Published in final edited form as:
Lab Phonol. 2017 ; 8(1): . doi:10.5334/labphon.25.

The Distribution of Talker Variability Impacts Infants' Word Learning

Carolyn Quam[1,2,*], **Sara Knight**[1,3], **LouAnn Gerken**[1]

[1]Department of Psychology, The University of Arizona, USA

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

[3]Department of Psychiatry, The University of Arizona, USA

Abstract

Infants struggle to apply earlier-demonstrated sound-discrimination abilities to later word learning, attending to non-constrastive acoustic dimensions (e.g., Hay et al., 2015), and not always to contrastive dimensions (e.g., Stager & Werker, 1997). One hint about the nature of infants' difficulties comes from the observation that input from multiple talkers can improve word learning (Rost & McMurray, 2009). This may be because, when a single talker says both of the to-be-learned words, consistent talker's-voice characteristics make the acoustics of the two words more overlapping (Apfelbaum & McMurray, 2011). Here, we test that notion. We taught 14-month-old infants two similar-sounding words in the Switch habituation paradigm. The same amount of overall talker variability was present as in prior multiple-talker experiments, but male and female talkers said different words, creating a gender-word correlation. Under an acoustic-similarity account, correlated talker gender should help to separate words acoustically and facilitate learning. Instead, we found that correlated talker gender impaired learning of word-object pairings compared with uncorrelated talker gender—even when gender-word pairings were always maintained in test—casting doubt on one account of the beneficial effects of talker variability. We discuss several alternate potential explanations for this effect.

Keywords

variability; word learning; infancy; phonetics; phonological development

1. Introduction

Infants learn an impressive amount about their native-language sound categories in the first year of life. This learning is reflected by infants' loss of discrimination for non-native sound contrasts that fall within native categories (Bosch & Sebastián-Gallés, 2003; Polka & Werker, 1994; Werker & Tees, 1984) vs. maintenance or even enhancement (Kuhl et al., 2006; Narayan, Werker, & Beddor, 2010) of discrimination for native-language sound contrasts. Infants' word-form recognition also becomes more robust over the first year to changes on phonologically irrelevant dimensions like talker's voice (Houston & Jusczyk,

[*]Corresponding author. 724 SW Harrison Street, Neuberger Hall 93, Portland, Oregon, 97201; cquam@pdx.edu; (503) 725-3558.

2000), pitch (Singh, White, & Morgan, 2008), and affect (Singh, Morgan, & White, 2004). These findings suggest that over the course of the first year, infants are pulling the relevant dimensions out of previously undimensionalized acoustic input (Jusczyk, 1993).

Despite this precocious development over the first year in infants' knowledge of native-language speech sounds, infants sometimes struggle to apply this knowledge when learning similar-sounding words. For example, infants at 14 months often fail to differentiate novel, similar-sounding words (e.g., (/bIn/-/dIn/), though they can discriminate the component sounds (/b/ vs. /d/; Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). At the same time, younger infants are more willing than older learners to attend to acoustic dimensions that are *not* contrastive in their language. For instance, before about 17 months, English-learning infants are willing to learn and differentiate words based on their pitch patterns (Singh, Hui, Chan, & Golinkoff, 2014; Hay, Graf Estes, Wang, & Saffran, 2015; see also Quam & Swingley, 2010). Thus, even after the early "perceptual reorganization" for native-language sound discrimination, infants are still learning to attend to contrastive dimensions and listen through non-contrastive dimensions in word learning.

Word learning at 14 months is typically assessed using the Switch habituation paradigm. In one version of this paradigm—the one used here—infants are habituated to two word-object pairings, and then their word learning is tested by presenting them with either intact word-object pairings ("Same" trials) or switched word-object pairings ("Switch" trials). One view of infants' failure to differentiate similar-sounding words like /bIn/ and /dIn/ in the Switch task is that infants fail to weigh phonologically relevant acoustic information more heavily than phonologically irrelevant information. On this view, talker's-voice characteristics like pitch and spectral information are fairly similar across single-talker exemplars, so two different words spoken by a single talker (where the difference is phonologically relevant) might be as similar acoustically as the same word spoken by two different talkers (where the difference is phonologically irrelevant). Thus, when infants must detect a switch in the word-object pairings, i.e., to reject /buk/ as an acceptable pronunciation of /puk/ or vice-versa, the *relevant* distinction between similar-sounding words, VOT, competes with the overall similarity between the words, and infants fail to detect the switch (e.g., Apfelbaum & McMurray, 2011).

Note that this acoustic-similarity account explains 14-month-olds' failure to differentiate similar-sounding words at an *acoustic-phonetic* level, arguing that infants are still learning which dimensions are relevant to word learning. However, factors beyond the acoustic-phonetic level likely also play a role in infants' ability to differentiate words. Minimal pairs are rare in children's early lexicons (Caselli et al., 1995), and introducing a sound contrast in minimal-pair words like /buk/ and /puk/ seems to make the sounds particularly difficult for infants to differentiate, compared with introducing them in more clearly differentiated words (Feldman, Myers, White, Griffiths, & Morgan, 2013; Thiessen, 2007; Swingley, 2009). Infants' limited prior experience with minimal pairs might make them less likely to accept words like /buk/ and /puk/ as two distinct words based on brief laboratory experience—even if they are paired with distinct objects. In some cases, pairing sounds with objects can help infants differentiate the sounds (Yeung & Werker, 2009). However, in a habituation paradigm like the one used here, the task of not only differentiating similar-sounding word-forms, but also encoding and remembering their assignments to objects, appears to increase the cognitive load on infants relative to purely discriminating word-forms (Stager & Werker, 1997; Werker & Curtin, 2005).

1.1. What can acoustic variability tell us about the nature of infants' word-form representations?

Recent evidence from Rost and McMurray (2009, 2010) indicates that increasing variability in talker's voice during the habituation phase of the Switch procedure leads to more robust differentiation of similar-sounding words (/buk/ and /puk/) at 14 months. Whereas infants fail to differentiate newly learned minimal-pair words when both are spoken by a single talker, they succeed when the words are spoken by 18 different talkers. An acoustic-similarity explanation (e.g., Apfelbaum & McMurray, 2011) would be that in the case of a single talker, talker characteristics are fairly stable (see, e.g., Johnson, Ladefoged, & Lindau, 1993; Heald & Nusbaum, 2015), so infants associate talker's-voice characteristics just as strongly with the words as VOT values, making the words more similar in multi-dimensional acoustic space and preventing infants from differentiating them. When talker variability is present, the two words are no longer made similar (pulled together in perceptual space) by sharing a single talker. Because the talker is varying, dimensions like pitch and formants vary much more across tokens, so infants associate them more weakly with the object. For the phonologically relevant dimension, VOT—which is more consistent within-word than between words—different acoustic values are associated with each object, and contribute strong associations. Thus, when the word switches from /buk/ to /puk/, infants are able to detect the switch.

A facilitative effect of talker variability for early word learning is consistent with evidence from many domains of category learning that high variability in training items facilitates learning (e.g., Posner & Keele, 1968). In language learning, children and adults learn non-adjacent dependencies (between A and B in AXB) only when there are many intervening items (X's; Gómez, 2002). Children with Specific Language Impairment learn real-language non-adjacencies (e.g., *is VERB-ing*, as in "is jumping") better when multiple word types are used in teaching (Plante et al., 2014).

Phonetic learning is also facilitated by training variability. Most relevant to the present study, typically developing children produce non-words with fewer errors when they had heard them spoken by multiple talkers (Richtsmeier, Gerken, Goffman, & Hogan, 2009). In perception of synthetic speech, listeners generalize better to novel stimuli when their training set is more varied (Greenspan, Nusbaum, & Pisoni, 1988). Teaching of second-language phonetic contrasts has been shown across many studies to benefit from high-variability training. Logan, Lively, and Pisoni (1991) developed a high phonetic variability training procedure for teaching Japanese-speaking participants the English /r/-/l/ contrast (see also Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997, and Iverson, Hazan, & Bannister, 2005). Variability was instantiated via different phonetic environments and five different talkers. Later work demonstrated that talker variability is necessary in training in order for participants to generalize learning of /r/-/l/ contrast to a new talker (Lively, Logan, & Pisoni, 1993). High-variability training has since been employed to teach English speakers to perceive Mandarin tones (Wang, Jongman, & Sereno, 1998) and to teach French speakers to perceive English vowels (Iverson, Pinet, & Evans, 2012).

As summarized above, there is a wealth of evidence that acoustic variability in training facilitates category learning across domains. Still, an alternative explanation for Rost and McMurray's (2009) findings of facilitative effects of talker variability on early word learning has also been proposed. Fennell and Waxman (2010) have suggested that hearing 18 talkers say both /buk/ and /puk/, consistently matched with the objects, might provide *social* evidence (rather than acoustic evidence) that /buk/ and /puk/ really are distinct (since 18 people think so). However, recent evidence indicates that even *within-speaker* acoustic variability can facilitate word learning if it is sufficiently large. Galle, Apfelbaum, & McMurray (2015) instantiated high acoustic variability within a single talker by instructing the talker to vary mean pitch, pitch contour, and duration of word tokens. Infants learned words under these conditions, suggesting that overall acoustic variability may be more important than talker variability per se.

It should be noted that even adults are still somewhat sensitive to talker variability. Whereas infants' word recognition can be fully disrupted by a change in talker (Houston & Jusczyk, 2000), adults' word recognition is still slower and less accurate when the talker changes from familiarization (Palmieri, Goldinger, & Pisoni, 1993; Goldinger, 1996). These findings have inspired exemplar models of the adult lexicon (Johnson, 1997, 2006; Pierrehumbert, 2001, 2002; Foulkes & Docherty, 2006).

