four pitch contours were presented twice each, for 8 total tokens.

Vowel variability included the original two vowel categories used in the original experiments (/i/, in /vidoʊ/— "veedo"—and /a/, in /vɑdoʊ/—"vahdo") and two additional vowels, /eɪ/, as in /veɪdoʊ/— "veydo"—and /u/, as in /vudoʊ/—"voodo." Within each habituation and test trial in the pitch-contrast condition, all 4 vowels were presented twice each, for 8 total tokens.

The auditory stimuli used in the test phase were identical to the training stimuli in each condition.

## 3 | RESULTS AND DISCUSSION

Figure 6 displays looking times in Same and Switch trials for children in the low-variability (left) and high-variability (right) conditions, learning words differing in their pitch contour or vowel. Raw looking times can sometimes exhibit a right-tailed distribution, so we checked for normality of residuals. Visual inspection of residuals and Shapiro-Wilk tests indicated that residuals (checked separately in Same and Switch trials) were normally distributed.

We first compared looking times in the post-test ("ball") trial to the Same and Switch trials to determine whether children were still on task by the end of the test phase. If so, they should

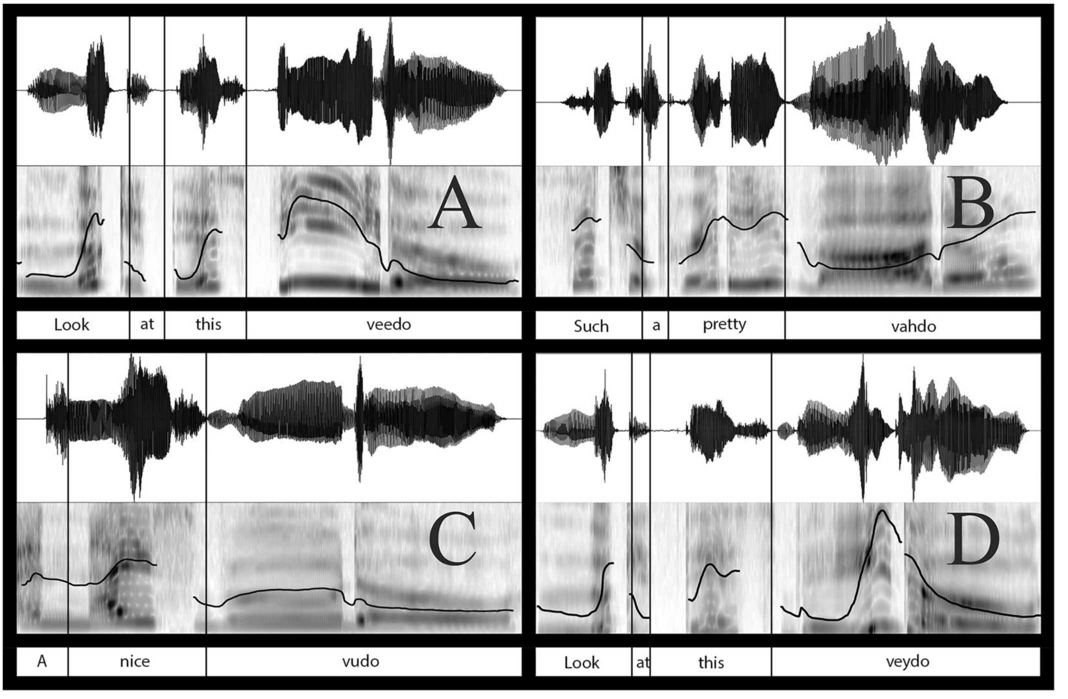


**FIGURE 5** Pitch contours in the high-variability condition of Experiment 3. Waveform, spectrogram, and pitch contour for the high-fall contour (a), rising contour (b), low-fall contour (c), and rise-fall contour (d), for the sentences "Look at the veedo," (a), "Such a pretty vahdo" (b), "A nice voodo" (c), and "Look at the veydo" (d).

