Perceptual flexibility in word learning: Preschoolers learn words with speech sound variability

Conor I. Frye, Sarah C. Creel*

Department of Cognitive Science, UC San Diego, La Jolla, CA, USA

ARTICLE INFO

Keywords:
Language development
Word learning
Input variability
Speech perception
Consonant bias

ABSTRACT

Children’s language input is rife with acoustic variability. Much of this variability may facilitate learning by highlighting unvarying, criterial speech attributes. But in many cases, learners experience variation in those criterial attributes themselves, as when hearing speakers with different accents. How flexible are children in the face of this variability? The current study taught 3–5-year-olds new words containing speech-sound variability: a single picture might be labeled both deer and teev. After learning, children’s knowledge was tested by presenting two pictures and asking them to point to one. Picture-pointing accuracy and eye movements were tracked. While children pointed less accurately and looked less rapidly to dual-label than single-label words, they robustly exceeded chance. Performance was weaker when children learned two distinct labels, such as voyte and foth, for a single object. Findings suggest moderate learning even with speech-sound variability. One implication is that neural representations of speech contain rich gradient information.

1. Introduction

A longstanding question in language development, and language processing generally, is how learners cope with high amounts of variability in their language input. One answer to this question has been that variability facilitates learning by outlining the criterial, non-varying attributes of language structure. Over development, learners become neurally committed to those attributes (Kuhl, 2004; Zhang, Kuhl, Imada, Kotani, & Tohkura, 2005), facilitating processing of spoken input. Neural commitment to criterial variation is substantiated by electrocorticography data indicating specific brain responses to phonetic features (Mesgarani, Cheung, Johnson, & Chang, 2014). However, as described below, learners commonly experience variability in those criterial attributes themselves. The current study investigates the effects of speech sound variability when children are learning new words.

Multiple studies in numerous domains suggest that variability helps define category structure. Variability is thought to facilitate category learning by delineating the range over which a particular category can vary, thus focusing learners on criterial attributes rather than irrelevant ones (Kovack-Lesh & Oakes, 2007; Perry, Samuelsone, Malloy, & Schiffer, 2010; Rost & McMurray, 2009, 2010; see Apelbaum & McMurray, 2011, for a computational account). One key achievement is children’s implicit perceptual learning of the contrastive speech sound categories of their language. By representing a word as its constituent speech sounds, children can both disregard irrelevant variation (such as who is talking) and focus on subtle but diagnostic information (such as differences in vowel formants). If two word tokens differ acoustically but contain the same speech sounds, such as one’s mother saying “dog” and one’s father saying “dog,” they are the same word. If two tokens differ in their constituent speech sounds, even if they are acoustically quite similar, such as one’s mother saying “dog” and “dug,” they are different words. Word recognition in similarity-based computational models such as TRACE (McClelland & Elman, 1986) or the Neighborhood Activation Model (Luce & Pisoni, 1998) is guided through phonetic features and speech sound categories. While those models were not created to address the current question, their structure builds in the assumption that a given word has a unitary phonetic form. If there are two phonetic forms, they should in principle compete with each other for recognition.

Thus, there is some evidence that children deal with speech variability by focusing selectively on cues to invariant speech sound categories. That is, traversing a speech sound category boundary strongly indicates a change in word identity. However, a learner who always uses sound change, or even category structure, as an indication of word change is likely to run into difficulty. In particular, inflexible adherence to the principle that different speech sounds mean different words is likely to create problems for child learners, because they experience

* Corresponding author at: Department of Cognitive Science, UC San Diego, La Jolla, CA 92093-0515, USA.
E-mail address: screel@ucsd.edu (S.C. Creel).
https://doi.org/10.1016/j.bandl.2022.105078
Received 5 December 2020; Received in revised form 21 December 2021; Accepted 3 January 2022
Available online 22 January 2022
0093-934X/© 2022 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
speech sound variation in a variety of language-input scenarios. Some of these scenarios are inherent to the phonological rules of the language. In English, for example, children must learn relationships between affixes that undergo context-conditioned phonological alternation, such as the plural marker -s (the nouns trip, buzz, and board become plural by adding /s/, /əz/, and /z/ respectively) and the regular past tense marker -ed (the verbs trip, buzz, board add /t/, /d/, /əd/ respectively). Children may also experience input variation due to language variation across or even within speakers, such as hearing more than one dialect or accent (e.g. Floccia et al., 2012; see related work by Hudson Kam et al., 2012; see related work by Miller, 2013; Miller & Schmitt, 2012); and speech errors, especially from other children (Smith, Bird, Freilinger, Berenthal, & Bird, 1990). Any of these situations give the impression of variability in constituent speech sounds. Children cannot solve this comprehension problem by simply ignoring some speech sound boundaries, as they would lose the ability to differentiate many pairs of words. For example, suppose an English-learning child hears a Romance language-accented speaker who conflates /i/ and /ɪ/. If the child accommodates to that speaker by treating /i/ and /ɪ/ as instances of the same sound, many English word pairs would become indistinguishable (sheep vs. ship, cheap vs. chip, leak vs. lick, etc.). Thus, children need flexibility in how they deal with variation.

A different perspective than a focus on speech sounds to strongly demarcate boundaries between words, and one that allows for perceptual flexibility, is that learners store detailed representations of experienced word tokens, allowing access to gradient similarity. That is, word tokens vary continuously in their degree of similarity to each other; a 10-millisecond difference in voice onset time for two instances of /ti/ to minimal differences in phonemes such as /t/ vs. /d/ to more dissimilar phonemes /t/ vs. /v/ to dissimilar sound sequences /t/ vs. /v/ vs. /l/ or /l/ as well as auditory differences that are not directly related to word identity, like fundamental frequency, voice quality (modal, breathy, creaky), or absolute differences in formant frequency. Recurring patterns—such as words, phonological relationships amongst similar forms, and speech sounds themselves—would emerge from large amounts of input. Unlike models such as TRACE (McClelland & Elman, 1986) or the Neighborhood Activation Model (Luce & Pisoni, 1998), on the current account, the input directly activates stored instances in proportion to their similarity to the input, as well as allowing representations to contain additional acoustic information that may affect recognition (see Creel, 2014b; Creel et al., 2008; Creel & Tumlin, 2011). The resultant summed activity constitutes recognition, without having to posit explicit category or feature structure— or even unitary phonetic forms for words— as present in TRACE or NAM, and potentially allowing cooperation (in the sense of coactivation) rather than competition amongst similar but nonidentical forms. On this account, which borrows from exemplar accounts of language acquisition and representation (e.g. Goldinger, 1996, 1998; Johnson, 1997), it is more straightforward to construe how children might detect and learn similarity relations between word tokens that are not phonetically identical. Incomplete match would not block activation of a stored instance of “ship” in both native (/ʃIp/) and foreign (/ʃ Ip/) accents, or instances that contain a childlike speech error (/ʃIp/).

Limited evidence speaks to the question of children’s sensitivity and flexibility to speech sound variability in words, either in everyday learning or lab experiments. Most evidence comes from studies of infants’ recognition of familiar words and of newly learned words. Below we briefly review evidence for sensitivity and flexibility to input variability.

After that, we review related work suggesting that sensitivity and flexibility to input variability may depend on the type of segment that varies (consonants vs. vowels).