1.2. Predictions when talker variability is correlated with novel words

The present study was designed to test the acoustic-similarity view that a single talker's voice makes minimal pairs too similar for infants to differentiate, an effect which is ameliorated by having multiple talkers. The study used the Switch habituation paradigm to teach 14-month-olds two word-object pairs spoken by 18 different talkers. Talker gender was either perfectly correlated with the words, by having male talkers say one word and female talkers say the other word, or was randomly varying with respect to the words (as in Rost & McMurray, 2009, 2010). To our knowledge, this is the first investigation of infants' processing of a talker gender as a correlated cue. In previous work, variability in talker's voice, for word learning (Rost & McMurray, 2009) or in intervening elements, for nonadjacent-dependency learning (Gómez, 2002) has been *unstructured* relative to the categories or pattern to be learned. When the distribution of variation has been bimodal, it has been on a phonological dimension (e.g., voicing; Maye, Werker, & Gerken, 2002; Maye, Weiss, & Aslin, 2008; or visual cues to phoneme identity; Teinonen, Aslin, Alku, & Csibra, 2008; see also Cristia, McGuire, Seidl, & Francis, 2011; though see Zhao, Al-Aidroos, & Turk-Browne, 2013, with adults).

There are two reasons to think that correlated talker gender might make infants' word learning more robust. First, in a different paradigm, both adults (Weiss, Gerfen, & Mitchel, 2009) and infants (Gonzales, Gómez, & Gerken, 2011) have been shown to segregate two differently structured artificial languages better when they are spoken by two different voices. Still, the task of segregating and learning artificial languages is quite different from the present word-learning task. Second, if words spoken by the same talker are pulled together in perceptual space (Apfelbaum & McMurray, 2011), then correlated talker gender should have the opposite effect and serve as an additional cue pulling words apart, thereby facilitating word differentiation. That is, words spoken by different talker genders are more acoustically distinct, differing in both VOT and pitch and spectral characteristics associated with male vs. female voices.

Evidence from early word learning suggests that English-learning infants at this age are fairly flexible about what they will consider relevant to word learning. They are willing to treat pitch contour as lexically contrastive (Singh, Hui, Chan, & Golinkoff, 2014; Frota, Butler, Correia, Severino, & Vigário, 2012; Hay, Graf Estes, Wang, & Saffran, 2015; see also Quam & Swingley, 2010). Until roughly 20 months of age, children are generally more willing than older learners to accept even non-word-like symbols, such as gestures, noise-maker sounds, and pictograms, as potential words (Namy, 2001; Namy & Waxman, 1998; Woodward & Hoyne, 1999). This greater flexibility in infancy and early toddlerhood about what can be relevant to word learning could lead infants in our study to differentiate the two words by talker gender, or perhaps by both talker gender and VOT, as might be predicted by a model of word learning in which infants employ both talker and VOT information to determine if two words are different (Apfelbaum & McMurray, 2011).

However, there are other reasons to predict that correlated talker gender might not facilitate word learning. In particular, it could be that correlated talker gender would operate very differently from linguistically relevant cues. It is possible that by 14 months of age, infants already have expectations that talker gender should *not* be relevant to word learning. This could cause the gender correlation to have no impact on learning. Or, it could lead infants to attempt to explain the surprising gender-word-object correlation, which could increase the task complexity and potentially impair word learning.

1.3. The present study

In both experiments, as in previous studies on word learning in the presence of talker variability (Rost & McMurray, 2009, 2010), 18 talkers produced words over the course of habituation. It is important to note that the correlation between talker gender and word was the primary difference between Experiment 1 and Experiment 2. The same set of 18 talkers was used, and the same number of tokens of each word was included.

To ensure that our correlated variability experiment was comparable—in all ways except the talker-gender correlation—with prior talker-variability studies, Experiment 1, which we refer to as the "Uncorrelated" case, aimed to replicate Rost and McMurray's (2009) uncorrelated-variability condition. In this condition, 18 talkers each said *both* words over the course of habituation, so variation in mean pitch and on spectral dimensions was *uncorrelated* with the words /buk/ and /puk/.

In Experiment 2, talker gender was perfectly correlated with the words: 9 male talkers said only one of the words (e.g., /buk/), while 9 female talkers said only the other word (/puk/), causing the words to differ on all the acoustic dimensions affected by talker gender (e.g., pitch mean and spectral characteristics). Note that as a result of the gender-word correlation, the number of talkers saying each word also differed from Experiment 1. Two conditions of

Experiment 2 were included to determine exactly what infants could learn from training with the talker-gender correlation. The habituation phase was identical in the two conditions, but they differed in the test phase. In one condition, the “Test of Learning,” we asked whether infants would show learning of the words when the gender-word pairings were maintained in the test phase. In the other condition, the “Test of Generalization,” we asked whether infants could learn words in the presence of correlated talker-gender information, and then *generalize* beyond the trained gender-word-object pairings, which were violated in half the test tokens.

2. Experiment 1

Experiment 1 was designed to verify whether the learning context in this study was comparable to previous word-learning studies that demonstrated facilitative effects of talker variability (Rost & McMurray, 2009, 2010). This was important to determine given some methodological differences between the present study and prior work (see Method below). Successful word learning here would provide a replication of Rost and McMurray’s (2009; 2010) talker-variability experiment in a slightly different training paradigm. This would allow us to then proceed to investigate word learning in the context of correlated talker variability, in Experiment 2.

2.1. Method

2.1.1. Participants—All infants included in the study were born at 37 weeks’ gestation or more, weighing at least 5 pounds, 8 ounces. Parents reported no history of speech or language problems in their nuclear family, nor significant foreign-language exposure—children had to hear English at least 70% of the time from birth. Infants were not given medication for an ear infection within one week of testing. Eighteen children (five girls) between the ages of 13 months, 23 days and 14 months, 28 days (mean age, 14 months, 9 days, *SD*, 9 days) were included in the analysis. Nine more infants participated but were excluded for fussiness (6) or equipment failure (3).

2.1.2. Experimental design—Our experimental design was modeled after Rost and McMurray (2009). The habituation phase of the experiment taught infants two different word-object pairs. During habituation, pairings of words with objects were always consistent. For example, /buk/ might always co-occur with a round metal toy with a plastic sail, while /puk/ might always co-occur with a juicer with a skirt around it (see Figure 2 for pictures of the objects). The word-object assignments were counterbalanced across participants.

Our design departed from Rost and McMurray’s in the following ways. First, each habituation and test trial contained a sequence of eight word-object presentations, rather than seven (Rost & McMurray, 2009). Using sequences of eight tokens enabled us to equate, within-trial, the number of tokens of each word and the number spoken by each gender. This was important because of our second change from Rost and McMurray’s design: within each trial, the eight word-object presentations pseudo-alternated between /buk/ and /puk/. Intermixing the two word-object pairs within each trial had two advantages: (1) It made the distribution of the two words more consistent over time. In prior work, up to 14 repetitions of a particular word-object pair could occur in sequence, e.g., when two /buk/ trials occurred in a row (Rost & McMurray, 2009), but here no more than two *tokens* of each word occurred in a row. (2) All test trials contained both words (rather than being, e.g., either a /buk/-“Same” trial or a /puk/-“Same” trial) making them more comparable to each other, and eliminating the need to consider test-word as a potential confounding factor in statistical analyses (as did Rost & McMurray, 2009). Across trials, objects pseudo-alternated in the following patterns: AABABBAB, ABAABABB, and ABABBABA, where A could be either /buk/ or /puk/—so that there were 6 different word-object presentation orders.

It is important to note that the within-trial pseudo-alternations may have changed the task complexity relative to Rost and McMurray’s (2009) design. A priori, it was not clear whether they would make the task more or less complex. On the one hand, since word-object pairs were alternating more frequently, this could have made word learning more difficult by increasing the complexity of each trial. On the other hand, alternations could have drawn infants’ attention more directly to the relevant contrast between words, improving word learning. However they impacted learning, they occurred in all three experiments, so could not explain any potential differences in performance across experiments.

Third, rather than static images, we used looming videos in which the object appeared at a small size, loomed to a large size, and then retracted. Moving objects are commonly used in the Switch paradigm (e.g., Fennell & Werker, 2003). However, it is possible that the use of looming videos here increased the task complexity relative to the task used by Rost and McMurray (2009, 2010). The presentation of each word token was temporally centered around the midpoint of the object presentation, so that it co-occurred with the point at which the object was largest on the screen. Each object was present for 1 2/3 seconds, followed by a 1/3-second blackscreen (so that the change in objects would not look abrupt), meaning each word-object presentation took 2 seconds, for a total trial length of 16 seconds.

2.1.3. Auditory stimuli—To generate habituation stimuli for use in both experiments, 18 different native English speakers—9 men and 9 women—produced the words /buk/ and /puk/ in an infant-directed register. All 18 were speakers of the west-coast dialect of English, and talkers were excluded if they appeared to have a different accent. In an informal judgment task conducted with 10 members of our laboratory, listeners were able to accurately identify the male talkers as male 94% of the time and female talkers as female 88% of the time.

Four tokens of each of the words for each talker were selected for their recording quality. In Experiment 1, each of the 18 habituation talkers said both words, so that talker gender was not linked with words. Only 2 tokens of each word from each talker were presented to each participant; the particular tokens used were counterbalanced across participants. Figure 1 confirms that in Experiment 1, one dimension related to talker’s voice, mean pitch across the entire word, was uncorrelated with the words /buk/ and /puk/.

Test stimuli were also the words /buk/ and /puk/, but were spoken by eight new talkers: four men and four women. Using new talkers circumvented issues about whether the habituation talkers were all equally represented in test. The test talkers were carefully selected so that they were as close as possible to average men's voices or average women's voices from habituation on all of the following acoustic dimensions: pitch mean, pitch maximum, standard deviation of pitch samples, and word duration. Identical test stimuli were used in Experiment 1 and Experiment 2. As in the Experiment 1 habituation, these consisted of only 2 tokens for each talker-word combination, since each talker said both words.

Table 1 reports pitch (Table 1a), formant (Table 1b), and VOT (Table 1c) measurements for each training and test talker, averaged over all tokens (4 /buk/ and 4 /puk/). Regarding VOT measurements, the mean VOT for /p/-initial stimuli across all talkers was 75.19 milliseconds, which is consistent with reports elsewhere (e.g., Zlatin & Koenigsknecht, 1976, reported mean VOT values for 20 adults of 78.99 milliseconds for "peas" and 83.77 milliseconds for "pear"). Hand-measurements of VOT found prevoicing in 16% of /b/-initial stimuli across habituation and test. Degree of pre-voicing varied across the sets of training vs. test talkers (see Table 1c), but the mean VOT for /b/-initial stimuli across all talkers was −2.23 milliseconds. Zlatin and Koenigsknecht (1976) reported mean VOT values of −23.17 milliseconds for "bees" and −12.02 milliseconds for "bear," thus, slightly more pre-voicing than we found overall. In an informal judgment task conducted with 10 members of our laboratory, listeners were able to accurately identify the word /buk/ with 97% accuracy and the word /puk/ with 98% accuracy.

