

Portland State University

PDXScholar

Speech and Hearing Sciences Faculty
Publications and Presentations

Speech and Hearing Sciences

10-2020

Infants' Discrimination of Consonant Contrasts in the Presence and Absence of Talker Variability

Carolyn Quam

Portland State University, cquam@pdx.edu

Lauren Clough

University of Arizona

Sara Knight

University of Arizona

LouAnn Gerken

University of Arizona

Follow this and additional works at: https://pdxscholar.library.pdx.edu/sphr_fac



Part of the [Speech and Hearing Science Commons](#), and the [Speech Pathology and Audiology Commons](#)

Let us know how access to this document benefits you.

Citation Details

Published as: Quam C, Clough L, Knight S, Gerken L. Infants' discrimination of consonant contrasts in the presence and absence of talker variability. *Infancy*. 2020;00:1–20. <https://doi.org/10.1111/infa.12371>

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Speech and Hearing Sciences Faculty Publications and Presentations by an authorized administrator of PDXScholar. Please contact us if we can make this document more accessible: pdxscholar@pdx.edu.

Infants' Discrimination of Consonant Contrasts in the Presence and Absence of
Talker Variability

*Carolyn Quam*¹

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

Departments of Speech, Language, and Hearing Sciences and Psychology,
University of Arizona, USA

Lauren Clough

Departments of Educational Psychology and Linguistics, University of Arizona,
USA

Sara Knight

Departments of Psychology and Psychiatry, University of Arizona, USA

and

LouAnn Gerken

Department of Psychology, University of Arizona, USA

Keywords: speech perception; infancy; phonetics; sound discrimination; variability

Wordcount: 9,918

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

Abstract

To learn speech-sound categories, infants must identify the acoustic dimensions that differentiate categories and selectively attend to them as opposed to irrelevant dimensions. Variability on irrelevant acoustic dimensions can aid formation of robust categories in infants through adults in tasks such as word learning (e.g., Rost & McMurray, 2009) or speech-sound learning (e.g., Lively, Logan, & Pisoni, 1993). At the same time, variability sometimes overwhelms learners, interfering with learning and processing. Two prior studies (Kuhl & Miller, 1982; Jusczyk, Pisoni, & Mullennix, 1992) found that irrelevant variability sometimes impaired early sound discrimination. We asked whether variability would impair or facilitate discrimination for older infants, comparing 7.5-month-old infants' discrimination of an early acquired native contrast, /p/ vs. /b/ (in the word-forms /pIm/ vs. /bIm/), in Experiment 1, with an acoustically subtle, non-native contrast, /n/ vs. /ŋ/ (in /nIm/ vs. /ŋIm/), in Experiment 2. Words were spoken by one or four talkers. Infants discriminated the native but not the non-native contrast and there were no significant effects of talker condition. We discuss implications for theories of phonological learning and avenues for future research.

Introduction

The present study investigates whether variability in talker voice impacts 7.5-month-old infants' speech-sound discrimination, an index of speech-sound knowledge. Infants must discover the speech-sound categories that differentiate words in their language(s). Speech-sound contrasts differing in a single phonetic feature that are present in many languages, like /b/ vs. /p/, are likely to be discriminated universally from birth. Between 4 to 6 months for vowels (Bosch & Sebastián-Gallés, 2003; Polka & Werker, 1994) and 10 to 12 months for consonants (Werker & Tees, 1984), infants undergo a perceptual reorganization of discrimination, exhibiting decreased sensitivity to non-native contrasts that fall within native categories. However, developmental trajectories of discrimination vary somewhat by contrast. While many non-native contrasts show decreases in sensitivity after 10-12 months, Japanese learners show improved discrimination between infancy and adulthood for German vowel contrasts (Mazuka, Hasegawa, & Tsuji, 2014). While many native-language contrasts maintain a high level of discrimination over the first year, others show a slower developmental trajectory (e.g., /l/ vs. /r/; Kuhl et al., 2006).