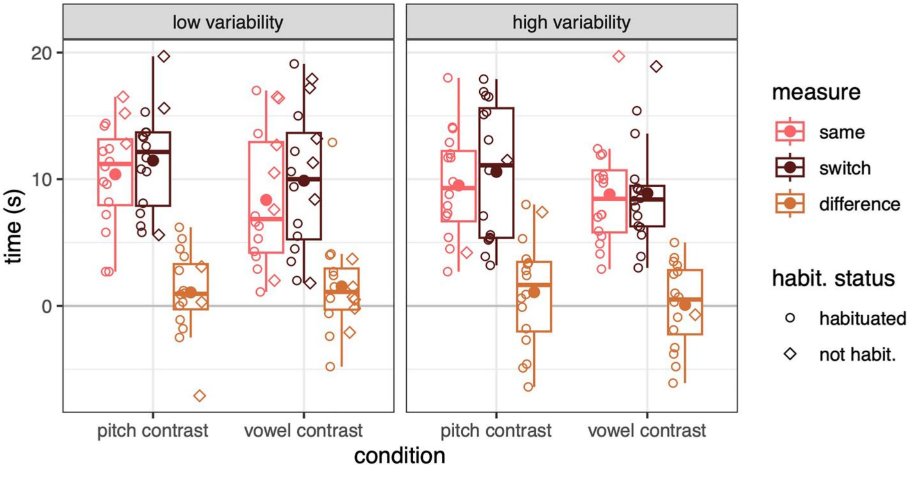


**FIGURE 6** 18-month-olds' looking times (in seconds) in the Switch procedure. Each point shows results from one child. Filled-in circles indicate condition means. The orange-colored points are by-child Switch-minus-Same difference scores. Switch > Same looking times mark recovery from habituation, indicating label-object learning. Learning was not significantly different across pitch- versus vowel-contrast conditions, though learning appears informally to not be evident for vowel distinctions (/i/-/a/) amid high pitch-contour variability.

"perk up" when the experimental trials gave way to a different word-object pair. Children's looking time in the post-test trial (*M* = 14.4s, *SD* = 4.8s) was significantly greater than their average looking time across the two Same trials (*M* = 9.3s, *SD* = 4.5s; $t(62) = 8.12$, $p < 0.001$) and across the two Switch trials (*M* = 10.2s, *SD* = 4.8s; $t(62) = 5.60$, $p < 0.001$). Infants therefore appeared to be on task during the test phase.

We next characterized children's patterns of habituation. Fifty-three of 64 children (83%) habituated by the final, 24$^{th}$ habituation trial. Children who habituated took on average 14.8 trials to habituate (*SD* = 5.1) and fixated the pictures for an average of 3.45 min (*SD* = 1.55 min) over the course of habituation. The proportion of habituators was somewhat different in the two variability conditions. For children with high-variability habituation, 30 of 32 children (94%) habituated. For children with low-variability habituation, only 23 of 32 (72%) habituated. Primary analyses (reported below) were conducted both with all 64 children included and with only the 53 who habituated (as inclusion of children who do not habituate can weaken novelty effects in habituation analyses; Oakes, 2010). Data patterns were similar in both cases.

In the test phase, children looked longer overall in Switch trials (*M* = 10.2s, *SD* = 4.8) than Same trials (*M* = 9.3s, *SD* = 4.5), an effect confirmed in an ANOVA conducted over all children (habituated or not), with Trial Type (Same vs. Switch) as a within-subjects factor, and Cue (pitch vs. vowel) and Variability (high vs. low) as between-subjects factors, $F(1,60) = 4.05$, $p = 0.049$, partial Eta squared = 0.063. There were no significant effects or interactions with Cue or Variability.

The pattern of results was similar in an ANOVA that included only the children who habituated (Switch *M* = 9.7s, SD = 4.5; Same *M* = 8.7s, SD = 4.0), $F(1,49) = 4.46$, $p = 0.04$, partial Eta squared = 0.083. There was a significant main effect of Cue, $F(1,49) = 4.36$, $p = 0.042$, partial Eta squared = 0.082, indicating higher overall looking in the Pitch condition (*M* = 10.2s, *SD* = 3.9s) than in the Vowel condition (*M* = 8.0s, *SD* = 3.5s). However, this overall difference did not interact with Trial Type or with Variability Condition, and there were no significant interactions or other main effects.

In the Switch habituation procedure, 18-month-olds learned minimal pair words, with learning holding over both pitch-contrasted and vowel-contrasted words, whether or not variability was present on the other dimension. The results provided no support for the hypothesis that under these teaching conditions 18-month-olds would consider variation in pitch contour differently from variation in vowel identity, nor did they indicate strong effects of variability on an irrelevant dimension. At 18 months, children seemed to be willing to learn a non-native pitch contrast whether or not vowels were varying. This surprising result adds to prior findings that, prior to 19 months, children are more willing to treat non-phonological variation as contrastive than older learners (Hay et al., 2015; Mulak et al., 2017; Singh et al., 2014). It may seem especially surprising that children learned the pitch contrast even when a *phonological* dimension—vowels—was varying. While within-word vowel-identity variation is disallowed in English (excepting rare cases like *goose/geese*), it occurs in many other languages. For example, in Spanish, grammatical-gender inflection causes the same adjective to surface with different vowels (such as *perfecto/perfecta*). Infants therefore need to accumulate linguistic evidence to determine whether/how within-word vowel-identity variation occurs in their language(s).