2.1.4. Apparatus and procedure—Infants came to the lab with their parents. In a playroom, they were given time to settle in and adjust to the lab environment while the experimenter described the study and the procedure to the parent. When both the infant and the parent were ready to proceed, they were led to a separate, sound-attenuated room containing a large screen, a projector, two side speakers, and a video camera to record the infant's looking responses. The infant was seated on the parent's lap facing the screen. The experimenter sat in a separate control room viewing a video of the child's face.

Audiovisual stimuli were presented using the Habit software (Cohen, Atkinson, & Chaput, 2004). At the start of each trial, an attention-getting stimulus drew the child's gaze to the screen. This stimulus was a baby jumping in a crib, with a pacifier-squeaking sound. The background of the attention-getter was black, so that the contrast between the attention-getter and the trials (where objects were placed on white backgrounds) could be used in offline, reliability coding to identify the start and end of each trial. Once the infant had oriented to the attention-getter, the experimenter pressed a button to initiate the trial. Each trial was 16 seconds long and consisted of 8 word-object presentations. During each trial, the experimenter pressed another button to indicate the start and end of each look to the screen.

All looks to the screen in each trial were summed to calculate the total looking time for each trial. Looking times were then summed over the first three trials to set a baseline, pre-habituation looking level. The Habit program automatically computed the cumulative looking time across every subsequent sequence of three trials (using a moving window) and compared that cumulative looking time to the baseline level. Once the cumulative looking time across three consecutive trials had decreased to 50% or less of the baseline level, the infant was considered to have habituated, and the Habit program presented the test trials. Note that our habituation criterion (which has been recommended for infant habituation research; Oakes, 2010) differed slightly from Rost and McMurray, who used a 4-trial moving window.

Most children who completed the experiment habituated within the 27 training trials; however, 2/18 children in Experiment 1 and 4/36 children in Experiment 2 (11% in both cases) did not meet the habituation criterion by the end of the training phase. These children's responses in all other ways looked comparable to children who did habituate, and 27 habituation trials, each 16 seconds long, amounted to over 7 minutes of exposure to the two words—a sizeable amount for an infant experiment—so these children were retained in the analysis. However, data patterns were similar when they were excluded.

During test, infants saw two "Same" trials, which maintained the original word-object pairs, and two "Switch" trials, which violated them. The presentation order was counterbalanced across children (there were four possible orders of test trials: SWSW, WSWS, SWWS, WSSW; 'S' indicates a Same trial and 'W' indicates a Switch). A post-test, novel trial was also included at the very end of the experiment to assess whether, as one would expect, children perked up their attention when they saw entirely new objects paired with "buk" and "puk." Note that Rost and McMurray's (2009) experiment included a single Same trial, a single Switch trial (order also counterbalanced), and then the Novel trial. Thus, for closest comparison with the experiment we aimed to replicate, only the first two test trials were analyzed. Analyzing only the first two trials reduced the possibility that looking-time differences between Switch and Same trials might have been contaminated by increases in fatigue toward the end of the experiment, or by exposure to previous Switch trials (e.g., Same trials could be contaminated because increased looking times might bleed over into the next trial; or Switch trials could be contaminated because children are less surprised the second time they are exposed to the switched word-object pairing). However, data patterns were numerically similar when all four trials were included (see Table 2 for means and standard deviations; see the Results and discussion section below for further discussion).

The looking times recorded online by the live experimenter were used in the analysis. However, since these looking times were recorded under time pressure, reliability coding was conducted on 12 participant videos (of 54 total) to evaluate the accuracy of the online coding. The 12 participants coded offline were selected to include a representative proportion from each of the three online coders, and to sample equally from the three participant groups (Experiment 1 and the two conditions of Experiment 2). Reliability was evaluated by computing the Pearson's correlation between trial-by-trial total looking times in the offline coding file and in the online coding file. Correlations between offline and online coding were quite high for all 12 participant videos (mean correlation coefficient: 0.87; range: 0.61-0.99; all $p < .005$).

2.2. Results and discussion

Visual inspection of residuals and Shapiro-Wilk tests of normality across experiments and trial-types revealed that residuals were not normally distributed. Residuals were computed for the cross-experiment ANOVA reported below on pages 23-24, for each experiment—trial-type pair, and then entered into Shapiro-Wilk tests of normality. The following groups exhibited significant non-normality of residuals: Experiment 1 Switch trials (W = 0.893, *p*

< .05), Experiment 1 Novel trials (W = 0.841, *p* < .01), and Experiment 2 Novel trials (W = 0.897, *p* < .005). Upon visual inspection, all three of these trial types exhibited a left-tailed distribution. However, a square transform on raw looking times was not appropriate; while it normalized the three trial types with left skew, it introduced right skew in the three trial-types in which residuals were already normally distributed. In order to avoid introducing bias via a normalization process, we instead conducted both parametric and nonparametric tests.

We first conducted analyses of variance, which have been shown to be fairly robust to moderate non-normality (Glass et al. 1972, Harwell et al. 1992). We then investigated significant main effects and interactions with both t-tests (parametric) and exact Fisher-Pitman permutation tests (nonparametric). Permutation tests are appropriate when data violate the normality assumption of parametric tests (Legendre & Legendre, 1998). Briefly, the exact Fisher-Pitman permutation test involves computing the mean difference between two groups, then scrambling the assignment of data-points to groups and recomputing the mean difference for every possible assignment of data-points to groups. The p-value reflects the fraction of permutations in which the difference between the group means exceeded the true mean difference. Throughout the paper, we investigate the within-subjects factor trial type using one-tailed, paired tests, for both t-tests and Fisher-Pitman permutation tests. Use of one-tailed tests is justified because the Switch procedure provides clear directional predictions that looking times in the Novel trial will exceed looking in Switch and Same trials, and that Switch looking will exceed Same looking.

An analysis of variance (ANOVA) on raw looking times revealed a significant effect of Trial Type (first Same, first Switch, and Novel; F(2,34) = 16.6, *p* < .001). Planned comparisons (one-tailed, paired t-tests; and one-tailed, paired exact Fisher-Pitman permutation tests) revealed that looking times in the Novel trial exceeded looking times in both the Switch trials (paired t(17) = 3.16, *p* < .005; Fisher-Pitman *p* < .005) and Same trials (paired t(17) = 5.99, *p* < .001; Fisher-Pitman *p* < .001). Looking times were also significantly higher in Switch vs. Same trials (paired t(17) = 2.10, *p* < .05; Fisher-Pitman *p* < .05). Figure 3 and Table 2 report mean looking times, and Figure 4 displays a scatterplot of Switch- minus Same-trial looking times for each participant. Note that Table 2 also reports mean looking times averaged across both trials of each type (both Same or both Switch trials). The difference between Switch and Same looking times was still in the predicted direction (Switch > Same), but was numerically smaller. We believe that this reduction in effect size could be due to increased fatigue or fussiness at the end of the experiment and/or contamination from the previous Switch trial on later Same and Switch looking.

Replicating the findings of Rost & McMurray (2009), we found that 14-month-olds learned /buk/ vs. /puk/ in the presence of uncorrelated talker variability (18 talkers, 9 males and 9 females, saying both words). We found the pattern that is predicted when children successfully learn words in the Switch paradigm: children looked longer in response to Switch trials (where word-object pairings were reversed from habituation stimuli) vs. Same trials (where word-object pairings were maintained).

3. Experiment 2

Experiment 2 built on prior findings that uncorrelated talker variability enhances word learning at 14 months (Rost & McMurray, 2009, 2010), which we replicated in Experiment 1. In Experiment 2, we asked whether word learning would proceed any differently when talker gender was perfectly correlated with the words /buk/ and /puk/. To create the gender-word correlation, one word was spoken only by males and the other only by females.

Infants were tested in two conditions, which together paint a complete picture of precisely what infants are able to learn in the presence of correlated talker gender. In the "Test of Learning" condition, we assessed whether infants could learn words when *not* required to generalize beyond the trained gender-word pairings. Test tokens were selected so that they always *maintained* the gender-word pairings from the training. In the "Test of Generalization" condition, we assessed whether infants could both learn words and *generalize* beyond trained gender-word pairings in the test phase. Infants were habituated to a perfect correlation between talker genders and words, but the test stimuli contained uncorrelated talker variability. As a result, word tokens in test maintained the familiarized gender-word pairings half the time, and violated them half the time. Both maintained and violated gender pairings occurred within each trial.

3.1. Method

3.1.1. Participants—Eligibility criteria matched Experiment 1. Thirty-six children (19 girls) between the ages of 13 months, 22 days and 15 months, 6 days (mean age, 14 months, 12 days; *SD,* 10 days) were included in the analysis: 18 in the "Test of Learning" condition and 18 in the "Test of Generalization" condition. Twenty-eight more infants participated but were excluded for fussiness (23), experimenter error (1), equipment failure (2), biasing maternal behavior (1), or eyes not visible by the camera (1).

3.1.2. Experimental design—See Experiment 1.

3.1.3. Auditory stimuli—Training stimuli were taken from the same set as was used in Experiment 1. However, within each trial-order of Experiment 2, each word was spoken by only males or only females. To equate the number of overall tokens between experiments (36 of each word across all 27 potential habituation trials), all 4 tokens of each word from each talker were included (as opposed to 2 tokens per talker in Experiment 1). The gender-word associations were counterbalanced across infants, but were completely consistent within each infant's training. For example, /buk/ might have been spoken by only male talkers, while /puk/ was only spoken by female talkers. This created a perfect correlation between not only the words and objects, but between genders, words, and objects. Figure 5 depicts mean pitch across the entire word for each of these sets of habituation tokens. The left panel shows mean pitch when /buk/ was spoken by males and /puk/ was spoken by females. The right panel shows mean pitch when /buk/ was spoken by females and /puk/ was spoken by males. The figure indicates that in both cases, mean pitch differed between the words as a result of the gender correlation.