1.1. Speech sound specificity and variability in familiar word recognition

Early research by Swingley and Aslin (2000, 2002; see also Mani & Plunkett, 2007) found that infants aged 15–23 months detect subtle sound changes to familiar words. Specifically, when infants saw pictures of a ball and an apple, if they heard “ball”, they fixated the ball more rapidly than if they heard a mispronunciation of ball (“gall”). A less appreciated aspect of these findings is that even when hearing “gall,” children still looked more to the ball than to the other picture (in this example, the apple). This looking pattern implies that infants are sensitive to the similarity of the form “gall” to their representation of “ball.”

Recognition of partial similarity extends to cases where a novel object is present (19 months: White & Morgan, 2008; 2 to 2.5 years, Swingley, 2016; 3-5 years, Creel, 2012). Creel (2012) found that eye gaze data and points to pictures indicated that children most often treated a single-feature mispronunciation of a word (e.g., “fesh” for fish) as the word itself, even if a novel object was present as a plausible referent for the novel word. Preferences to treat mispronunciations as (nearly) equivalent to a known word are not the result one would expect if speech sound identity implied a new word. It is more consistent with a gradient similarity account of word form recognition.

Related work has considered infant and child word recognition in unfamiliar accents. Infants show difficulty in recognizing speech in unfamiliar accents (Mulk, Best, Tyler, Kitamura, & Irwin, 2013; Schmale, Cristia, Seidl, & Johnson, 2010; Schmale & Seidl, 2009; van Heugten & Johnson, 2014), with difficulty easing after two years of age (van Heugten & Johnson, 2015; van Heugten, Krieger, & Johnson, 2015). Further research (Bent, 2014; Bent & Atagi, 2015; Bent & Holt, 2018; Creel, Rojo, & Paullada, 2016; Holt & Bent, 2017; Nathan, Wells, & Donlan, 1998) suggests that children well past infancy, ages 3–7 years, still have difficulty recognizing words in unfamiliar accents. Nonetheless, 3- to 7-year-olds performed well above floor, suggesting that they can detect similarities to familiar words, consistent with a strong role for gradient similarity in word form recognition.

1.2. Speech sound variability in word learning

While a number of studies have assessed sensitivity to phonemic specificity in familiar words, relatively few have looked at effects of phonemic variability during the word-learning process itself. Muench and Creel (2013) found that learning two highly similar novel labels (such as “ziv” and “zev” both referring to the same picture) was relatively easy for adults. As the two labels for a given picture became increasingly phonetically dissimilar, learning accuracy declined, reaching its lowest level when learners had to learn two distinct words for each single object such as “deeye” and “vig”. This finding suggests that phonetic similarity—rather than generally good ability to learn dual labels for pictures— influenced adults’ learning.

A handful of studies have asked whether children as young as one year of age can recognize the equivalence between similar but non-identical sounds in certain contexts, with varying outcomes. Schmale and colleagues (Schmale, Holslich, & Seidl, 2011) find that 2-year-olds have some difficulty generalizing newly-learned words from one accent to another but with brief exposure to the unfamiliar accent, they succeed in generalizing recognition of the newly-learned word (Schmale, Cristia, & Seidl, 2012), as measured by visual fixations to pictures. Further studies have examined accent-like variability present during word learning. White, Peperkamp, Kirk, and Morgan (2008) found that children ages 8.5–12 months could learn similarity of phonologically alternating sound patterns (see also Chong & Sundara, 2015, with 18-month-olds). However, other evidence suggests that exposure to multiple forms presents difficulty: Floccia, Delle Luche, Durrant, Butler, and Goslin (2012) found that 20-month-olds exposed to two different dialects only represent the sound patterns of the socially-dominant dialect; 24-month-olds hearing multiple-accent input recognize words more slowly (Buckler, Oczak-Arsic, Siddiqui, & Johnson,
than their single-accent-learning peers; and young children learning a phonetically-variable plural morpheme learn it more slowly than children learning a consistent plural morpheme in another dialect of the same language (Miller & Schmitt, 2012).

Finally, in an artificial-accent word learning study, Creel (2014a) asked children aged 3–5 years to learn novel words that could vary in their constituent phonemes. For instance, a child might learn that one object was called both geef and gif (/gif/ and /g/), and another was called both keeb and kib (/kib/ and /kib/). Children who learned variable words were less accurate than children in a different experiment who learned words with only one label.

In sum, research on sound variability in children’s word learning is sparse, making it unclear how much learning difficulty children experience due to sound inconsistency in their input. Still, naturalistic evidence suggests that children as young as 3 years can produce variable phonetic forms depending on social context (such as play vs. discipline; Smith et al., 2007). While it is not necessarily the case that production and perception pattern identically, this production finding implies that variation is acquireable over the long term provided it is contextually conditioned.

1.3. Consonant biases in word learning

If phonemic variability slows word learning, is all phonemic variability equally detrimental? Several researchers have theorized that listeners and learners weight consonants more strongly in word recognition (Bonatti, Peña, Nespor, & Mehler, 2005; Flocia, Nazzi, Delle Luche, Poltrock, & Goslin, 2014; Nazzi, 2005), often referred to as a consonant bias. The consonant bias leads to the prediction that phonemic variability in word learning will be more disruptive when consonants vary than when vowels vary. That is, it may be easier to consider vowel-differing forms to be equivalent. Accounts of the consonant bias range from innate prespecification (Bonatti, Peña, Nespor, & Mehler, 2005) to consonants to be perceived more categorically than vowels (Flocia, Nazzi, Delle Luche, Poltrock, & Goslin, 2014), to lexically driven learning which indicates that consonants are more reliable cues to word identity (Keidel, Jenison, Klucher, & Seidenberg, 2007). Regardless of the account, evidence for a consonant bias is widespread in adults speaking European languages (English, French, Spanish, Dutch, Italian; though see Wiener & Turnbull, 2016 on Mandarin), in a wide range of tasks (segmentation: Bonatti et al., 2005; Toro, Shukla, Nespor, & Endress, 2008; though see Newport & Aslin, 2004; word learning: Creel, Aslin, & Tanenhaus, 2006; primed lexical decision: Delle Luche, Poltrock, Goslin, New, Flocia, & Nazzi, 2014; New et al., 2008; similar-word-generative: Cutler, Sébastien-Gallés, Soler-Vilageliu, & van Ooijen, 2000; van Ooijen, 1996).

Further, numerous developmental studies suggest that as early as 12 months, children are more sensitive to consonant differences in words than vowel differences (see Nazzi, Poltrock, & Von Holzen, 2016, for a review). Much of this evidence comes from French and Italian but some is from English, the language tested in the current work. The major exception in the largely European languages tested thus far is Danish, which has an unusually large vowel repertoire, and where infants show greater sensitivity to vowels than consonants (Elojen & Nazzi, 2016). Interestingly, the developmentally earliest evidence for this effect in English-speaking children is at about 30 months (Nazzi, Flocia, Moquet, & Butler, 2009), and studies of older children (Creel, 2012; Swingley, 2015) suggest roughly equivalent pronunciation sensitivity to vowel vs. consonant changes (fish as “fūsh” vs. puzzle as “buzzle”). Still, these children presumably eventually acquire a consonant bias, given that adult English speakers show consonant biases (Creel et al., 2006; Delle Luche et al., 2014; van Ooijen, 1996). In any case, the interest here is that learning words with sound inconsistency allows an additional test of consonant biases in English-learning children.