## 4 | GENERAL DISCUSSION

Three experiments assessed differentiation of sounds in newly learned words at three ages using two experimental methods: the language-guided looking procedure at 3–5 years and 24 months, and the Switch habituation procedure at 18 months. The first result was that 3–5-year-old English-learning children attended to vowel but not pitch changes, replicating prior findings with 30-month-olds and adults (Quam & Swingley, 2010); helping to paint a continuous developmental picture between these ages (Creel & Quam, 2015); and adding to the knowledge base about word learning beyond toddlerhood (e.g., McGregor et al., 2007).

The second result was that 24-month-old English learners' recognition of newly taught words was not hindered by changes to the pitch contour. This suggests that by 24 months, children resist attributing a lexical-contrast function to pitch-contour variation. However, in a parallel condition involving vowel distinctions, children were inconsistent, showing an effect of vowel mispronunciation after one set of training stimuli (with the rise-fall contour) but not the other (the low-fall contour). It is possible that there is a systematic explanation for this particular pattern of vowel learning versus non-learning. Perhaps children trained with the rise-fall contour received an overall attentional benefit from its higher emotional-prosodic arousal level (see, e.g., Russell et al., 2003) compared with the relatively more subdued-sounding low-fall contour (similar to an apparent attentional boost from a pitch change at 30 months observed by Quam & Swingley, 2010). However, this is highly speculative.

It is unclear why 24-month-olds did not consistently detect the highly salient vowel change from /i/ to /a/. Using similar experimental procedures, Singh et al. (2014) found that English-learning 24-month-olds responded to vowel but not tone changes; and Swingley (2007, preexposure condition) found that Dutch-learning 18-month-olds detected one-feature mispronunciations of initial consonants. Curtin et al. (2009) found 15-month-olds learned only one of three vowel contrasts in Switch (/i/ vs. /ɪ/). However, toddlers in that study seemed to rely on F1 (tongue height), a dimension on which /i/ and /ɑ/ (the vowels used here) are cleanly differentiated. Quam & Swingley, 2023, using the same language-guided looking training used here, reported that 24-month-olds did not detect changes from "deebo" to "teebo." Thus, inconsistent sensitivity to mispronunciations at 24 months in this paradigm is not limited to vowels.

One possibility is that the demands of the rich teaching context were not ideally tailored to the cognitive and linguistic abilities of 24-month-olds. However, 24-month-olds did recognize words robustly in CP trials (63%–69%). In addition, across two pilot experiments (reported in *Supplemental Materials*), 24-month-old children did not differentiate vowel-contrasted words in the Switch habituation procedure. Thus, task demands may not fully explain 24-month-olds' inconsistent detection of vowel changes. Another possibility is that perhaps insufficient variability was provided on non-phonological dimensions during the training, as words were pronounced by a single talker with a consistent pitch contour. Fourteen-month-old children differentiate similar-sounding words better in the Switch procedure when they are spoken by several voices (e.g., Quam et al., 2017; Rost & McMurray, 2009), apparently thanks to increased phonetic variability (Rost & McMurray, 2010). The inclusion of pitch variability in training in Experiment 3 did not meaningfully impact 18-month-olds' sensitivity to vowel changes. Still, future work could incorporate phonetic variability in training, and/or include a production task as part of the training, to support perceptual encoding at 24 months (Davis & Redford, 2023; Velleman & Vihman, 2002).

The final result was that 18-month-old English learners tested in the Switch procedure treated both pitch and vowel changes as relevant to word learning, regardless of the degree of irrelevant variability introduced during habituation. Of course, this interpretation of looking patterns presupposes that the Switch task is a good model for word learning, but it has shown sensitivity to native-language phonological distinctions (e.g., Dietrich et al., 2007) and convergent findings with studies using related methods (e.g., Swingley & van der Feest, 2019).

The null effect of variability at 18 months may be surprising—in particular, the successful learning of pitch-differentiated words even in the presence of phonologically relevant vowel variability. Effects of variability on language learning and processing are inconsistent across the literature (see Quam & Creel, 2021, for a review). In studies testing infants, children, and adults, variability that is irrelevant to the particular task has sometimes facilitated learning (e.g., for 14-month-olds' minimal-pair word learning in Switch; Rost & McMurray, 2009), impaired learning (e.g., for adults' learning and generalization of non-native speech-sound contrasts including lexical tones; Antoniou & Wong, 2016), or had null effects (e.g., for 7-month-olds' speech-sound discrimination; Quam et al., 2021). Our finding of null effects of variability on word learning at 18-month-olds adds to this literature.