To verify that the distribution of talker gender caused the two words to differ on the predicted acoustic dimensions (pitch and formants), we computed acoustic measurements of each talker's productions of /buk/ and /puk/ (averaged across all 4 tokens). Each token was measured (via a Praat script; Boersma & Weenick, 2001) on several acoustic dimensions, which were then entered into separate analyses of variance (ANOVAs) as dependent variables. The ANOVAs were conducted on the full set of habituation stimuli (/buk/ tokens and /puk/ tokens for each of the 18 talkers) with predictors Talker Gender (M vs. F; between-subjects) and Word (/buk/ vs. /puk/; within-subjects), and dependent variables (in separate tests) pitch mean, pitch maximum, standard deviation of pitch samples, first formant (F1) frequency, second formant (F2) frequency, and third formant (F3) frequency (see Table 1, above, for means for each talker and each gender).

All six ANOVAs revealed significant effects of Talker Gender and none of Word. Talker Gender (M vs. F) impacted pitch means (F(1,16) = 21.82, *p* < .001; see Figure 5), pitch maxima (F(1,16) = 22.17, *p* < .001), and the standard deviation of pitch samples (F(1,16) = 15.16, *p* < .005), with females showing higher means and maxima and greater variability. The ratio of female/male f0 was 1.66 for habituation talkers and 1.69 for test talkers, both of which are quite close to the 1.70 found by Peterson and Barney (1952). Talker Gender also impacted F1 frequency (F(1,16) = 21.33, *p* < .001), F2 frequency (F(1,16) = 16.35, *p* = .001), and F3 frequency (F(1,16) = 119.8, *p* < .001). The female/male formant ratios for our habituation talkers (1.37, 1.69, and 1.30, for F1, F2, and F3, respectively) and the F1 ratio for test talkers (1.37) were higher than those found by Peterson and Barney (1.16, 1.19, and 1.16, for F1, F2, and F3, respectively), but were all in the correct direction. For test talkers, the F2 ratio (1.16) and F3 ratio (1.17) were comparable to Peterson and Barney's findings. Taken together, the ANOVAs indicate that males vs. females significantly differed on pitch and spectral dimensions.

We also evaluated whether the distribution of talker gender might have impacted the words' voice-onset times (VOT). We conducted an additional ANOVA on VOT, again with predictors Talker Gender and Word. The ANOVA revealed a significant main effect of Word, with *puk* exhibiting a higher mean VOT (*M* = 74.1 ms, *SD*=17.6 ms) than *buk* (*M* = 5.4 ms, *SD* = 27.5 ms; F(1,16) = 79.41, *p* < .001) but no main effect of or interaction with Talker Gender (Fs < 1, *p*'s > 0.5), indicating that male and female habituation talkers produced comparable VOTs (see Table 1 for means by talker). We also considered whether VOT variability might have differed between training sets. Both experiments contained the same set of word tokens across training sets, but divided the set of tokens up differently. We compared the VOT variability of 8 training sets, each composed of 36 tokens each. Four training sets were from Experiment 1: (1) *buk* spoken by all 18 talkers, tokens 1 & 2. (2) *puk* spoken by all 18 talkers, tokens 1 & 2. (3) *buk* spoken by all 18 talkers, tokens 3 & 4. (4) *puk* spoken by all 18 talkers, tokens 3 & 4. Four training sets were from Experiment 2: (1) *buk* spoken by 9 females, tokens 1-4 (2) *puk* spoken by 9 males, tokens 1-4. (3) *buk* spoken by 9 males, tokens 1-4. (4) *puk* spoken by 9 females, tokens 1-4. Six of the eight training sets failed the Shapiro-Wilk test for normality of VOT residuals. We therefore used the non-parametric Levene test to compare the variances of the 8 groups. This test found no significant differences between the variances of the 8 groups (L(7,280) = 1.15, *p* = .331).

Thus, we do not believe that difference in VOT variance could be a confound between experiments.

Test stimuli in the "Test of Generalization" condition were identical to Experiment 1. However, test stimuli in the "Test of Learning" condition were selected so that they included only word tokens that *matched* the gender-word pairings from familiarization. For example, if an infant was familiarized to male-"buk" and female-"puk," the test trials included only male-"buk" and female-"puk" tokens. To equate the amount of token variability, 4 tokens were included for each talker-word combination. Note that although gender-word pairings were always maintained, Switch trials still violated the familiarized pairings of word and object (as is always the case in the Switch procedure), while in "Same" trials these pairings were kept the same.

One possibility we needed to address is that in the case in which female talkers say "buk" and male talkers say "puk," the pairing of gender and word might have introduced cue conflict between the onset f0 cue to voicing (lower for /buk/ and higher for /puk/) and the typical f0 of the talker gender (higher for females and lower for males).[1] To investigate this possibility, we conducted an informal experiment with undergraduate research assistants from our laboratory (N=10) who were not familiar with our stimuli. We asked them to identify the word as "bewk" or "pewk," for each of the 208 tokens used in the infant experiment. Word-identification accuracy was high overall (M = 97.4%, SD = 3.4%). Visual inspection and Shapiro-Wilk tests of normality of residuals indicated that accuracy distributions for each gender-word pair were left-tailed because of ceiling effects (female-*buk:* W = .80, $p < .05$; female-*puk*: W = .78, $p < .01$; male-*buk*: W = .77, $p < .01$; male-*puk*: W = .87, $p = .089$). Thus, both parametric tests (paired t-tests) and non-parametric tests (paired approximate, or Monte-Carlo, permutation tests) were used for planned comparisons. In an analysis of variance with factors Word (/buk/ vs. /puk/), Talker Gender (male vs. female), and their interaction, there were no significant main effects. However, there was a trend (F(1,9) = 4.1, $p = .07$) for an effect of the interaction of Word and Talker Gender. Planned comparisons revealed that for the word /puk/, word-identification accuracy was higher for females (M = 98.7%, SD = 1.6%) than for males (M = 96.9%, SD = 2.9%; paired t(9) = 2.38, $p < .05$, Monte-Carlo $p = .08$). For the word /buk/, there was a numerical but non-significant effect in the other direction (females: M =95.8%, SD = 5.3%; males: M = 98.1%, SD = 2.4%; n.s.; Fisher-Pitman n.s.). Thus, there appears to be a tendency for adults to identify words slightly more accurately when the f0 cues to gender and voicing converge. While the effect size with adults is very modest (a roughly 2% difference in word-identification accuracy), it could be that ceiling effects reduced the effect size. It is also possible that infants could be more strongly affected by cue conflict than adults. Thus, in the Results and discussion section below, we investigate whether cue convergence/conflict might have impacted infants' word learning.

3.1.4. Apparatus and procedure—See Experiment 1.

[1] We thank anonymous reviewers for this suggestion.

3.2. Results and discussion

We first conducted an ANOVA to compare word learning across the two conditions of Experiment 2, including the between-subjects factor Condition ("Test of Learning," in which trained gender-word pairings were maintained in test, vs. "Test of Generalization," in which they were sometimes violated) and the within-subjects factor Trial Type (Same, Switch, and Novel). We also included the between-subjects factor, Gender-Word Pairing (conflicting or convergent), and the interactions of all three predictors. As discussed in the *Auditory stimuli* section, word learning in Experiment 2 might have been impaired by the pairing of women's voices with /buk/ and men's voices with /puk/. The higher fundamental frequency (f0) of women's voices could potentially be in conflict with the onset-f0 cue (a secondary cue to voicing), which is lower for voiced /b/. Likewise, pairing lower-f0 men with voiceless /p/—for which the onset-f0 cue is higher—could create similar conflict.

The ANOVA revealed a significant main effect of Trial Type (F(2,64) = 9.02, $p < .001$). Planned comparisons (one-tailed, paired t-tests and one-tailed, paired exact Fisher-Pitman permutation tests) revealed that looking times in the Novel trial exceeded looking times in both the Switch trials (paired t(35) = 3.90, $p < .001$; Fisher-Pitman $p < .001$) and Same trials (paired t(35) = 3.40, $p = .001$; Fisher-Pitman $p < .001$). However, looking times were not greater in Switch than Same trials (paired t(35) = −.20, n.s.; Fisher-Pitman $p = 0.58$). Table 2 and Figure 3, above, report mean looking times, and Figure 4, above, displays a scatterplot of switch- minus same-trial looking times for each participant. The ANOVA revealed no other significant main effects or interactions (all Fs <= 1.5, all $p > .2$), indicating that looking patterns were equivalent across conditions and across gender-word pairings. In other words, word learning was not impacted by whether gender-word pairings from habituation were maintained in test—nor was it impacted by whether gender-word pairings would have led to cue convergence or cue conflict between the onset f0 cue to voicing and the f0 tendency of male vs. female talkers.

Finally, we asked whether looking patterns differed significantly in Experiment 1 (where talker gender varied randomly across words) and Experiment 2 (where talker gender was correlated with training words). As the previous ANOVA revealed no differences in looking patterns across the two conditions of Experiment 2, we collapsed across conditions in this analysis. An ANOVA predicting raw looking times included factors Experiment (between-subjects; 1 vs. 2) and Trial Type (within-subjects; Same, Switch, Novel). There was again a significant main effect of Trial Type (F(2,104) = 26.08, $p < .001$), indicating that again Novel-trial looking exceeded Switch (one-tailed paired t(53) = 5.04, $p < .001$; Fisher-Pitman $p < .001$) and Same looking (one-tailed paired t(53) = 5.90, $p < .001$; Fisher-Pitman $p < .001$). Overall, Switch and Same looking did not differ. However, Trial Type significantly interacted with Experiment (F(2,104) = 3.97, $p < .05$; the main effect of Experiment was not significant). Comparisons of looking times within each trial type across experiments (unpaired and two-tailed, because there was no clear directional prediction) revealed that children in the two experiments did not differ in their Novel-trial looking (t(52) = −1.17, $p = .25$; Fisher-Pitman $p = .25$) or Switch looking (t(52) = −0.01, $p = .99$; Fisher-Pitman $p = .99$) but did differ significantly in their Same looking (t(52) = 2.02, $p < .05$; Fisher-Pitman $p < .05$). Thus, the cross-experiment comparison suggests that the equivalent looking in

Switch vs. Same trials in Experiment 2 was driven by longer looking in Same trials (relative to children in Experiment 1), rather than failure to "perk up" in Switch trials (see Table 2 and Figure 3, above, for means in each experiment). Longer looking in Same trials might reflect the greater complexity (or even the surprising nature) of the learning situation when talker gender was correlated with words in Experiment 2.