Across many domains, learners must establish categories robust to irrelevant, within-category variability (Sloutsky, 2010). Learning speech-sound categories requires identifying and selectively attending to acoustic-phonetic dimensions that differentiate categories, while disregarding irrelevant changes across productions. Variability on dimensions not criterial to phonological-learning tasks may help learners down-weight those dimensions and zero in on criterial dimensions (e.g., Apfelbaum & McMurray, 2011). However, work on speech-sound learning and other phonological tasks has revealed facilitative, inhibitory, and null effects of non-criterial variability.

There is an extensive literature on adults' learning of L2 speech sounds, which has demonstrated facilitative effects of talker variability for Japanese speakers' identification of

English /l/ and /r/ (Lively, Logan, & Pisoni, 1993), Dutch speakers' identification and generalization of a Japanese singleton/geminate consonant contrast (Sadakata & McQueen, 2013), and Dutch speakers' learning of Mandarin tonal patterns (Sadakata & McQueen, 2014). However, for identification of Mandarin tones, facilitation held only for learners with high perceptual aptitude; learners with low perceptual aptitude experienced inhibition (Sadakata & McQueen, 2014; see also Davis, 2015, and Perrachione, Lee, & Wong, 2011). Antoniou and Wong (2016) also reported inhibitory effects of irrelevant phonetic variability. In addition, across many studies (e.g., Magnuson & Nusbaum, 2007), listeners' identification of native-language speech sounds is impaired when they have to adjust cognitively to multiple talkers. Thus, adult learning and processing of sound categories is sometimes facilitated and sometimes inhibited by variability.

Infants exploit variation on dimensions criterial to the learning task (Maye, Werker, & Gerken, 2002; Teinonen, Aslin, Alku, & Csibra, 2008; see also Weatherhead & White, 2016; van der Feest & Johnson, 2016). However, evidence is more limited regarding the impact of *non-criterial* variability on infants' speech-sound learning. Kuhl and Miller (1982), in the high-amplitude-sucking (HAS) procedure, found that pitch discrimination in 4- to 16-week-old infants was impaired by vowel variation, but not vice-versa. Jusczyk, Pisoni, and Mullennix (1992) tested 2-month-olds' detection of changes to syllables (/b^g/ vs. /d^g/) in the HAS procedure. When testing immediately followed familiarization, talker variability did not impact discrimination. With a 2-minute delay before test, children familiarized to six female and six male talkers did not detect the change in syllable, while children familiarized to a single talker did. Finally, Kuhl (1983) found that 6-month-old infants trained to discriminate a vowel pair produced by a single synthesized "talker" successfully generalized discrimination to multiple simulated men's, women's, and

children's voices, suggesting variability did not disrupt discrimination. Thus, young infants' sound discrimination is sometimes inhibited and sometimes not impacted by variability.

At 7.5 months (the age tested here), infants fail to recognize familiarized word forms across changes in talker gender (Houston & Jusczyk, 2000), pitch (Singh, White, & Morgan, 2008), or affect (Singh, Morgan, & White, 2004). However, if acoustic variability is incorporated into training, it can aid formation of robust representations that generalize to a broader range of stimuli (Singh, 2008). Thus, training variability is simultaneously challenging for young infants and essential to a robust phonological-learning process.

While studies of infant sound discrimination, unlike some adult studies, have not indicated facilitative effects of variability, facilitation has been found in other phonological-learning tasks. Seidl, Onishi, and Cristia (2014) reported that learning of phonotactic strings at 4 and 11 months was facilitated when strings were spoken by multiple training talkers. Facilitation has also been found for early word learning in the Switch habituation procedure. In the Switch procedure, 14-month-old infants often fail to detect differences between similar-sounding words (such as /bI/ and /dI/) despite distinguishing the individual sounds (/b/ and /d/, Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002). Rost and McMurray (2009, 2010; see also Höhle et al., 2020) found that 14-month-olds differentiated /buk/ vs. /puk/ with 18 habituation talkers, but not with a single talker. Notably, the paradigm used by Rost and McMurray (2009) is the same one used in the present study, except that Rost and McMurray paired word forms with distinct visual referents to probe word learning.