The fact that 18-month-olds learned both pitch- and vowel-differentiated words is consistent with findings that 12-month-olds are sensitive to both indexical and linguistic changes to sounds (Mulak et al., 2017)—a broad sensitivity that putatively supports eventual normalization of indexical variation for speech-sound recognition and differentiation. The extant literature on sensitivity to intonational and tonal patterns in word learning by English-learning 17- to 19-month-olds is inconclusive. Thus, the sensitivity to intonational variation that we found at this age is consistent with some prior studies while conflicting with others. Table 7 contextualizes our findings in a sampling of the prior literature, indicating that children's success or failure to differentiate pitch-contrasted words cannot be easily tied to use of different pitch contours. There are two possible explanations for these apparent inconsistencies across studies, which are not mutually exclusive. One is that 17- to 19-month-old English learners are just at

**TABLE 7** One-year-olds' responses to pitch-differentiated words across a sampling of studies.

| Experimental paradigm | Pitch contours: English intonation: Rise-fall versus low fall | Pitch contours: Mandarin tones: Rise versus fall |
|---|---|---|
| Switch procedure | Experiment 3 reported here: 3 familiar-word pre-trials, typical phonetic variability contrasted with high variability<br>✓ at 18 months | Hay et al. (2015): No familiar-word pre-trials<br>✓ at 14 months<br>*Intermediate at 17 months*<br>✗ at 19 months |
| Language-guided looking | Quam and Swingley (2023, supplemental experiment S3): Exposure in narration and ostensive labeling<br>✗ at 19 months | Singh et al. (2014): Ostensive labeling<br>✓ at 18 months<br>✗ at 24 months |

*Note*: Check marks indicate successful word differentiation; X's indicate failure to differentiate words.

the cusp of ruling out pitch as lexically contrastive, so patterns across studies are more variable than they would be for older age groups whose interpretations have stabilized.

Subtle differences between experimental paradigms may also contribute to discrepancies in results. Eighteen-month-olds may be most willing to attend to pitch differences in words when the task difficulty is low (as in the Switch procedure). With this in mind, the methodological discontinuity (Creel & Quam, 2015) between Experiment 3 (Switch at 18 months) and Experiments 1 and 2 (language-guided looking with older children) must be acknowledged. By teaching one word and testing responses to mispronunciations, the language-guided looking procedure leaves open the (non)importance of pitch for children's and adults' interpretation. By contrast, the Switch procedure intentionally presents pitch-contrasted (or vowel-contrasted) words, so it is more 'heavy handed' in suggesting to children that pitch contours (or vowels) are relevant.

The present instantiation of Switch included 3 familiar-word trials (ball, car, and shoe) intended to help get across to children the referential nature of the task. This clarification of the referential nature of the task could have boosted infants' word learning overall (Fennell & Waxman, 2010). Perhaps infants at 18 months are open enough to attending to lexical pitch that when the referential nature of the task is clarified (and task difficulty is relatively low), they attend to both tonal and vowel differences.

The continuous developmental picture painted by this series of 3 experiments is that 18-month-olds are willing to treat both pitch and vowel changes as relevant to word recognition, regardless of variability in other phonetic features; 24-month-olds in the first stages of learning a new word recognize it despite changes to pitch contour and sometimes even to the vowel; and preschoolers show phonologically constrained responses, treating vowel changes as relevant and pitch changes as irrelevant. Thus, consistent native-language phonological distinction is only observed beyond 24 months. This result contrasts with findings of early precocity in infants' phonological learning (e.g., Polka & Werker, 1994; Werker & Tees, 1984), adding to emerging evidence of protracted development in children's ability to apply phonological knowledge to word learning and word recognition (Quam & Swingley, 2023; Singh & Chee, 2016).

## ACKNOWLEDGMENTS

We are tremendously grateful to the parents and children who participated in this study. We thank members of the Infant Language Center at the University of Pennsylvania who assisted with tasks including participant scheduling and testing or manuscript preparation, including Jane Park, Sara Clopton, Kristin Vindler Michaelson, Alba Tuninetti, Allison Britt, Rebecca Mead, Gabriela Garcia, Anna Runova, and Sophia Heiser. Additional students from the Child Language Learning Center at Portland State University assisted with manuscript preparation, including Natalie Robison, Genesis Ocegueda, Anna Zhen, Rachel Atkinson, Josie Johnson, Joey Lim, and Abigail Tolomei. We thank Melissa Redford and members of the University of Oregon linguistics department for helpful discussion. The authors declare no conflicts of interest with regard to the funding sources for this study. This manuscript has been made open access through support provided by Portland State University Library.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available for download at https://pdxscholar.library.pdx.edu/sphr_data/2/.