In the present experiment, when words were perfectly correlated with talker gender, children failed to learn the words. The fact that children failed to show evidence of word learning even in the "Test of Learning" condition indicates that it was *not* the case that children learned the three-way combination of gender, word, and object (e.g., male-"puk"-Juicer vs. female-"buk"-Sail), but did not recognize words when the gender-word pairings were violated. Instead, the talker-gender distribution seems to have interfered with learning completely, so that children did not learn words at all, in contrast to Experiment 1 and previous studies in which talker variability was uncorrelated with words (Rost & McMurray, 2009, 2010).

4. General discussion

In the Switch habituation paradigm, we first verified, in Experiment 1, that infants successfully learn similar-sounding words when all 18 talkers say both words during habituation, consistent with prior evidence that uncorrelated talker variability facilitates word learning at this age (Rost & McMurray, 2009; 2010). In Experiment 2, we next found that 14-month-old infants did *not* learn the similar-sounding words /buk/ and /puk/ when talker gender was perfectly correlated with the word-object pairs (i.e., 9 male talkers said only one word, and 9 female talkers said only the other word). This finding contrasts with prior work in which uncorrelated talker variability facilitated word learning (Rost & McMurray, 2009, 2010). Children failed to differentiate the test words regardless of whether the gender-word pairings from habituation were always maintained (Experiment 2 "Test of Learning") or sometimes violated (Experiment 2 "Test of Generalization"), indicating that it was not the case that infants learned the word-object pairs but treated talker gender as an important aspect of words' sounds. Instead, it appears that when talker gender was correlated with words, this additional complexity inhibited infants' word learning. A cross-experiment analysis indicated that Same-trial looking times were significantly longer in Experiment 2 than in Experiment 1, potentially reflecting the greater complexity of the training stimuli in Experiment 2, where talker-gender was correlated with words.

4.1. Implications of the results for our understanding of early word learning

There were two primary reasons to predict that correlated talker gender might *facilitate* infants' word learning—particularly in Experiment 2's "Test of Learning," when infants were not required to generalize beyond trained gender-word pairings. First, pairing different voices with different language input sets has been shown to help both adults (Weiss, Gerfen, & Mitchel, 2009) and infants (Gonzales, Gómez, & Gerken, 2011) to segregate the languages. Second, correlated cues generally facilitate language learning. Pairing language categories with objects can help learners differentiate sounds (Yeung & Werker, 2009), learn phonological rules (Frank, Slemmer, Marcus, & Johnson, 2009; Thiessen, 2012; van den

Bos, Christiansen, & Misyak, 2012), and segment words (at least in adulthood; Thiessen, 2010). Statistical information like transitional probabilities (Graf Estes, Evans, Alibali, & Saffran, 2007) or distributional cues (Lany & Saffran, 2010; Thiessen, 2007) can facilitate word learning. Christiansen, (2013a,b) has argued that cue redundancy is a crucial component of language.

Despite the above-mentioned reasons to predict that a correlated cue would facilitate learning, we found that infants failed to learn words in the presence of correlated talker gender, even when not forced to generalize beyond the trained gender-word pairings. One possible explanation concerns the task complexity. Since infants at this age are willing to consider non-phonological acoustic dimensions as potentially relevant to word learning, they may have detected that talker gender was correlated with the words, increasing the complexity of the learning task (Gerken, Dawson, Chatila, & Tenenbaum, 2015). On this view, infants needed to store two types of acoustic information with each word – talker gender and VOT. Research by Gerken et al. (2015) suggests that storing two stimulus dimensions is more demanding on infant pattern learning than storing a single dimension. Task demands are also known to impact word learning (Fennell & Werker, 2003, 2004; Yoshida, Fennell, Swingley, & Werker, 2009).

Rost and McMurray (2009) have called task complexity into question as a full explanation of 14-month-olds' word learning performance in the Switch paradigm, since increasing talker variability, which presumably increases task complexity, actually *improves* word learning. However, there is a large difference between uncorrelated talker variability—as in Rost and McMurray's work (2009; 2010) and our Experiment 1—and *correlated* talker variability, as in Experiment 2. Uncorrelated variability appears to make the learning task easier for infants by helping them to rule out irrelevant dimensions and focus on the relevant dimension(s) of contrast. Inversely, it is plausible that the consistent pairing of words and genders increases the task complexity, because it makes *more*—not fewer—kinds of information potentially relevant to the learning task.

Gender-word-object correlations are also surprising, occurring rarely in the real world. Infants' experience with language may already have indicated that gender is a non-contrastive dimension. Therefore, the surprising structure of the habituation stimuli could also have increased the task complexity. These possibilities are potentially consistent with task-complexity explanations for early failures to learn minimal-pair words (e.g., the PRIMIR model; Werker & Curtin, 2005). Recent work by Hay, Graf Estes, Wang, and Saffran (2015) indicates that 14-month-old English-learning infants will differentiate words using tones in the Switch paradigm, even though English does not use tone contrastively. This might suggest that infants at this age can learn words using dimensions that are contrastive cross-linguistically (even if not contrastive in their native language), but not dimensions that are never contrastive, like gender. However, Hay et al.'s study and the present study differ in other ways, such as the amount of overall acoustic variability, as Hay et al. used a single talker. It would be interesting in future research to directly compare infants' interpretations of talker gender to a dimension that is potentially contrastive.

Another possible explanation is that, as an additional cue for differentiating objects, talker gender may have introduced cue competition. For example, at an acoustic-phonetic level, the correlation of acoustic dimensions related to talker gender with VOT might have caused talker-gender-related dimensions (e.g., pitch) to compete for explanatory power with the VOT dimension. Under this account, it is somewhat surprising that infants did not recognize trained word-object pairings when gender correlations were maintained in the Experiment 2 “Test of Learning.” However, it may be that infants never fully resolved the cue competition.

Another possible explanation for infants’ failure to differentiate words in the presence of correlated talker gender is that they might have attributed acoustic differences between words to talker gender and thereby decreased the likelihood of two distinct words. That is, they might have concluded that they were hearing a *single word* that men pronounce one way and women another. This hypothesis could have “explain[ed] away” (Pearl, 1988; Dawson & Gerken, 2011) the VOT contrast,[2] causing infants to treat /buk/ and /puk/ as instances of the same word. This account might seem unlikely given that the words were being associated with distinct objects. However, associating words with objects increases the task complexity (Werker & Curtin, 2005). In addition, infants’ experience with minimal pairs is limited (Caselli et al., 1995), and minimal pairs are particularly difficult to learn (Feldman, Myers, White, Griffiths, & Morgan, 2013; Thiessen, 2007).

We cannot currently know which, if any, of the abovementioned accounts of infants’ failure in Experiment 2 might be correct. We *can* say that their failure in Experiment 2’s “Test of Learning” and success in Experiment 1 would not have been predicted by a model of word learning focused solely on the acoustic similarity between words (Apfelbaum and McMurray, 2011). Such a model attributes the failure of infants to differentiate minimal word pairs in a single-talker condition to the voice characteristics of the single talker making the words more similar. Under that account, the gender-word-object pairs in Experiment 2 should have been maximally different from each other, with pitch information *reinforcing* VOT information to highlight word differences. This should have led infants to show robust differentiation of the two gender-word-object pairs in the Experiment 2 “Test of Learning,” when they were not required to generalize beyond the trained gender correlation. The fact that the correlated acoustic information did not behave in the predicted way suggests that infants may have been using both VOT and talker voice, but crucially, treating the two as distinct cues that were competing for explanatory power.

4.2. Alternative explanations

One possible alternative explanation for the present results concerns the amount of within-word variability. To instantiate correlated talker gender, each word had to be spoken by members of only one gender. This meant that 9 females or 9 males (gender-word pairings were counterbalanced across infants) said “buk” in the correlated-gender experiments, whereas 9 females *and 9 males* said “buk” in the uncorrelated-gender experiment. We attempted to equate the variability across the correlated- and uncorrelated-gender cases by habituating infants to the same number of *tokens* of each word in the two cases (36) and the

[2] Thanks to Colin Dawson for this suggestion.

same overall number of talkers (9 males and 9 females). However, it is possible that either male+female variability is necessary within-word to facilitate word learning at 14 months, or that there is a threshold somewhere between 9 and 18 talkers that is "enough" variability.[3] These explanations seem unlikely given that Galle, Apfelbaum, & McMurray (2015) have demonstrated successful word learning at this age when sufficient acoustic variability was instantiated within a single talker. Still, it is possible that 9 male talkers and 9 female talkers did not reach some more abstract "acoustic variability" threshold. These possibilities should be addressed in future work that might manipulate the within-word variability in the habituation set, e.g., by comparing different numbers of male vs. female talkers (e.g., 1 male vs. 1 female; 18 males vs. 18 females). It is not clear whether more within-word variability is better for learning, or whether in the present study it contributed to increased task complexity and inhibited learning.

Another potential alternative explanation for the present results is that when males said /puk/ and females said /buk/, the overall fundamental frequency (f0) of the word (lower for males and higher for females) might have conflicted with the onset-f0 cue to voicing (higher for voiceless /p/ and lower for voiced /b/; Ohde, 1984). This could have impaired learning for the children who learn the male-/puk/ and female-/buk/ pairing. However, as gender-word assignments were counterbalanced across children, we were able to investigate this possibility in Experiment 2, and we found no effect of cue convergence in an analysis of variance. Still, when we presented our stimuli to adult listeners in an informal experiment (reported in Experiment 2 *Auditory Stimuli*), cue convergence did lead to a slight advantage in word-identification accuracy, so the impact of cue convergence on infants' and adults' learning should be investigated further in future research.