1.4. The current study

To expand understanding of children’s sensitivity to speech sound variability during word learning, we presented each child two types of to-be-learned words. In Experiment 1, children learned one set of words with pronunciation variability (dual-label words)—for example, the same pictured cartoon character was referred to as both deev and teev. Each child also learned a different set of words without pronunciation variability (single-label words). Words were learned two at a time, and the two words presented in a round of learning and testing were dissimilar from each other (dual-label example: deev/teev, fayfe/vayfe; single-label example: boove, sode, reducing the possibility that the two words would be phonologically confused with each other. For half of the children, the dual-label words were produced with two different consonants; for the rest, two different vowels. This design allowed assessment of consonant vs. vowel difficulty. Experiment 2 assessed whether children have additional difficulty learning pairs of unrelated words, for example fawfe and fosh, as labels for the same cartoon character.

This study expands on the question of form variability in word learning initially raised by Creel (2014). However, that study used only four word forms. Here we provide a stronger test using an improved design. First, we used a set (Table 1) of 32 (vs. four) to-be-learned words (distributed across children), which allowed for 16 dual-consonant words (4 consonant pairs, all of which differed in voicing: /d/, /b/, /s/, /z/, /t/, /v/) and 16 dual-vowel words (4 vowel pairs, all of which differed along the tense-lax distinction: /i/, /e/, /a/, /u/). This larger set of stimuli both increases generalizability and allows comparison of consonant variation vs. vowel variation to assess the presence of consonant bias. Second, we included a within-subjects baseline of word learning ability, by testing each child’s performance in learning two words that had only a single pronunciation.

If phoneme boundaries strongly indicate word-form boundaries, then variability in a word’s phonemes should make word learning quite difficult. That is, pointing accuracy and visual fixations to pictures should be very low for dual-label words. However, if children are simply learning forms on a continuum of graded similarity, then variability should present only mild difficulty.

A second set of predictions concerns the segment type (consonant, vowel) of the variable element. If children are more sensitive to consonants than vowels as a cue to word identity, as in many previous studies, then it should be easier for children to treat dual-vowel words equivalently than to treat dual-consonant words equivalently. That is, pointing and visual fixations to pictures should be greater for dual-vowel words than for dual-consonant words.

2. Experiment 1

2.1. Method

2.1.1. Participants

We tested 64 monolingual English-learning preschool-aged children (33 F; M = 4.33 years, SD = 0.70, range: 2.87–5.92). The study was novel word stimuli. Consonant voicing minimal pairs are left/right of each other, vowel minimal pairs are above/below each other. For each word, English pseudophonetic spelling appears on the left, International Phonetic Alphabet in slashes on the right.

| beesh /bʃ/ | peen /pɛn/ | vosh /vɒʃ/ | fosh /fɒʃ/ |
| bish /bɪʃ/ | pish /pɪʃ/ | vush /vʊʃ/ | fush /fʊʃ/ |
| boove /bʊv/ | poove /pʌv/ | vayfe /vɛɪf/ | fayfe /fɛɪf/ |
| buhv /bʊv/ | pahv /pæv/ | veɪh /vɛɪh/ | fɛh /fɛh/ |
| dayge /deɪdʒ/ | tayge /tædʒ/ | zodʒ /zɒdʒ / | sodʒ /sɒdʒ / |
| dedge /dɛdʒ/ | tʃedʒ /tʃdʒ/ | zodʒ /zɒdʒ/ | sodʒ /sɒdʒ/ |
| deev /dɪv/ | tɪev /tɪv/ | zʊf /zʊf/ | sʊf /sʊf/ |
| dihv /dɪv/ | tɪhv /tɪv/ | zʊf /zʊf/ | sʊf /sʊf/ |

Table 1
approved by the IRB at the researchers’ university. Prior to participation, each child’s parent consented to participation in writing, and each child verbally assented to participate. Additional children were tested but replaced for the following reasons: computer error led to loss of eye tracking data (5); lack of interest led child to end experiment early (2); lack of cooperativeness (1); glare interfered with eye tracking (1); exposure to a language besides English (1); incorrect contents of a sound file were discovered after data collection, necessitating replacement of 12 participants.

2.1.2. Stimuli
A set of 32 novel words (Table 1) were recorded by a male native speaker of English from the western United States.

2.1.3. Procedure
Children wore child-sized KidzGear headphones (www.gearforkidz.com). Each child learned the names for four novel cartoon creatures, two at a time. For one pair of creatures, the names were variable in one sound position (dual-label words). For example, one picture was labeled both deev (/diːv/) and teev (/tɪv/), while the second picture was labeled both fayfe (/faɪf/) and vayfe (/væɪf/). The two label pairs were dissimilar from each other, mismatching in all three consonant–vowel-consonant positions. For the other pair, the names were fixed, such as one picture being labeled boove (/buːv/) and the other sodge (/sɒdʒ/). For both dual-label and single-label pictures, the label(s) for one picture were dissimilar from the label(s) for the other. The order of the dual-label block and the single-label block was counterbalanced across participants. For half the participants, variable names varied in consonants; for the remainder, they varied in vowels. The large set of novel word stimuli also allowed the word pairs used to vary across children, so that a large number of dual-consontant sets (such as deev/teev, vayfe/fayfe; 8) and dual-vowel sets (8) were tested, as well as 32 single-label pairs. Using multiple word and sound pairs ensured that results were not limited to the idiosyncrasies of a small set of sounds.

Each child received two learning-testing cycles, presented via custom code written in MATLAB using Psychtoolbox-3 (Brainard, 1997; Pelli, 1997) and the Eyelink Toolbox (Cornelissen, Peters, & Palmer, 2002). In each cycle, during the learning phase, children saw each cartoon character appear onscreen eight times (16 trials total). On each trial, the character moved onto the screen using one of several types of motion (e. g., moving horizontally, bouncing), paused at screen center, and then the character was labeled twice, for example, “Deev. Deev.” with a 1000-millisecond (ms) pause between the two labelings. Even for dual-label blocks, the name heard on a single trial was the same (there were no “Deev. Teev.” trials). Four of the eight learning trials per picture presented one name, and the other four trials presented the other name. Trial order was randomized.

In the testing phase of each cycle (16 trials, 8 per object), the two characters appeared side-by-side, and the name of one of them was spoken as children’s eye movements were tracked by an Eyelink 1000 eye tracker (SR Research, Mississauga, Ontario, Canada) in remote mode. For dual-label blocks, half of trials for a target object used one label, half the other. Children were instructed to point to the one that was named, and an experimenter entered their pointing response via mouse-click on the selected picture. Thus, both pointing accuracy and visual fixations to pictures were assessed as converging measures of word recognition.

2.2. Results

2.2.1. Pointing accuracy
Children’s pointing accuracy (Fig. 1) was assessed via mixed-effects logistic regression. Logistic regression is designed for binomially distributed variables such as pointing accuracy, and mixed-effects models permit both participants and items random effects in a single model.

![Fig. 1. Experiment 1, pointing accuracy by condition. Bars are means with standard errors, with individual participant scores represented as superimposed points.](image-url)
SE = 0.21, z = 3.84, p < 0.001) and dual-label trials (B = 0.56, SE = 0.18, z = 3.09, p < 0.002) showed improvements with age.

2.2.2. Visual fixations

2.2.2.1. Data processing. Fixation data (Fig. 2) were processed as follows. First, we wanted to address a problem that occurs when obtaining pointing and fixation measures simultaneously: children’s pointing movements sometimes obscure the eye tracker camera, making it appear that they are not fixating anything at a time when they are most likely fixating one of the two pictures onscreen, because they are looking to the selected picture in order to guide their pointing movement. Accordingly, we processed the data trial-by-trial so that, when the gaze record reflected lack of knowledge of gaze location, it was filled in with the last known look location. Note that the fixation measure remains independent of the pointing accuracy measure, as we did not use any information about point location to fill in look location. Next, for occasional trials where children’s responses were made (and entered by the experimenter) earlier than 2000 ms after word onset—the end of our targeted time window—we extended the last looking position of the trial out to 2000 ms so that all trials contributed equally to each time bin for graphing and analysis. Finally, we then dropped trials with less than 50% of the time spent looking to the screen, as indicating a poor tracking quality or child inattention to the display (9% of trials). After these exclusions, 13.1% of data points were those filled in based on previous look location. We retained both correct and error trials in analysis, making our data comparable to younger age groups where looking time is the only measure.