## ORCID

*Carolyn Quam* https://orcid.org/0000-0003-2607-8597

## REFERENCES


Antoniou, M., & Wong, P. C. M. (2016). Varying irrelevant phonetic features hinders learning of the feature being trained. *Journal of the Acoustical Society of America*, *139*(1), 271–278. https://doi.org/10.1121/1.4939736

Bion, R. A. H., Borovsky, A., & Fernald, A. (2013). Fast mapping, slow learning: Disambiguation of novel word-object mappings in relation to vocabulary learning at 18, 24, and 30 months. *Cognition*, *126*(1), 39–53. https://doi.org/10.1016/j.cognition.2012.08.008

Brooks, M. E., Kristensen, K., Benthem, K. J. van, Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., Mächler, M., & Bolker, B. M. (2017). GlmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*(2), 378–400. https://doi.org/10.32614/rj-2017-066

Burnham, D., Singh, L., Mattock, K., Woo, P. J., & Kalashnikova, M. (2018). Constraints on tone sensitivity in novel word learning by monolingual and bilingual infants: Tone properties are more influential than tone familiarity. *Frontiers in Psychology*, *8*, 2190. https://doi.org/10.3389/fpsyg.2017.02190

Chen, A., & Kager, R. (2016). Discrimination of lexical tones in the first year of life. *Infant and Child Development*, *25*(5), 426–439. https://doi.org/10.1002/icd.1944

Chen, A., Stevens, C. J., & Kager, R. (2017). Pitch perception in the first year of life, a comparison of lexical tones and musical pitch. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00297

Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies.

Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, *113*(4), 487–509. https://doi.org/10.1016/j.jecp.2012.07.007

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

Curtin, S., Fennell, C., & Escudero, P. (2009). Weighting of vowel cues explains patterns of word-object associative learning. *Developmental Science*, *12*(5), 725–731. https://doi.org/10.1111/j.1467-7687.2009.00814.x

Davis, M., & Redford, M. A. (2023). Learning and change in a dual lexicon model of speech production. *Frontiers in Human Neuroscience*, *17*. https://doi.org/10.3389/fnhum.2023.893785

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. https://doi.org/10.1073/pnas.0705270104

Edwards, J., Munson, B., & Beckman, M. E. (2011). Lexicon–phonology relationships and dynamics of early language development – a commentary on Stoel-Gammon's Relationships between lexical and phonological development in young children. *Journal of Child Language*, *38*(1), 35–40. https://doi.org/10.1017/S0305000910000450

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. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, *8*(2), 181–195. https://doi.org/10.1016/s0163-6383(85)80005-9

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papousek, U. Jurgens, & M. Papousek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches* (pp. 262–282). Cambridge University Press.

Fernald, A., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherese speech. *Infant Behavior and Development*, *10*(3), 279–293. https://doi.org/10.1016/0163-6383(87)90017-8

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. https://doi.org/10.1080/15475441.2014.895249

Gray, S., Lancaster, H., Alt, M., Hogan, T. P., Green, S., Levy, R., & Cowan, N. (2020). The structure of word learning in young school-age children. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1446–1466. https://doi.org/10.1044/2020_JSLHR-19-00186

Hay, J. F., Cannistraci, R. A., & Zhao, Q. (2019). Mapping non-native pitch contours to meaning: Perceptual and experiential factors. *Journal of Memory and Language*, *105*, 131–140. https://doi.org/10.1016/j.jml.2018.12.004

Hay, J. F., Graf Estes, K., Wang, T., & Saffran, J. R. (2015). From flexibility to constraint: The contrastive use of lexical tone in early word learning. *Child Development*, *86*(1), 10–22. https://doi.org/10.1111/cdev.12269

Hazan, V., & Barrett, S. (2000). The development of phonemic categorization in children aged 6–12. *Journal of Phonetics*, *28*(4), 377–396. https://doi.org/10.1006/jpho.2000.0121

Hofrichter, R., Siddiqui, H., Morrisey, M. N., & Rutherford, M. D. (2021). Early attention to animacy: Change-detection in 11-month-olds. *Evolutionary Psychology*, *19*(2), 147470492110282. https://doi.org/10.1177/14747049211028220

Hollich, G. (2005). Supercoder: A program for coding preferential looking.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy*, *13*(2), 128–157. https://doi.org/10.1080/15250000701795598

Humphrey, C., & Swingley, D. (2018). Regression analysis of proportion outcomes with random effects. *ArXiv:1805.08670 [Stat]* http://arxiv.org/abs/1805.08670

Jusczyk, P. W. (1992). Developing phonological categories from the speech signal. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 17–64). York Press, Inc.

Katz, G. S., Cohn, J. F., & Moore, C. A. (1996). A combination of vocal f0 dynamic and summary features discriminates between three pragmatic categories of infant-directed speech. *Child Development*, *67*(1), 205–217. https://doi.org/10.2307/1131696

Law, F., & Edwards, J. R. (2015). Effects of vocabulary size on online lexical processing by preschoolers. *Language Learning and Development*, *11*(4), 331–355. https://doi.org/10.1080/15475441.2014.961066

Law, F., Mahr, T., Schneeberg, A., & Edwards, J. (2017). Vocabulary size and auditory word recognition in preschool children. *Applied PsychoLinguistics*, *38*(1), 89–125. https://doi.org/10.1017/S0142716416000126

Liu, L., & Kager, R. (2014). Perception of tones by infants learning a non-tone language. *Cognition*, *133*(2), 385–394. https://doi.org/10.1016/j.cognition.2014.06.004