5. Conclusion

We found that when similar-sounding words were each spoken by a different gender, this correlated talker-gender information appeared to impair 14-month-olds' word learning relative to uncorrelated talker variability. Structure on a non-phonological dimension that is usually not relevant to word learning—talker gender—appears to have increased the task complexity (Yoshida, Fennell, Swingley, & Werker, 2009; Fennell, 2012; Werker & Curtin, 2005), perhaps by introducing competition between cues or creating a surprising correlation that infants needed to explain (Gerken et al., 2015). However, two other explanations are also possible. First, reduced within-word variability, caused by having only male talkers say one word and only female talkers say the other word, could have impaired learning. Second, for roughly half the infants, male talkers said /puk/ and females said /buk/, creating potential cue conflict between the talker's mean f0 and the onset-f0 cue to voiceless vs. voiced consonants. While we found no evidence that this impacted infants' learning, future research should consider it more directly. Thus, the precise nature of the interference effect must be investigated further in future research. The fact that the interfering dimension (talker gender) was a non-phonological one indicates that infants are still learning to apply native-language dimensions to the task of word learning. However, our results suggest that infants seem to

[3] Thanks to Rebecca Gómez for this suggestion.

treat the two dimensions (VOT and talker gender) as distinct cues, rather than simply integrating them to separate words in multidimensional acoustic space.

Acknowledgements

We thank members of the Tweety Language Development Lab at the University of Arizona, including Jeannette Bell, Erin Huff, Celeste Gallegos, and Devin McGrane, who recruited and tested participants and conducted reliability coding. We also thank Colin Dawson and Rebecca Gómez for insightful suggestions, and Jessamyn Schertz and Natasha Warner for phonetics consultations. Finally, thanks go to Sarah Creel for helping us pilot this work with preschool children. Funding was provided by NIH grants F32 HD065382 and R00 DC013795 to CQ and NSF grant 0950601 to LAG.

7. References

Apfelbaum KS, & McMurray B (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. Cognitive Science, 35, 1105–1138. [PubMed: 21609356]

Bosch L, & Sebastián-Gallés N (2003). Simultaneous bilingualism and the perception of a language-specific vowel contrast in the first year of life. Language and Speech, 46, 217–243. [PubMed: 14748445]

Bradlow A, Pisoni D, Akahane-Yamada R, & Tohkura Y (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. Journal of the Acoustical Society of America, 101, 2299–2310. [PubMed: 9104031]

Caselli MC, Bates E, Casadio P, Fenson J, Fenson L, Sanderl L, & Weir J (1995). A cross-linguistic study of early lexical development. Cognitive Development, 10, 159–199.

Christiansen MH (2013a). Language has evolved to depend on multiple-cue integration In Botha R and Everaert M (Eds.). The evolutionary emergence of language: Evidence and inference (pp. 42–61). Oxford: Oxford University Press.

Christiansen MH (2013b). Modeling culteral evolution: Language acquisition as multiple-cue integration In Lefebvre C, Comrie B & Cohen H (Eds.), New Perspectives on the origins of language (pp. 487–504). Amsterdam, the Netherlands: John Benjamins.

Cohen LB, Atkinson DJ, & Chaput HH (2004). Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). Austin: University of Texas.

Cristia A, McGuire GL, Seidl A, & Francis AL (2011). Effects of the distribution of acoustic cues on infants' perception of sibilants. Journal of Phonetics, 39, 388–402. [PubMed: 21804656]

Dawson C, & Gerken LA (2011). When global structure "Explains Away" local grammar: A Bayesian account of rule-induction in tone sequences. Cognition, 120, 350–359. [PubMed: 21257161]

Feldman NH, Myers EB, White KS, Griffiths TL, & Morgan JL (2013). Word-level information influences phonetic learning in adults and infants. Cognition, 127, 427–438. [PubMed: 23562941]

Fennell CT (2012). Object familiarity enhances infants' use of phonetic detail in novel words. Infancy, 17, 339–353.

Fennell CT, & Waxman SR (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. Child Development, 81, 1376–1383. [PubMed: 20840228]

Fennell CT, & Werker JF (2003). Early word learners' ability to access phonetic detail in well-known words. Language and Speech, 46, 245–264. [PubMed: 14748446]

Fennell CT, & Werker JF (2004). Infant attention to phonetic detail: Knowledge and familiarity effects. Proceedings of the 28th Annual Boston University Conference on Language Development, 165–176.

Foulkes P, & Docherty G (2006). The social life of phonetics and phonology. Journal of Phonetics, 34, 409–438.

Frank MC, Slemmer JA, Marcus GF, & Johnson SP (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. Developmental Science, 12, 504–509. [PubMed: 19635078]

Frota S, Butler J, Correia S, Severino C, & Vigário M (2012). Pitch first, stress next? Prosodic effects on word learning in an intonation language. Proceedings of the 36th Annual Boston University Conference on Language Development, 190–201.

Galle M, Apfelbaum K, & McMurray B (2015). The role of single talker acoustic variation in early word learning. Language Learning and Development, 11, 66–79. [PubMed: 27594811]

Gerken LA, Dawson C, Chatila R & Tenenbaum J (2015). Surprise! Infants consider possible bases of generalization for a single input example. Developmental Science, 18, 80–89. [PubMed: 24703007]

Glass GV, Peckham PD, & Sanders JR (1972). Consequences of failure to meet assumptions underlying fixed effects analyses of variance and covariance. Review of Educational Research, 42, 237–288.

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

Gómez RL (2002). Variability and detection of invariant structure. Psychological Science, 13, 431–436. [PubMed: 12219809]

Gómez RL (2002). Variability and detection of invariant structure. Psychological Science, 13, 431–436. [PubMed: 12219809]

Gonzales K, Gómez RL, & Gerken LA (2011). 12-month-olds use voice and temporal cues to extract structure that only one of two speakers follows consistently. Paper presented at the 36th Annual Boston University Conference on Language Development, Boston, MA.

Graf Estes K, Evans JL, Alibali MW, & Saffran JR (2007). Can infants map meaning to newly segmented words? Statistical segmentation and word learning. Psychological Science, 18, 254–260. [PubMed: 17444923]

Greenspan SL, Nusbaum HC, & Pisoni DB (1988). Perceptual learning of synthetic speech produced by rule. Journal of Experimental Psychology: Learning, Memory & Cognition, 14, 421–433.

Harwell MR, Rubinstein EN, Hayes WS, & Olds CC (1992). Summarizing Monte Carlo results in methodological research: the one- and two-factor fixed effects ANOVA cases. Journal of Educational Statistics, 17, 315–339.

Hay J, Graf Estes K, Wang T, & Saffran JR (2015). From flexibility to constraint: The contrastive use of lexical tone in early word learning. Child Development, 86, 10–22. [PubMed: 25041105]

Heald SLM, & Nusbaum HC (2015). Variability in vowel production within and between days. PLOS One, 10. doi: 10.1371/journal.pone.0136791

Houston DM, & Jusczyk PW (2000). The role of talker-specific information in word segmentation by infants. Journal of Experimental Psychology: Human Perception and Performance, 26, 1570–1582. [PubMed: 11039485]

Iverson P, Hazan V, & Bannister K (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. Journal of the Acoustical Society of America, 118, 3267–3278. [PubMed: 16334698]

Iverson P, Pinet M, & Evans B (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. Applied Psycholinguistics, 33, 145–160.

Johnson K (1997). The auditory/perceptual basis for speech segmentation. OSU Working Papers in Linguistics, 50, 101–113.

Johnson K (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. Journal of Phonetics, 34, 485–499.

Johnson K, Ladefoged P, & Lindau M (1993). Individual differences in vowel production. Journal of the Acoustical Society of America, 94, 701–714. [PubMed: 8370875]

Jusczyk PW (1993). From general to language-specific capacities: The WRAPSA Model of how speech perception develops. Journal of Phonetics, 21, 3–28.

Kuhl PK, Stevens E, Hayashi A, Deguchi T, Kiritani S, & Iverson P (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. Developmental Science, 9, F13–F21. [PubMed: 16472309]

Lany J, and Saffran JR (2010). From statistics to meaning: Infants' acquisition of lexical categories. Psychological Science, 21, 284–291. [PubMed: 20424058]

Lively SE, Logan JS, & Pisoni DB (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. Journal of the Acoustical Society of America, 94, 1242–1255. [PubMed: 8408964]

Logan IS, Lively SE, & Pisoni DB (1991). Training Japanese listeners to identify English /r/ and /1/: A first report. Journal of the Acoustical Society of America, 89, 874–886. [PubMed: 2016438]

Maye J, Weiss DJ & Aslin RN (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. Developmental Science, 11, 122–134. [PubMed: 18171374]

Maye J, Werker JF, & Gerken LA (2002). Infant sensitivity to distributional information can affect phonetic discrimination. Cognition, 82, B101–B111. [PubMed: 11747867]

Namy LL (2001). What's in a name when it isn't a word? 17-month-olds' mapping of nonverbal symbols to object categories. Infancy, 2, 73–86.

Namy LL, & Waxman SR (1998). Words and gestures: Infants' interpretations of different forms of symbolic reference. Child Development, 69, 295–308. [PubMed: 9586206]

Narayan CR, Werker JF, & Beddor PS (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. Developmental Science, 13, 407–420. [PubMed: 20443962]

Oakes LM (2010). Using habituation of looking time to assess mental processes in infancy. Journal of Cognitive Development, 11, 255–268.

Ohde Ralph. 1984 Fundamental frequency as an acoustic correlate of stop consonant voicing. Journal of the Acoustical Society of America, 75, 224–230. [PubMed: 6699284]

Palmieri TJ, Goldinger SD, & Pisoni DB (1993). Episodic encoding of voice attributes and recognition memory for spoken words. Journal of Experimental Psychology: Learning, Memory, and Cognition, 19, 309–328.

Pearl J (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. San Francisco: Morgan Kaufmann.

Peterson GE, & Barney HL (1952). Control methods used in a study of the vowels. The Journal of the Acoustical Society of America, 24, 175–184.

Pierrehumbert JB (2001). Exemplar dynamics: Word frequency, lenition and contrast In Bybee J & Hopper P (Eds.), Frequency and the emergence of linguistic structure (pp. 337–359). Amsterdam, Netherlands: John Benjamins Publishing Company.