We used a time window beginning at 200 ms and ending at 2000 ms. This window is a roughly 2-second-long interval following word onset, but with the start point shifted forward in time to 200 ms to account for the time needed to plan and execute an eye movement based on an external (verbal) signal (see Salverda, Kleimschmidt, & Tanenhaus, 2014). We chose this time window based on similarity to time windows used in studies with infants and toddlers (e.g., Swingley & Aslin, 2000, 2002).

2.2.2.2. Analysis. Although it is currently in vogue to analyze visual fixation data with a mixed-effects linear regression that includes individual trials as data points, we regard using individual trial data as inadvisable for the following reason: individual trial target fixations are not normally distributed, nor are they binomially distributed. Therefore, we instead attain approximately normal distributions by aggregating both over participants and over items. We then report both by-participants and by-items analyses. After aggregation, we created a target advantage score by subtracting non-target picture looks from target-picture looks to obtain a measure centered at 0 (that is, 0 = equivalent looks to target picture and to incorrect picture).

The predictors were the same as in the pointing accuracy analyses: Variability, Segment Type Varied, and their interaction. The effect of Variability was significant (F1(1,62) = 11.46, p < 0.001; F2(1,31) = 8.87, p = .006), with greater looking proportions for single-label word learning (boove, sode) than for dual-label word learning (deev/teev, vayle/fayle). There was no effect of Segment Type Varied (F1(1,62) = 0.02, p = .89; F2(1,31) = 0.73, p = .40), and there was no Variability × Segment Type Varied interaction (F1(1,62) = 0.26, p = .61; F2(1,31) = 0.04, p = .84), suggesting that decreases in looking for variable vowels and the variable consonants were roughly equivalent.

As with pointing accuracy data, we tested our a priori hypothesis of differences in consonant variability and vowel variability by comparing the dual-consonant and dual-vowel conditions directly, leaving out the single-label condition. There was no difference between the two (F1(1,62) = 0.14, p = .71; F2(1,31) = 0.77, p = .39), though each exceeded chance individually (dual consonants: t(31) = 3.98, p = .0004; t(2) = 7.46, p < .0001; dual vowels: t(31) = 3.88, p = .0005; t(2(31) = 4.91, p < .0001), as did the single-label condition (t(1,63) = 9.22, p = .0004; t(2(31) = 10.31, p < .0001).

Finally, we assessed effects of age on looks, by rerunning analyses with Age (mean-centered) and its interactions as additional predictors in the by-subjects analysis (age did not vary systematically over items). The effect of Variability was still present (F1(1,60) = 11.55, p = .001). There was also an effect of Age (F1(1,60) = 11.85, p < .0001), with an increase in target advantage as age increased, suggesting higher looking proportions at later ages. No interactions approached significance. The effect of Age held for both dual-label conditions (F1(1,62) = 4.78, p = .03) and single-label conditions (F1(1,62) = 13.30, p = .0005).

2.2.3. Bayes tests of the consonant bias

We found no evidence of consonant biases in our data. However, one might reasonably ask whether our design was insufficiently sensitive to detect the presence of a consonant bias. Bayes tests are designed to address this question, by taking into account the relative evidence for the null vs. alternative hypotheses. Note that the consonant bias provides a directional hypothesis that scores will be higher in the dual-vowel condition because vowel differences are thought to be easier to ignore. Accordingly, we computed Bayes factors for dual-consonant vs. dual-vowel comparisons, as described by Dienes (2014) and using the associated calculator. In each case, we assumed a normal distribution of the sampling mean, and entered the signed mean difference; standard error; and upper and lower bounds for the alternative hypothesis (uniform distribution assumed). The lower bound was always 0, because, on the alternative (consonant bias) hypothesis, the dual-vowel case should have equal or higher pointing accuracy or looking proportions. The upper bound, i.e., the largest reasonable expected difference between conditions, was set to the average score achieved in the single-label condition minus the chance score. (Setting the lower bound above 0, or using a higher upper bound such as 1.0 accuracy or 100% looks, both tend to increase evidence for the null, so our strategy is more conservative.)

The Bayes factor is a ratio, with the likelihood of the alternative hypothesis in the numerator and the likelihood of the null hypothesis in the denominator. A typical guideline for interpreting the Bayes factor (Dienes, 2014; Jeffreys, 1939/1961, as cited in Dienes, 2014) is that B$\geq$3 represents evidence for the alternative hypothesis, while the reciprocal boundary B$\leq$1/3, represents evidence for the null hypothesis. Numbers in between represent insensitivity of the test due to high levels of noise. For pointing accuracy (Mdiff = –0.082, SE = 0.061, bounds: 0-0.282), the Bayes factor was B = 0.12. For looking proportions by participants (Mdiff = –0.020, SE = 0.054, bounds: 0-0.240), the Bayes factor was B = 0.21. For looking proportions by items (Mdiff = –0.029, SE = 0.033, bounds: 0-0.240), the Bayes factor was B = 0.10. As all numbers are less than 1/3 (≈0.33), our null results can be interpreted as generally consistent with the null hypothesis that vowel variability is no
easier to ignore than consonant variability.

2.3. Discussion

Consonant and vowel variability did not prevent children from showing learning of word forms. Further, consonant voicing variability did not appear to be more disruptive than vowel tenseness variability, which does not support the presence of a consonant bias in English. Nonetheless, a potential explanation of children’s fairly good learning in the dual-label condition is that they are simply good at learning any two labels, regardless of the similarity between the two labels. If so, they should be quite good at learning phonologically unrelated labels for the same picture. This prediction was tested in Experiment 2.

3. Experiment 2

3.1. Method

3.1.1. Participants

Thirty-two preschool-aged children (age $M = 4.32$ years, $SD = 0.77$; 14 F) took part.

3.1.2. Stimuli and procedure

Recorded words were the same as in Experiment 1, but were differently assigned to pictures. Like Experiment 1, each child took part in two learning-testing cycles: single-label and dual-label. Unlike Experiment 1, the dual-label condition presented two disyllabic words for each object, such as one picture being labeled /vet/ and /fet/, and the other /dæd/ and /tæv/. Order of single-label and dual-label conditions was counterbalanced across children.

3.2. Results

3.2.1. Pointing accuracy

As in Experiment 1, pointing accuracy (Fig. 3) was assessed via mixed-effects logistic regression with maximal random effects for participants and items with predictor Variability (single-label, dual-labels). If it is more difficult to learn variable words for the same picture (/vet/ and /fet/), there should be a main effect of Variability. Variability was significant ($B = 0.34, SE = 0.15, z = 2.23, p = .03$), suggesting that learning dual labels ($M = 0.59$) is more difficult than learning single labels ($M = 0.68$). Individually, both dual labels ($B = 0.38, SE = 0.14, z = 2.76, p = .006$) and single labels ($B = 1.07, SE = 0.28, z = 3.87, p = .0001$) exceeded chance performance.