Mahr, T., & Edwards, J. (2018). Using language input and lexical processing to predict vocabulary size. *Developmental Science*, *21*(6), e12685. https://doi.org/10.1111/desc.12685

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. https://doi.org/10.1016/0010-0285(88)90017-5

McGregor, K. K., Sheng, L., & Ball, T. (2007). Complexities of expressive word learning over time. *Language, Speech, and Hearing Services in Schools*, *38*(4), 353–364. https://doi.org/10.1044/0161-1461(2007/037

Moore, D. S., Spence, M. J., & Katz, G. S. (1997). Six-month-olds' categorization of natural infant-directed utterances. *Developmental Psychology*, *33*(6), 980–989. https://doi.org/10.1037//0012-1649.33.6.980

Mulak, K. E., Bonn, C. D., Chládková, K., Aslin, R. N., & Escudero, P. (2017). Indexical and linguistic processing by 12-month-olds: Discrimination of speaker, accent and vowel differences. *PLoS One*, *12*(5), e0176762. https://doi.org/10.1371/journal.pone.0176762

Munro, N., Baker, E., Mcgregor, K., Docking, K., & Arciuli, J. (2012). Why word learning is not fast. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00041

Nittrouer, S., Miller, M. E., Crowther, C. S., & Manhart, M. J. (2000). The effect of segmental order on fricative labeling by children and adults. *Perception & Psychophysics*, *62*(2), 266–284. https://doi.org/10.3758/bf03205548

Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognitive Development*, *11*(3), 255–268. https://doi.org/10.1080/15248371003699977

Oakes, L. M., Sperka, D., DeBolt, M. C., & Cantrell, L. M. (2019). Habit2: A stand-alone software solution for presenting stimuli and recording infant looking times in order to study infant development. *Behavior Research Methods*, *51*(5), 1943–1952. https://doi.org/10.3758/s13428-019-01244-y

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (2011). A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting. *Journal of Phonetics*, *39*(2), 156–167. https://doi.org/10.1016/j.wocn.2011.01.002

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. https://doi.org/10.1037//0096-1523.20.2.421

Quam, C., Clough, L., Knight, S., & Gerken, L. (2021). Infants' discrimination of consonant contrasts in the presence and absence of talker variability. *Infancy: The Official Journal of the International Society on Infant Studies*, *26*(1), 1–17. https://doi.org/10.1111/infa.12371

Quam, C., & Creel, S. C. (2021). Impacts of acoustic-phonetic variability on perceptual development for spoken language: A review. *WIREs Cognitive Science*, *12*(5), e1558. https://doi.org/10.1002/wcs.1558

Quam, C., Knight, S., & Gerken, L. (2017). The distribution of talker variability impacts infants' word learning. *Journal of the Association for Laboratory Phonology*, *8*(1), 1–27. 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. (2012). Development in children's interpretation of pitch cues to emotions. *Child Development*, *83*(1), 236–250. https://doi.org/10.1111/j.1467-8624.2011.01700.x

Quam, C., & Swingley, D. (2014). Processing of lexical stress cues by young children. *Journal of Experimental Child Psychology*, *123*, 73–89. https://doi.org/10.1016/j.jecp.2014.01.010

Quam, C., & Swingley, D. (2023). A protracted developmental trajectory for English-learning children's detection of consonant mispronunciations in newly learned words. *Language Acquisition*, *30*(3–4), 256–276. https://doi.org/10.1080/10489223.2022.2069026

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. http://dx.doi.org.proxy.lib.pdx.edu/10.1016/j.cogpsych.2009.02.002

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

Russell, J. A., Bachorowski, J.-A., & Fernández-Dolz, J.-M. (2003). Facial and vocal expressions of emotions. *Annual Review of Psychology*, *54*(1), 329–349. https://doi.org/10.1146/annurev.psych.54.101601.145102

Ryalls, B. O., & Pisoni, D. B. (1997). The effect of talker variability on word recognition in preschool children. *Developmental Psychology*, *33*(3), 441–452. https://doi.org/10.1037/0012-1649.33.3.441

Shi, R., Santos, E., Gao, J., & Li, A. (2017). Perception of similar and dissimilar lexical tones by non-tone-learning infants. *Infancy*, *22*(6), 790–800. https://doi.org/10.1111/infa.12191

Singh, L., & Chee, M. (2016). Rise and fall: Effects of tone and intonation on spoken word recognition in early childhood. *Journal of Phonetics*, *55*, 109–118. https://doi.org/10.1016/j.wocn.2015.12.005

Singh, L., Fu, C. S. L., Seet, X. H., Tong, A. P. Y., Wang, J. L., & Best, C. T. (2018). Developmental change in tone perception in Mandarin monolingual, English monolingual, and Mandarin–English bilingual infants: Divergences between monolingual and bilingual learners. *Journal of Experimental Child Psychology*, *173*, 59–77. https://doi.org/10.1016/j.jecp.2018.03.012