Pierrehumbert JB (2002). Word-specific phonetics In Laboratory Phonology VII, (pp. 101–139). Berlin: Mouton de Gruyter.

Plante E, Ogilvie T, Vance R, Aguilar JM, Dailey NS, Meyers C, Lieser AM, & Burton R (2014). Variability in the language input to children enhances learning in a treatment context. American Journal of Speech-Language Pathology, 23, 530–545. [PubMed: 24700145]

Polka L, & Werker JF (1994). Developmental changes in perception of nonnative vowel contrasts. Journal of Experimental Psychology: Human Perception and Performance, 20, 421–435. [PubMed: 8189202]

Posner MI, & Keele SW (1968). On the genesis of abstract ideas. Journal of Experimental Psychology, 77, 353–363. [PubMed: 5665566]

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, 135–150. [PubMed: 20161601]

Richtsmeier PT, Gerken LA, Goffman L, & Hogan T (2009). Statistical frequency in perception affects children's lexical production. Cognition, 111, 372–377. [PubMed: 19338981]

Rost GC, & McMurray B (2009). Speaker variability augments phonological processing in early word learning. Developmental Science, 12, 339–349. [PubMed: 19143806]

Rost GC, & McMurray B (2010). Finding the signal by adding noise: The role of nonconstrastive phonetic variability in early word learning. Infancy, 15, 608–635.

Singh L, Hui TJ, Chan C, & Golinkoff RM (2014). Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. Developmental Science, 17, 94–109. [PubMed: 24118787]

Singh L, Morgan JL, & White KS (2004). Preference and processing: The role of speech affect in early spoken word recognition. Journal of Memory and Language, 51, 173–189.

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

Stager CL, & Werker JF (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. Nature, 388, 381–382. [PubMed: 9237755]

Stevens KN, & Klatt DH (1973). Role of formant transitions in the voiced-voiceless distinction for stops. Journal of the Acoustical Society of America, 55, 653–659.

Swingley D (2009). Contributions of infant word learning to language development. Philosophical Transactions of the Royal Society, 364, 3617–3632.

Teinonen T, Aslin RN, Alku P, & Csibra G (2008). Visual speech contributes to phonetic learning in 6-month-old infants. Cognition, 108, 850–855. [PubMed: 18590910]

Thiessen ED (2007). The effect of distributional information on children’s use of phonemic contrasts. Journal of Memory and Language, 56, 16–34.

Thiessen ED (2010). Effects of visual information on adults’ and infants’ auditory statistical learning. Cognitive Science, 34, 1093–1106. [PubMed: 21564244]

Thiessen ED (2012). Effects of Inter- and Intra-modal Redundancy on Infants’ Rule Learning. Language Learning and Development, 8, 197–214.

van den Bos E, Christiansen MH, & Misyak JB (2012). Statistical learning of probabilistic nonadjacent dependencies by multiple-cue integration. Journal of Memory and Language, 67, 507–520.

Wang Y, Spence MM, Jongman A, & Sereno JA (1999). Training American listeners to perceive Mandarin tones. Journal of the Acoustical Society of America, 106, 3649–3658. [PubMed: 10615703]

Weiss DJ, Gerfen C & Mitchel AD (2009) Speech segmentation in a simulated bilingual environment: A challenge for statistical learning? Language Learning and Development 5, 30–49. [PubMed: 24729760]

Werker JF, & Curtin S (2005). PRIMIR: A developmental framework of infant speech processing. Language Learning and Development, 1, 197–234.

Werker JF, & Tees RC (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development, 7, 49–63.

Werker JF, Fennell CT, Corcoran KM, & Stager CL (2002). Infants’ ability to learn phonetically similar words: Effects of age and vocabulary size. Infancy, 3, 1–30.

Woodward AL, & Hoyne KL (1999). Infants’ learning about words and sounds in relation to objects. Child Development, 70, 65–77. [PubMed: 10191515]

Yeung HH, & Werker JF (2009). Learning words’ sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. Cognition, 113, 234–243. [PubMed: 19765698]

Yoshida KA, Fennell CT, Swingley D, & Werker JF (2009). Fourteen-month-old infants learn similar-sounding words. Developmental Science, 12, 412–418. [PubMed: 19371365]

Zhao J, Al-Aidroos N, & Turk-Browne NB (2013). Attention is spontaneously biased toward regularities. Psychological Science, 24, 667–677. [PubMed: 23558552]

Zlatin MA, & Koenigsknecht RA (1976). Development of the voicing contrast: A comparison of voice onset time in stop perception and production. Journal of Speech and Hearing Research, 19, 93–111. [PubMed: 1271805]

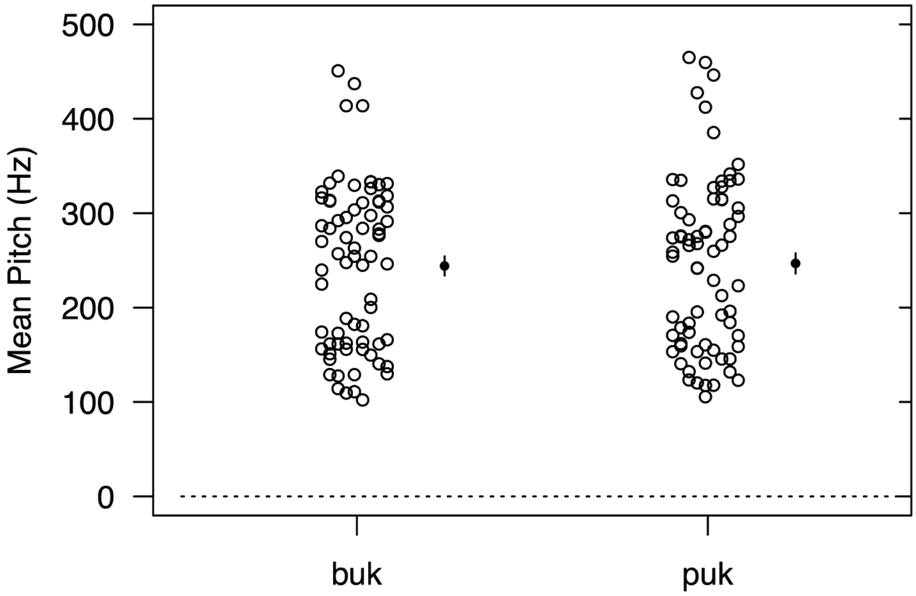

Figure 1.

Mean pitch for 72 tokens each of /buk/ and /puk/ in Experiment 1, where talker gender was uncorrelated with the words. Filled-in black circles indicate means and vertical lines through them indicate standard errors. To improve visibility, points were plotted with x-axis jitter.

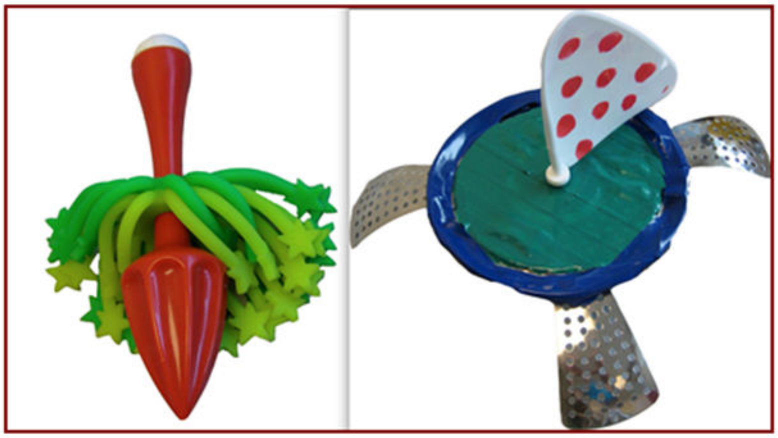

Figure 2.
/buk/ and /puk/ objects

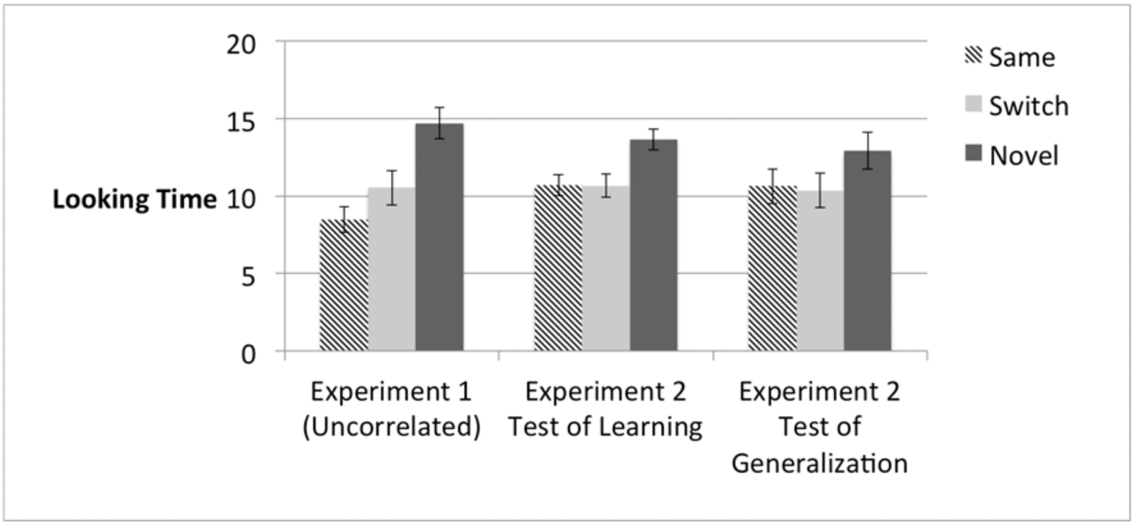

Figure 3.