We reran the analysis with Age (centered) and Age × Variability as additional predictors, to gauge changes with age. Variability was again significant ($B = 0.37, SE = 0.16, z = 2.35, p = .02$). The effect of Age was significant ($B = 0.48, SE = 0.13, z = 3.68, p = .0002$), suggesting improved accuracy with Age. There was an interaction of Age × Variability ($B = 0.29, SE = 0.12, z = 2.40, p = .02$), resulting from improvement with Age in the single-label condition ($B = 0.74, SE = 0.22, z = 3.41, p = .0007$) but not the dual-label condition ($B = 0.19, SE = 0.12, z = 1.59, p = .11$). This lack of improvement in the dual-label condition may imply that learning dual labels is relatively difficult even for older children.

Of interest is whether dual-label learning is more difficult when the dual labels are more phonetically distant, as is the case in the current experiment. To assess this, we computed a model with Variability (single-label, dual-labels) and Experiment (1, 2) and their interaction as predictors. We collapsed over the consonant–vowel factor since it was not significant in Experiment 1. The single-label condition controls for possible differences in overall word-learning ability in the two samples. If there is such a difference, there should be an interaction of Variability and Experiment, with a larger drop in performance for dual labels in Experiment 2 than in Experiment 1. While the main effect of Variability was significant ($B = 0.36, SE = 0.10, z = 3.57, p = .0004$), indicating that learning dual labels is more difficult than learning single labels overall, the interaction with Experiment was not significant ($B = 0.02, SE = 0.10, z = 0.16, p = .87$), suggesting that difficulty learning dual labels was comparable between variable segments and variable words. Given our a priori hypothesis that dual-label learning would be especially difficult, we considered the dual-word condition alone. Here there were differences between Experiment 1 and Experiment 2 ($B = 0.31, SE = 0.13, z = 2.43, p = .02$). In the single-word condition, Experiment missed significance ($B = 0.33, SE = 0.18, z = 1.87, p = .06$). While this pattern of results should be interpreted with caution given the lack of interaction, it is consistent with dual-word learning being more difficult than dual-segment learning.

However, the main effect of Experiment was significant ($B = 0.33, SE = 0.12, z = 2.73, p = .0006$), with higher accuracy overall in Experiment 1 than Experiment 2. We explored the reasons behind the Experiment effect. One possibility is that children in Experiment 2 were simply less-skilled word learners overall. One way to address this “poor group of learners” explanation is to identify the better learners in this experiment based on their performance in the single-label condition control. If they show better learning of dual-word labels also, the implication is that poor dual-label performance here simply results from weaker overall word learning abilities. If they do not, it suggests that even if we had had a group of stronger word learners, dissimilar-label learning would still have been difficult. Accordingly, we looked at learners whose one-label performance was 0.75 or higher ($n = 13$). Their mean single-label performance was 0.93, but mean dual-label performance was 0.60, quite similar to overall dual-label performance of 0.59. A similar calculation for Experiment 1 data (Table 2) resulted in...
Table 2

| Segment Varied | Accuracy | Looks |
|----------------|----------|-------|
|                | Single label | Dual label | Single label | Dual label |
| Cs vary (Exp. 1) | 0.919 | 0.812 | 0.321 | 0.229 | 20 |
| Vs vary (Exp. 1) | 0.932 | 0.741 | 0.339 | 0.199 | 21 |
| All vary (Exp. 2) | 0.928 | 0.601 | 0.350 | 0.068 | 13 |

A comparable single-label accuracy but higher dual-label accuracy. An analysis limited to these high scorers did show the predicted interaction effect ($B = 0.27, SE = 0.14, z = 2.01, p = .04$), though the reader should bear in mind that this is an exploratory finding. In short, even the better word-learners in the current experiment, those comparable to Experiment 1, were not good at learning two distinct words as labels.

A second consideration is block order. Specifically, children may have found dual-label learning sufficiently confusing that they “tuned out,” or perhaps began to view the experimental speaker as unreliable (e.g., Koenig & Harris, 2005), and as a result performed worse in the single-label learning task. Children in Experiment 2 were numerically less accurate on single-label trials when they did these trials in the second block ($M = 0.64$), after completing a dual-label block, than when they completed single-label trials first ($M = 0.71$). (Dual-label performance was quite similar, at 0.58 and 0.59 respectively.) This performance mirrors a pattern in adult studies by Muench and Creel (2013, Experiment 1) where adults who learned two very-different words for the same picture exceeded chance ($t(1,30) = 8.35, p = .007$). The effect of Age was significant ($F(1,30) = 12.81, p = .001$), suggesting that looking performance improves with age. Individually, age was significant for single-label ($F(1,30) = 5.19, p = .03$) and dual-label ($F(1,30) = 4.24, p = .048$) conditions.

As with pointing accuracy, we compared findings to Experiment 1, with predictors Variability, Experiment, and their interaction. Variability was significant ($F(1,94) = 23.75, p < .0001; F(2,31) = 18.37, p = .0002$), suggesting less looking to correct pictures when pictures had dual labels. The main effect of Experiment was significant ($F(1,94) = 6.39, p = .01; F(2,31) = 10.69, p = .003$), with higher looks in Experiment 1 overall. However, the interaction of Variability and Experiment was not significant ($F(1,94) = 1.44, p = .23; F(2,31) = 1.58, p = .22$). Considering the dual-word condition alone, there were differences between Experiment 1 and Experiment 2 ($F(1,94) = 9.07, p = .003; F(2,31) = 19.14, p = .0001$), while in the single-word condition, there was no effect of Experiment ($F(1,94) = 1.81, p = .18; F(2,31) = 1.53, p = .23$). While this pattern of results should be interpreted with caution given the lack of interaction, it is consistent with dual-word learning being more difficult than dual-segment learning.

Again, one possible interpretation of the main effect but lack of interaction is that children in Experiment 2 were simply less good at word learning overall. If so, then comparing children who showed high accuracy on single-label trials in both experiments should show a similar pattern, with similarly high looking proportions for dual-label word learners here as for dual-label word learners in Experiment 1. However, as shown in Table 2 (right), the high-accuracy learners here showed weak looks on dual-label trials compared to high-accuracy learners in Experiment 1. As with the accuracy data, analyses limited to these high-accuracy learners showed a significant interaction with Experiment ($F(1,52) = 6.22, p = .02; F(2,31) could not be calculated because once many participants were dropped, not all words occurred in all cells). Keeping in mind that this is an exploratory analysis that bears confirmation in future research, this outcome suggests that even if children in Experiment 2 had been comparable word learners, they would have performed poorly on dual-label trials where the two labels were dissimilar words.

3.2.2. Visual fixations

Fixations (Fig. 4) were preprocessed as in Experiment 1. After exclusions as in Experiment 1, 13.9% of data points were those filled in based on previous look location. Data were analyzed in ANOVAs with the predictor Variability. Variability was significant ($F(1,31) = 12.64, p = .001; F(2,31) = 11.00, p = .002$), with greater looking proportions for single-label word learning (boove, sodge) than for dual-label word learning (vayfe/fosh, dedge/tiv). Individually, looks to single-labeled pictures exceeded chance ($t(31) = 4.35, p = .0001, t(31) = 4.91, p < .0001$), while looks to dual-labeled pictures did not ($t(31) = 1.49, p = .15, t(31) = 1.70, p = .10$). These findings support the accuracy data in suggesting that learning two words for an object is more difficult for children than learning one word for an object.

We reran the analysis with Age (centered) and Age × Variability as additional predictors, to gauge changes with age. Variability was again significant ($F(1,30) = 8.35, p = .007$). The effect of Age was significant ($F(1,30) = 12.81, p = .001$), suggesting that looking performance improves with age. Individually, age was significant for single-label ($F(1,30) = 5.19, p = .03$) and dual-label ($F(1,30) = 4.24, p = .048$) conditions.