Singh, L., Fu, C. S. L., Tay, Z. W., & Golinkoff, R. M. (2018). Novel word learning in bilingual and monolingual infants: Evidence for a bilingual advantage. *Child Development*, *89*(3), e183–e198. https://doi.org/10.1111/cdev.12747

Singh, L., Hui, T. J., Chan, C., & Golinkoff, R. M. (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science*, *17*(1), 94–109. https://doi.org/10.1111/desc.12097

Singh, L., & Tan, A. R. Y. (2021). Beyond perceptual narrowing: Monolingual and bilingual infants discriminate Hindi contrasts when learning words in the second year of life. *Developmental Psychology*, *57*(1), 19–32. https://doi.org/10.1037/dev0001137

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. https://doi.org/10.1038/41102

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., & van der Feest, S. V. H. (2019). A cross-linguistic examination of toddlers' interpretation of vowel duration. *Infancy*, *24*(3), 300–317. https://doi.org/10.1111/infa.12280

Tsao, F.-M. (2017). Perceptual improvement of lexical tones in infants: Effects of tone language experience. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00558

Velleman, S. L., & Vihman, M. M. (2002). Whole-word phonology and templates: Trap, bootstrap, or some of each? *Language, Speech, and Hearing Services in Schools*, *33*(1), 9–23. https://doi.org/10.1044/0161-1461(2002/002

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

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, *68*(2), 123–139. https://doi.org/10.1016/j.jml.2012.09.004

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. https://doi.org/10.1111/j.1467-7687.2008.00789.x

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Quam, C., & Swingley, D. (2024). Developmental change in English-learning children's interpretations of salient pitch contours in word learning. *Infancy*, *29*(3), 355–385. https://doi.org/10.1111/infa.12587

## APPENDIX

**TABLE A1** Acoustics of the auditory stimuli from Experiment 3. Means (and standard deviations) for duration in seconds, pitch maximum (max) in Hz, pitch mean in Hz, and formant frequencies for F1, F2, and F3 in Hz, are computed over all tokens of each word type and pitch contour from the low-variability and high-variability conditions. Rows in ***bold and italics*** report the grand means across all 8 tokens in each of 4 high-variability cases (for comparison with the 4 low-variability cases reported in the first 4 rows).

| Variability | Variability type | Word | Pitch contour | Duration (SD) | Pitch max (SD) | Time of max (SD) | Pitch mean (SD) | F1 (SD) | F2 (SD) | F3 (SD) |
|---|---|---|---|---|---|---|---|---|---|---|
| Low | N/A | Veedo | Rise-fall | 1.0973 (0.1024) | 685.21 (59.84) | 0.48184 (0.08818) | 328.40 (24.93) | 340.26 (25.29) | 2510.4 (56.0) | 3412.5 (102.4) |
| Low | N/A | Veedo | Low fall | 1.1204 (0.0972) | 274.65 (47.06) | 0.17332 (0.17289) | 216.61 (4.79) | 351.68 (40.11) | 2483.3 (40.4) | 3261.9 (76.2) |
| Low | N/A | Vahdo | Rise-fall | 1.1101 (0.1178) | 666.5 (57.52) | 0.50236 (0.07257) | 326.31 (17.79 | 857.83 (74.22) | 1359.6 (73.1) | 3020.5 (136.5) |
| Low | N/A | Vahdo | Low fall | 1.102 (0.1135) | 254.76 (22.86) | 0.28966 (0.20857) | 215.64 (4.78) | 878.71 (39.34) | 1389 (44) | 3054.7 (116.8) |
| High | Pitch | Veedo | Rise-fall | 1.391 (0.075) | 736.79 (26.1) | 0.55889 (0.09540) | 321.65 (16.64) | 423.03 (16.23) | 2526.2 (43.2) | 3309.1 (159.6) |
| High | Pitch | Veedo | Low fall | 1.272 (0.01) | 348.69 (47.22) | 0.28697 (0.37412) | 249.3 (5.02) | 349.06 (25.96) | 2274.9 (324.6) | 3128.8 (107.8) |
| High | Pitch | Veedo | High fall | 1.144 (0.071) | 555.45 (19.04) | 0.20854 (0.06507) | 325.6 (8.75) | 546.49 (15.11) | 2490.2 (137.8) | 3009.9 (324.8) |
| High | Pitch | Veedo | Rise | 0.909 (0.049) | 550.48 (8.31) | 0.87323 (0.06844) | 331.34 (11.46) | 409.81 (0.02) | 2530.5 (17.7) | 3407.3 (1.1) |

(Continues)