Raw looking times (with standard-error bars) across trial types in each experiment

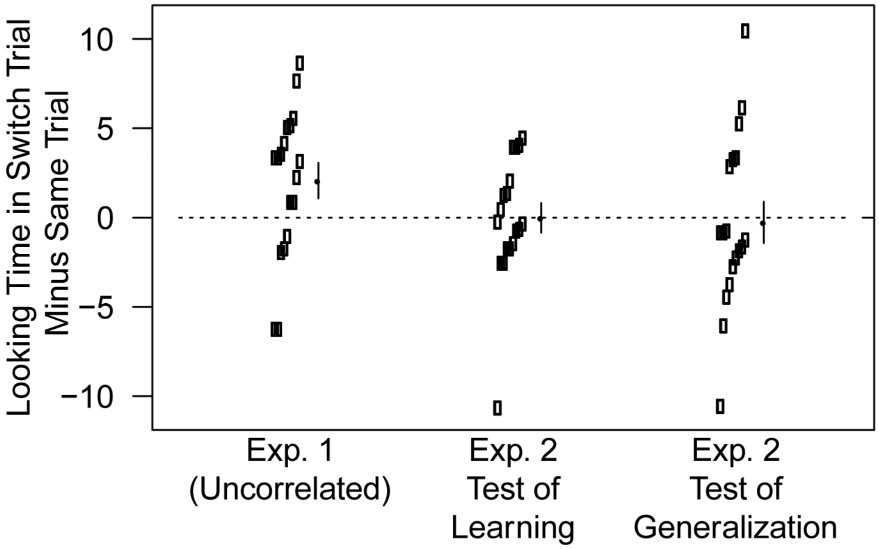

Figure 4.

Scatterplot of raw looking times in Switch trials minus Same trials for participants in each experiment. Filled circles and vertical lines indicate means and standard errors. Points are jittered on the x-axis to improve visibility.

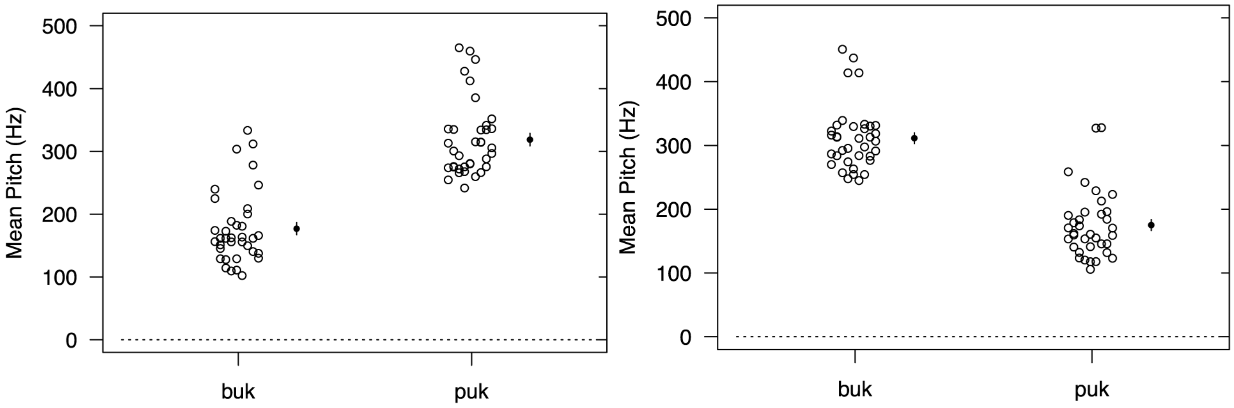

Figure 5.

Mean pitch for all tokens of /buk/ vs. /puk/ in Experiment 2, for male-/buk/ and female-/puk/ training (left box) and female-/buk/ and male-/puk/ training (right box). Filled-in black circles indicate means and vertical lines through them indicate standard errors. Points are jittered on the x-axis to make them more visible.

Table 1a.

Pitch measurements for each talker (with standard deviations)

	Talker	Pitch mean	Pitch max	Pitch SD
Training females	**1**	288 (18) Hz	339 (36) Hz	44 (13) Hz
	2	277 (31)	338 (46)	52 (16)
	3	340 (39)	388 (26)	45 (19)
	4	313 (29)	372 (46)	52 (17)
	5	261 (13)	300 (23)	33 (9)
	6	289 (18)	318 (17)	31 (6)
	7	439 (20)	512 (16)	67 (21)
	8	320 (18)	390 (37)	58 (15)
	9	308 (25)	388 (47)	60 (23)
	Mean	***315 (52) Hz***	***372 (62) Hz***	***49 (12) Hz***
Training males	**1**	183 (34)	204 (38)	17 (7)
	2	295 (48)	341 (34)	43 (18)
	3	264 (79)	301 (98)	28 (14)
	4	183 (34)	209 (52)	19 (19)
	5	194 (42)	239 (53)	36 (12)
	6	116 (9)	135 (23)	13 (8)
	7	211 (35)	269 (43)	48 (11)
	8	144 (16)	159 (15)	13 (3)
	9	117 (9)	137 (12)	14 (6)
	Mean	***190 (61) Hz***	***222 (73) Hz***	***26 (14) Hz***
Training female/male ratio		***1.658***	***1.676***	***1.885***
Test females	**1**	268 (6)	299 (10)	32 (9)
	2	325 (13)	390 (17)	59 (5)
	3	313 (42)	384 (83)	47 (27)
	4	304 (16)	416 (25)	82 (7)
	Mean	***303 (24) Hz***	***372 (51) Hz***	***55 (21) Hz***
Test males	**1**	177 (19)	193 (22)	12 (6)
	2	229 (26)	293 (61)	43 (21)
	3	160 (23)	211 (33)	32 (8)
	4	149 (8)	159 (8)	6 (2)
	Mean	***179 (36) Hz***	***214 (57) Hz***	***23 (17) Hz***
Test female/male ratio		***1.693***	***1.738***	***2.391***

Table 1b

Mean formant measurements for each talker (with standard deviations), computed 1/3 of the way into the vowel using the Praat "Get formant" function

	Talker	F1	F2	F3
Training females	**1**	613 (354) Hz	1916 (246) Hz	3025 (139) Hz
	2	549 (47)	1995 (55)	2892 (264)
	3	427 (11)	1504 (47)	2986 (62)
	4	561 (56)	1910 (128)	2876 (106)
	5	476 (35)	1666 (55)	2828 (109)
	6	624 (36)	1590 (83)	2753 (68)
	7	673 (228)	1320 (163)	3033 (228)
	8	558 (49)	1705 (163)	2883 (69)
	9	482 (143)	1585 (211)	2937 (122)
	Mean	***552 (162) Hz***	***1688 (249) Hz***	***2913 (163) Hz***
Training males	**1**	385 (19)	1431 (82)	2181 (33)
	2	345 (19)	1130 (54)	2243 (82)
	3	362 (63)	1414 (130)	2283 (28)
	4	414 (21)	1298 (100)	2271 (75)
	5	499 (353)	1609 (308)	2453 (411)
	6	388 (16)	1051 (56)	2519 (122)
	7	356 (16)	1308 (106)	2342 (47)
	8	489 (43)	1384 (59)	2225 (98)
	9	388 (20)	1064 (61)	2513 (122)
	Mean	***403 (126) Hz***	***1299 (216) Hz***	***2337 (192) Hz***
Training female/male ratio		***1.37***	***1.299***	***1.246***
Test females	**1**	495 (33)	1608 (115)	2819 (35)
	2	638 (64)	1640 (96)	2844 (80)
	3	541 (136)	1568 (172)	3007 (477)
	4	445 (19)	1581 (73)	2856 (61)
	Mean	***530 (103) Hz***	***1599 (117) Hz***	***2882 (244) Hz***
Test males	**1**	394 (18)	1256 (68)	2469 (37)
	2	494 (193)	1326 (325)	2677 (330)
	3	273 (126)	1346 (227)	2527 (205)
	4	387 (24)	1574 (62)	2192 (106)
	Mean	***387 (136) Hz***	***1376 (228) Hz***	***2466 (262) Hz***
Test female/male ratio		***1.37***	***1.162***	***1.168***

Table 1c

Mean voice-onset time hand-measurements for each talker/word combination (with standard deviations).

	Talker	***/buk/* (voiced)**	***/puk/* (voiceless)**
Training females	**1**	−35 (100) milliseconds	90 (23) milliseconds
	2	13 (2)	60 (10)
	3	19 (3)	73 (7)
	4	24 (6)	77 (6)
	5	29 (4)	77 (19)
	6	17 (2)	67 (9)
	7	24 (11)	104 (13)
	8	25 (3)	70 (7)
	9	−67 (59)	93 (25)
	Mean	***5 (47) milliseconds***	***79 (19) milliseconds***
Training males	**1**	10 (5)	47 (8)
	2	2 (31)	55 (16)
	3	−44 (127)	57 (9)
	4	6 (29)	56 (3)
	5	−13 (74)	62 (12)
	6	25 (10)	82 (12)
	7	24 (2)	101 (12)
	8	18 (6)	62 (9)
	9	21 (6)	100 (17)
	Mean	***6 (50) milliseconds***	***69 (22) milliseconds***
Test females	**1**	18 (5)	79 (19)
	2	10 (1)	63 (12)
	3	17 (10)	103 (8)
	4	−13 (49)	108 (2)
	Mean	***8 (26) milliseconds***	***88 (22) milliseconds***
Test males	**1**	6 (21)	53 (9)
	2	−45 (52)	97 (14)
	3	−166 (50)	67 (11)
	4	17 (0)	52 (11)
	Mean	***−47 (82) milliseconds***	***67 (21) milliseconds***

Table 2.

Mean looking times in seconds (with standard deviations) in each trial type across experiments

Trial Type	**Exper. 1**	**Exper. 2 overall**	**Exper. 2 “Test of Learning”**	**Exper. 2 “Test of Generalization”**
Same-1st trial	8.46 (3.50)	10.66 (3.90)	10.68 (2.84)	10.63 (4.81)
Switch-1st trial	10.53 (4.70)	10.51 (3.95)	10.67 (3.18)	10.36 (4.69)
Novel	14.70 (4.33)	13.28 (4.12)	13.64 (2.93)	12.92 (5.11)
Same-Both trials	8.79 (3.32)	10.11 (2.85)	10.20 (2.28)	10.02 (3.39)
Switch-Both trials	9.72 (3.17)	9.80 (3.42)	10.34 (2.83)	9.26 (3.93)