As with pointing accuracy, we compared findings to Experiment 1, with predictors Variability, Experiment, and their interaction. Variability was significant ($F(1,94) = 23.75, p < .0001; F(2,31) = 18.37, p = .0002$), suggesting less looking to correct pictures when pictures had dual labels. The main effect of Experiment was significant ($F(1,94) = 6.39, p = .01; F(2,31) = 10.69, p = .003$), with higher looks in Experiment 1 overall. However, the interaction of Variability and Experiment was not significant ($F(1,94) = 1.44, p = .23; F(2,31) = 1.58, p = .22$). Considering the dual-word condition alone, there were differences between Experiment 1 and Experiment 2 ($F(1,94) = 9.07, p = .003; F(2,31) = 19.14, p = .0001$), while in the single-word condition, there was no effect of Experiment ($F(1,94) = 1.81, p = .18; F(2,31) = 1.53, p = .23$). While this pattern of results should be interpreted with caution given the lack of interaction, it is consistent with dual-word learning being more difficult than dual-segment learning.

Again, one possible interpretation of the main effect but lack of interaction is that children in Experiment 2 were simply less good at word learning overall. If so, then comparing children who showed high accuracy on single-label trials in both experiments should show a similar pattern, with similarly high looking proportions for dual-label word learners here as for dual-label word learners in Experiment 1. However, as shown in Table 2 (right), the high-accuracy learners here showed weak looks on dual-label trials compared to high-accuracy learners in Experiment 1. As with the accuracy data, analyses limited to these high-accuracy learners showed a significant interaction with Experiment ($F(1,52) = 6.22, p = .02; F(2,31) could not be calculated because once many participants were dropped, not all words occurred in all cells). Keeping in mind that this is an exploratory analysis that bears confirmation in future research, this outcome suggests that even if children in Experiment 2 had been comparable word learners, they would have performed poorly on dual-label trials where the two labels were dissimilar words.

3.3. Discussion

On the question of whether children are good at learning any two arbitrary pairs of labels as compared to a single label, the answer appears to be no. When learning distinct words, they are above chance in accuracy, but worse than when learning single labels, and looks to pictures do not exceed chance accuracy. Comparison to dual-segment learning gives a slightly different result.
learning in Experiment 1 is somewhat less clear. The full pattern of results is open to interpretation in that children in Experiment 2 performed less well overall than children in Experiment 1. On the one hand, this lower performance might result from testing a separate group of children who are less-good word learners. If so, the better single-label learners should have shown better dual-word learning, comparable to that in Experiment 1, but they did not, instead showing strong difficulty in dual-word learning. On the other hand, the lower learning accuracy in these children even in the single-label learning condition may indicate that dual-word learning is so difficult that it spills over into later learning tasks. Either of these alternative explanations suggests that learning two dissimilar words is more difficult than learning two similar-sounding words for the same referent.

4. General discussion

Our initial question was whether children had difficulty learning words when those words contained variable phonemes. The answer is partly yes and partly no. They experienced lower accuracy and smaller looking proportions relative to learning words with consistent phonemes. However, they still pointed and looked well above chance even when words’ labels were variable. By contrast, learning two distinct words for the same object appeared more difficult, with even-lower accuracy than for similar-sounding dual labels, and visual fixations that did not exceed chance. (Recall that this interpretation should be regarded with caution due to the lack of interaction with Experiment.) Further, we detected learning with a range of consonant voicing pairs and vowel tense/lax pairs, suggesting that ability to learn from variable input extends across particular speech sound contrasts and segment types. This relative ease of learning may explain how children can learn the equivalence of morphemes with related but non-identical phonological forms in particular phonological contexts due to dialect or accent variation, or due to language-specific phonological variation (see also White et al., 2008): these variants may be sufficiently phonetically similar that they naturally group together. At a broader level, these findings suggest that child learners may represent gradient similarity across word forms, rather than limiting themselves to phoneme category matching.

We also asked if there were differences in learning variable consonants vs. variable vowels. On theories that consonants are more central to word identity (Bonatti et al., 2005; Nazzi et al., 2016; Toro et al., 2008), dual-consonant words should be less like the same word than dual-vowel words, and thus dual-consonant words should be harder to learn. This is not what we found. Instead, vowel and consonant variability had effects that were not distinguishable from each other, and Bayes factors were fairly supportive of a lack of difference. These findings place conditions on theories holding that consonants are more central to word identity than vowels from early in development. We return to these below.

4.1. Discriminability: necessary, but neither sufficient nor obligatory for word learning

The finding that children can learn dual-pronunciation words fairly well implies that speech sound boundaries are not a strong barrier to word learning. Especially interesting is that children can learn across stop-consonant voicing boundaries, which have long been held up as an example of an early-maturing or early-learned speech sound perceptual boundary (Aslin, Pisoni, Hennessy, & Perrey, 1981; Elmas, Siqueland, Jusczyk, & Vigorito, 1971). This learning across stop-consonant boundaries raises questions about what the function of such early discrimination ability is, if not to distinguish words. Infant findings notwithstanding, findings from Treiman, Broderick, Tincoff, and Rodriguez (1998) on 5-year-olds’ word-similarity judgments suggest that voicing may be a less-salient consonant feature to young children than other consonantal articulatory features such as place or manner (see Jusczyk, Goodman, & Baumann, 1999, for infant sensitivity to onset-consonant manner similarity). The current results, along with the just-mentioned earlier research, suggest it may be fruitful to reframe the relationship between discrimination and word-form learning to encompass a full spectrum of variability.

Along the same lines, recent work suggests that, just as discriminability does not block learning of two forms as the same word, it also does not necessarily facilitate learning of two forms as different words. For example, Pajak, Creel, and Levy (2016) replicated a finding (Pajak & Levy, 2014) of perceptual discriminability advantages among certain listener groups for certain second-language sounds. For instance, speakers of a language with vowel length contrasts (Korean) showed perceptual advantages for consonant length contrasts. However, this perceptual advantage did not translate into good learning of words differentiated by those sounds: consonant length-contrasting words still showed very low accuracy in a word-learning task (Pajak et al., 2016). In a related nonspeech-auditory test, Creel (2016) reported that 4–5-year-olds who could distinguish rising vs. falling pitch contours in a same-different task were unable to learn distinct visual associations with the two pitch contours. That is, they could not associate rising pitch with one cartoon character and falling pitch with another cartoon character. As with Pajak et al. (2016), discrimination does not automatically confer learnability. Finally, children can readily make auditory distinctions between word tokens, such as vocal gender differences, yet do not necessarily treat these word tokens as different words (see Creel, 2014b, for further consideration of voice differences in words). In short, while discriminability is presumably necessary for word learning and word recognition, it is not sufficient.

A final point that bears on speech sound learning concerns age-related improvements. One might expect that if children’s speech sound resolution improves with age, the task of learning dual-label words would become more challenging with age. However, we found the opposite age pattern in our data: older children were better at learning dual-label words than were younger children. Further, at least one previous study suggests that adults are good at learning dual-label words (Muench & Creel, 2013). That research in combination with the current findings suggests that, if anything, learners improve over developmental time at learning words with inconsistent speech sounds. It is not clear whether a putative developmental improvement in inconsistent-sound learning might result from age-related improvements in learning processes, age-related increases in accent perception experience, or some other source, but it does suggest that dual-label learning is not driven by poor discriminability.