**TABLE A1** (Continued)

| Variability | Variability type | Word | Pitch contour | Duration (SD) | Pitch max (SD) | Time of max (SD) | Pitch mean (SD) | F1 (SD) | F2 (SD) | F3 (SD) |
|---|---|---|---|---|---|---|---|---|---|---|
| ***High*** | ***Pitch*** | ***Veedo*** | ***[Variable]*** | ***1.179 (0.196)*** | ***547.85 (148.42)*** | ***0.48191 (0.31658)*** | ***306.97 (36.79)*** | ***432.1 (77.72)*** | ***2455.5 (175.4)*** | ***3213.8 (218.2)*** |
| High | Pitch | Vahdo | Rise-fall | 1.072 (0.102) | 703.57 (23.07) | 0.44887 (0.07811) | 342.29 (6.79) | 989.12 (24.15) | 1355.3 (22.2) | 2966.7 (55.6) |
| High | Pitch | Vahdo | Low fall | 1.22 (0.052) | 285.31 (10.34) | 0.46305 (0.13311) | 234.64 (5.65) | 992.71 (65.91) | 1325 (54.7) | 3121 (14.1) |
| High | Pitch | Vahdo | High fall | 1.111 (0.086) | 498.25 (25.73) | 0.26952 (0.04924) | 310.68 (16.3) | 866.77 (34.19) | 1421.8 (41.5) | 2209.8 (576.3) |
| High | Pitch | Vahdo | Rise | 0.938 (0.004) | 475.41 (0.71) | 0.90057 (0.01723) | 292.26 (4.53) | 857.41 (75.54) | 1427.6 (84.2) | 3027.9 (100.2) |
| ***High*** | ***Pitch*** | ***Vahdo*** | ***[Variable]*** | ***1.085 (0.12)*** | ***490.63 (158.96)*** | ***0.52050 (0.25587)*** | ***294.97 (42.48)*** | ***926.5 (80.26)*** | ***1382.4 (62.7)*** | ***2831.3 (447.2)*** |
| High | Vowel | Veedo | Rise-fall | 1.365 (0.039) | 728.35 (13.62) | 0.56093 (0.10610) | 322.52 (15.76) | 422.04 (14.64) | 2532.5 (46.5) | 3304.3 (141.5) |
| High | Vowel | Vahdo | Rise-fall | 1.215 (0.012) | 674.34 (34.79) | 0.52522 (0.03038) | 319.08 (16.49) | 964.9 (1.07) | 1378.6 (30) | 2926.9 (31.7) |
| High | Vowel | Veydo | Rise-fall | 1.356 (0.029) | 651.66 (8.42) | 0.50392 (0.00010) | 309.79 (19.31) | 407.05 (4.33) | 1313.2 (66.7) | 2519.2 (14.8) |
| High | Vowel | Voodo | Rise-fall | 1.209 (0.011) | 677.11 (51.29) | 0.52092 (0.05984) | 293.4 (0.96) | 513.87 (20.68) | 2590.6 (219.5) | 3105.6 (294.5) |

**TABLE A1** (Continued)

| Variability | Variability type | Word | Pitch contour | Duration (SD) | Pitch max (SD) | Time of max (SD) | Pitch mean (SD) | F1 (SD) | F2 (SD) | F3 (SD) |
|---|---|---|---|---|---|---|---|---|---|---|
| ***High*** | ***Vowel*** | ***[Variable]*** | ***Rise-fall*** | ***1.287 (0.082)*** | ***682.86 (38.53)*** | ***0.46911 (0.18261)*** | ***311.2 (16.53)*** | ***576.97 (243.59)*** | ***1953.7 (656.7)*** | ***2964 (333.4)*** |
| High | Vowel | Veedo | Low fall | 1.265 (0.041) | 308.87 (2.25) | 0.18072 (0.35087) | 244.42 (3.89) | 367.71 (0.53) | 2537.4 (51.1) | 3256.8 (71.8) |
| High | Vowel | Vahdo | Low fall | 1.198 (0.11) | 280.94 (7.19) | 0.43373 (0.12106) | 234.32 (10.25) | 988.47 (79.27) | 1343.9 (124.8) | 3198.6 (103.7) |
| High | Vowel | Veydo | Low fall | 1.281 (0.175) | 308.26 (20.86) | 0.29618 (0.38518) | 234.44 (5.13) | 450.16 (1.5) | 1372.5 (22.4) | 2540.6 (13.3) |
| High | Vowel | Voodo | Low fall | 1.194 (0.025) | 289.71 (7.17) | 0.33863 (0.06934) | 229.06 (4.85) | 544.88 (37.11) | 2535.9 (105.2) | 2993.7 (175.2) |
| ***High*** | ***Vowel*** | ***[Variable]*** | ***Low fall*** | ***1.235 (0.091)*** | ***296.94 (15.59)*** | ***0.33490 (0.21436)*** | ***235.56 (7.72)*** | ***587.81 (258.35)*** | ***1947.4 (633.4)*** | ***2997.4 (311.6)*** |