4.2. Implications for roles of consonants and vowels in word identity

We had predicted that if young English-learning children weigh consonants more strongly than vowels in word learning, similar to children in other languages (reviewed in Nazzi et al., 2016) and adults in a range of languages including English (Bonatti et al., 2005; Creel et al., 2006; Cutler et al., 2000; Deli Luche et al., 2014; Toro et al., 2008; Van Ootjen, 1996), they should have greater difficulty learning dual-consonant words than dual-vowel words. However, we did not find differences in dual-consonant and dual-vowel learning.

Why this finding differs from other results with young children is not immediately obvious, but there are several possibilities. First, it may reflect differences in test paradigms used. Yet the paradigms that have shown consonant biases vary so widely (see Introduction) that it seems unlikely that paradigm differences can explain effects. Second, it may reflect different patterns in the variety of languages tested in developmental studies (multiple European languages, but French and Italian predominates; see Nazzi et al., 2016). Variability across languages seems a more likely explanation, as previous studies in English have found mixed results early in development (consonant bias: Nazzi et al., 2009; no consonant bias: Creel, 2012; Floccia et al., 2014; Mani & Plunkett, 2007; Swingley, 2016). Still, studies of English-speaking adults suggest a
consonant bias by adulthood (Creel et al., 2006; Delle Luche et al., 2014; Van Ooijen, 1996). These findings together may imply that in English, differential vowel and consonant weightings diverge slowly over development compared to other languages tested. Returning to the varying accounts of the consonant bias, our work considered in the context of prior research would seem to favor the lexical hypothesis for consonant biases in English—perhaps consonant biases in English emerge only as vocabulary size increases from childhood to adulthood (though children of the ages tested already know many hundreds of words; Anglin, 1993). Other languages may show consonant biases earlier for other reasons, such as phonetic category structure differences (Floccia et al., 2014). That is, consonants may have better category separation and thus may be perceived more categorically than vowels.

4.3. Implications for neurobiological and neurocognitive bases of spoken language

As we noted earlier, a prominent theory in language development is that very young learners become neurally committed to their native language or languages (Kuhl, 2004). While we agree that neural commitment must take place, our research raises questions about the exact nature of that commitment. If learners truly committed to symbol-like phoneme representations that function to keep sounds and words distinct from each other, then rapid learning of gradually-similar forms should be quite difficult. If instead learners are neurally committing to a collection of experienced instances, over which flexible, gradient similarity computations can occur, then learning may occur flexibly with partial similarity. Of course, an additional interpretation of our findings is that children can learn gradually-similar forms, but this comes at a cost to discriminability. This interpretation makes the strong prediction that neural mismatch responses measured to a variable sound distinction—say, before and after learning that deev and teev refer to the same object—would show a loss of perceptual sensitivity between those two particular word forms. It remains to be seen which of these is the case.

4.4. Limitations

A limitation of the current work is that we only used a subset of all possible consonant and vowel contrasts. Consonants all differed in voicing, and vowels all differed in tenseness/laxness. This choice of stimuli may affect both sets of conclusions. Perhaps learning that two labels are equivalent is especially easy for just the consonant and vowel contrasts that we chose, but other contrasts (manner or place differences in consonants, height or backness differences in vowels) might generate greater difficulty. Another limitation is that the current work only considered consonant variation at syllable onset, so it is not clear that ease of learning phonemically variable words would generalize to coda consonant variation (for more on coda consonants, see Creel et al., 2006; Nazzi & Bertoncini, 2009; Redford & Diehl, 1999).

On the question of consonant vs. vowel comparisons, perhaps consonant voicing contributes less to word identity, or is more difficult to detect, than consonant manner or place. Had we used, say, manner difference, perhaps we would have seen a consonant bias (more errors for dual-consonant labels than dual-vowel labels). On the other hand, voicing contrasts, in particular those on stop consonants, are one of the most-studied cases of early acquisition of speech sound discrimination (Astin et al., 1981; Eimas et al., 1971). Our demonstration of reasonable success in assimilating variability over this boundary is particularly interesting because one would tend to assume that early-acquired boundaries would be especially stable.

Still, we may have inadvertently landed on a consonant feature that is especially slow to reach adulthood processing levels (see, e.g., Hazan & Barrett, 2000; McMurray, Danzel, Rigler, & Seedorff, 2018) or that is less discriminable to young children (Treiman et al., 1998), necessitating tests of additional consonant features. Similar questions obtain for vowel variability and syllable position. Perhaps tense/lax variants like those we used here are especially admissible to interchangeability, while other contrasts are more resistant. Again, assessing ease of learning variable sounds of different types will require additional studies. Notably, Nazzi et al. (2009), one of the few studies to show a consonant bias in English-learning children (30 months), used place contrasts rather than voicing contrasts. Still, if the consonant bias only appears in English-learning children for certain consonant and vowel contrasts but not others, it suggests limits on the extent of consonant biases in young learners of English.

5. Conclusion

Children ages 3–5 years experienced less difficulty learning words that contained variable phonemes (e.g., deev and teev) than they did when learning variable words (e.g., vayfe and fosh). This finding suggests that phoneme boundaries are not inviolable during word learning and is consistent with word representation in terms of gradient similarity. Despite hypotheses that consonants are more central to word identity than vowels, children had no more difficulty learning words with consonant variation than words with vowel variation. In concert with other findings, our work suggests that consonant biases may emerge more slowly in English relative to other languages, and that consonant biases seen in English-speaking adults may reflect slow reweighting processes shaped by lexical learning.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Thanks to Elizabeth Groves, Alicia Escobedo, and Creel lab research assistants for data collection. SCC and CIF were supported by grants from the National Science Foundation (BCS-1230003 and BCS-1057080).

References

Anglin, J. M. (1993). Vocabulary development: A morphological analysis. Monographs of the Society for Research in Child Development, 58(10), i. https://doi.org/10.2007/1166112

Aptelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. Cognitive Science, 35(6), 1165–1188. https://doi.org/10.1111/j.1551-6709.2011.01181.x

Astin, R. N., Pisoni, D., Hennessy, B. L., & Perrey, A. (1981). Discrimination of voice onset time by human infants: New findings and implications for the effects of early experience. Child Development, 52(6), 1135–1145.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language, 68(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bent, T. (2014). Children’s perception of foreign-accented words. Journal of Child Language, 41(6), 1334–1355. https://doi.org/10.1017/S0143038014000457

Bent, T., & Atagi, E. (2015). Children’s perception of nonnative-accented sentences in noise and quiet. The Journal of the Acoustical Society of America, 138(6), 3985–3993. https://doi.org/10.1121/1.4938225

Bent, T., & Holt, R. F. (2016). Shhh… I Need Quiet! Children’s Understanding of American, British, and Japanese-accented English Speakers. Language and Speech, 61(4), 657–673. https://doi.org/10.1177/0023830917754598

Bonatti, L. L., Feno, M., Nespor, M., & Mehler, J. (2005). Linguistic constraints on statistical computations. Psychological Science, 16(6), 451–459.

Brainard, D. H. (1997). The psychophysics toolbox. Spatial Vision, 10, 433–436.

Buckler, H., Ocak-Arsic, S., Siddiqui, N., & Johnson, E. K. (2017). Input matters: Speed of word recognition in 2-year-olds exposed to multiple accents. Journal of Experimental Child Psychology, 164, 87–100. https://doi.org/10.1016/j.jecp.2017.06.017

Chong, A. J., & Sundara, M. (2015). 18 Month-olds compensate for a phonological alternation. Proceedings of the 39th Boston University Conference on Language Development.
C.I. Frye and S.C. Creel

Brain and Language 226 (2022) 105078

McMurray, B., Danitz, A., Rigler, H., & Seedorff, M. (2018). Speech categorization develops slowly through adolescence. Developmental Psychology, 54(8), 1472–1491.

Mergner, T., Uyan, O., Chung, C., Johnson, K., & Chang, E. F. (2014). Phonetic feature encoding in human superior temporal gyrus. Science, 343(6174), 1006–1010. https://doi.org/10.1126/science.1245994

Miller, K., & Schmitt, C. (2012). Variable input and the acquisition of plural morphology. Language Acquisition, 19(2), 233–261. https://doi.org/10.1080/10493852.2012.685026

Miller, K. (2013). Acquisition of variable rules:s/ʃ-inetion in the speech of Chilean Spanish-speaking children and their caregivers. Language Variation and Change, 25(3), 311–340. https://doi.org/10.1017/S104938521300005X

Muench, K. L., & Creel, S. C. (2013). Gradient phonological inconsistency affects vocabulary learning. Journal of Experimental Psychology. Learning, Memory, and Cognition, 39(5), 1585–1596. https://doi.org/10.1037/a0034244

Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., & Irwin, J. R. (2013). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-regional accent. Child Development, 1–15. https://doi.org/10.1111/cdev.12087

Nathan, L., Wells, B., & Donlan, C. (1998). Children’s comprehension of unfamiliar accentual accents: A preliminary investigation. Journal of Child Language, 25(2), 343–363. https://doi.org/10.1017/s0305000997080344

Newport, E. L. (2005). Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. Cognition, 98(1), 13–30. https://doi.org/10.1016/j.cognition.2004.10.005

Nazzi, T., & Bertoncini, J. (2009). Phonetic specificity in early lexical acquisition: New evidence from consonant confusion. Journal of Language and Speech, 52(4), 463–480. https://doi.org/10.1017/S0305000909356854

Nazzi, T., Flocia, C., Moquet, B., & Butler, J. (2009). Bias for consonantal information over vocalic information in 30-month-olds: Cross-linguistic evidence from French and English. Journal of Experimental Child Psychology, 102(4), 522–537. https://doi.org/10.1016/j.jecp.2008.05.003

Nazzi, T., Poltrock, S., & Von Holzen, K. (2016). The developmental origins of the consonant bias in lexical processing. Current Directions in Psychological Science, 25(4), 417–422. https://doi.org/10.1177/0963721416649206

New, B., Araújo, V., & Nazzi, T. (2008). Differential processing of consonants and vowels in lexical access through reading. Psychological Science, 19(12), 1223–1227. https://doi.org/10.1111/j.1467-9280.2008.02228.x

Newport, E. L., & Aulin, R. N. (2004). Learning at a distance I. Statistical learning of non-adjacent dependencies. Cognitive Psychology, 48(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00028-2

Pajuk, B., & Levy, R. (2014). The role of abstraction in cross-linguistic speech perception. Journal of Phonetics, 46, 147–160. https://doi.org/10.1016/jAPH.2014.07.001

Pajuk, B., Creel, S. C., & Levy, R. (2016). Difficulty in learning similar-sounding words: A developmental stage or a general property of learning? Journal of Experimental Psychology. Learning, Memory, and Cognition, 42(9), 1377–1399. https://doi.org/10.1037/aph0000247

Pelli, D. G. (1997). The Video Toolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision, 10, 437–442.

Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. Psychological Science, 21(2), 1894–1902. https://doi.org/10.1177/0956797610368919

Reflifa, A., & Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVVC syllables. Journal of the Acoustical Society of America, 106(S1 Pt 1), 1555–65. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10489711.

Rist, 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

Rist, 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

Salverda, A. P., Kleinschmidt, D., & Tanenhaus, M. K. (2014). Immediate effects of anticipatory coarticulation in spoken-word recognition. Journal of Memory and Language, 71(1), 145–163. https://doi.org/10.1016/j.jml.2013.11.002

Schmale, R., Cristia, A., & Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. Developmental Science, 15(6), 732–738. https://doi.org/10.1111/j.1467-7687.2012.01175.x

Schmale, R., Cristia, A., Seidl, A., & Hanson, E. K. (2010). Developmental changes in infants’ ability to cope with dialect variation in word recognition. Infancy, 15(6), 650–662. https://doi.org/10.1111/j.1532-7950.2010.00023.x

Schmale, R., Helllich, G., & Seidl, A. (2011). Contending with foreign accent in early word learning. Journal of Child Language, 38(5), 1096–1108. https://doi.org/10.1017/S0305000910000619

Schmale, R., & Seidl, A. (2009). Accommodating variability in voice and foreign accent: Flexibility of early word representations. Developmental Science, 12(4), 583–601. https://doi.org/10.1111/j.1467-7687.2009.00809.x

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms project and its Nebraska replication. Journal of Speech and Hearing Disorders, 55(Supplement), S79–S78.

Smith, J., Durham, M., & Fortune, L. (2007). “Mam, my trousers is fa’ in doon!”: Community, caregiver, and child in the acquisition of variation in a Scottish dialect. Language Variation and Change, 19(1), 63–99. https://doi.org/10.1017/S1047785207070044

Swingley, D. (2016). Two-year-olds interpret novel phonological neighbors as familiar words. Developmental Psychology, 52(7), 1011–1023.
Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. Cognition, 76, 147–166.
Swingley, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. Psychological Science, 13(5), 480–484. https://doi.org/10.1111/1467-9280.00485
Toro, J. M., Shukla, M., Nespor, M., & Endress, A. D. (2008). The quest for generalizations over consonants: Asymmetries between consonants and vowels are not the by-product of acoustic differences. Perception & Psychophysics, 70(8), 1515–1525. https://doi.org/10.3758/PP
Treiman, R., Broderick, V., Tincoff, R., & Rodriguez, K. (1998). Children’s phonological awareness: Confusions between phonemes that differ only in voicing. Journal of Experimental Child Psychology, 68, 3–21.
van Heugten, M., & Johnson, E. K. (2014). Learning to contend with accents in infancy: Benefits of brief speaker exposure. Journal of Experimental Psychology: General, 143(1), 340–350. https://doi.org/10.1037/a0032192
van Heugten, M., & Johnson, E. K. (2015). Toddlers’ word recognition in an unfamiliar regional accent: the role of local sentence context and prior accent exposure. Language and Speech, 1–11. https://doi.org/10.1177/0023830915600471
van Heugten, M., Krieger, D. R., & Johnson, E. K. (2015). The developmental trajectory of toddlers’ comprehension of unfamiliar regional accents. Language Learning and Development, 11, 41–65. https://doi.org/10.1080/15475441.2013.879636
Van Ooijen, B. (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. Retrieved from Memory & Cognition, 24(5), 573–583 https://www.ncbi.nlm.nih.gov/pubmed/8870528.
White, K. S., & Morgan, J. L. (2008). Sub-segmental detail in early lexical representations. Journal of Memory and Language, 59(1), 114–132. https://doi.org/10.1016/j.jml.2008.03.001
White, K. S., Peperkamp, S., Kirk, C., & Morgan, J. L. (2008). Rapid acquisition of phonological alternations by infants. Cognition, 107(1), 238–265. https://doi.org/10.1016/j.cognition.2007.11.012
Wiener, S., & Turnbull, R. (2016). Constraints of tones, vowels and consonants on lexical selection in Mandarin Chinese. Language and Speech, 59(1), 59–82. https://doi.org/10.1177/0023830915578000
Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. Neuroimage, 26(3), 703–720. https://doi.org/10.1016/j.neuroimage.2005.02.040