In a mechanistic model of facilitative impacts of phonetic variability on early word learning (Apfelbaum & McMurray, 2011), relevant acoustic-phonetic dimensions are identified during word learning. Variability reduces associations between non-phonological dimensions (e.g., pitch)

and visual referents. The tasks of associative word learning and speech-sound learning both require identifying relevant dimensions and attending to them while disregarding irrelevant dimensions. However, as variability in Apfelbaum and McMurray's model operates on associative strengths between non-criterial phonological dimensions and visual objects, it is not obvious mechanistically how talker variability would help differentiate speech-sound (or syllable) representations in the non-referential task used here.

The Present Study

The present study investigated potential impacts of talker variability on 7.5-month-olds' speech-sound discrimination. Infants were randomly assigned to two between-subjects conditions: one with a single habituation talker and one with four talkers. Talker variability during habituation could facilitate robust sound categorization and differentiation (à la Rost & McMurray, 2009 or Singh, 2008) or overwhelm learners with additional complexity (à la Houston & Jusczyk, 2000 or Kuhl & Miller, 1982). While a Switch discrimination task might seem to index existing knowledge more than dynamic learning (compared with tasks described above that have taught infants words or adults L2 categories), the process of habituating involves both learning about the laboratory stimulus set and drawing on existing knowledge (Oakes, 2010). Thus, learning principles that apply in other tasks (e.g., word learning) might also apply to sound-discrimination tasks.

Whether increased variability aids or hinders discrimination could depend on how the complexity it introduces interacts with the age group's processing abilities and the difficulty of the contrast (Kuhl & Miller, 1982; Werker & Curtin, 2005; Fennell & Werker, 2003; Fennell & Waxman, 2010; Yoshida, Fennell, Swingley, & Werker, 2009; see also Kidd, Piantadosi, & Aslin, 2012). As training variability sometimes inhibits and sometimes has no effect on younger infants'

sound discrimination (Kuhl & Miller, 1982; Jusczyk, Pisoni, & Mullenix, 1992), but facilitates the formation of more robust word-form representations at 7.5 months (Singh, 2008), we could not make clear *a priori* predictions about how variability would impact discrimination performance. As a first step in this line of research, we tested English-learning infants' discrimination of two contrasts representing quite distinct cases.

Experiment 1 tested discrimination of /b/ vs. /p/, a contrast that is attested in onset position in English, relatively acoustically salient, and early acquired (Eimas, Siqueland, Jusczyk, & Vigorito, 1971). The sounds differ in voicing: initial /p/ is voiceless and aspirated, with a positive voice-onset time (VOT), while /b/ is voiced and unaspirated, with a VOT of roughly 0. Sounds were embedded in the word forms /pIm/ vs. /bIm/.

Experiment 2 tested discrimination of /n/-/ŋ/, embedded in the word forms /nIm/ vs. /ŋIm/. These are nasal sounds differing in their place of articulation in the oral cavity: /n/ is alveolar, while /ŋ/ is velar. The contrast is attested in English in coda position (e.g., in the minimal pair /sIn/, “sin,” vs. /sInj/, “sing”), but /ŋ/ is unattested in syllable-initial position in English (and most other languages), though attested in some languages, including Filipino. As a result, different discrimination trajectories across development have been found for children learning Filipino vs. English. Narayan et al. (2010) found Filipino-learning infants failed to discriminate /n/ vs. /ŋ/ at 6-8 months, not succeeding until 10-12 months, a slower time-course than for /n/ vs. /m/. They attributed this slower time-course to low acoustic salience of the contrast, which also leads Filipino-speaking adults to show significantly worse discrimination for /n/-/ŋ/ than /n/-/m/ (Narayan, 2008).

English-speaking adults, who have undergone perceptual attunement to native contrasts, do not successfully discriminate /n/ vs. /ŋ/ in onset position, while Filipino-speaking adults do

(Narayan, 2008). However, evidence about English-learning infants' discrimination is mixed. Narayan et al. found that English-learning infants did not successfully discriminate the contrast at 6-8 or 10-12 months. However, in what they argued was a more sensitive habituation paradigm, Sundara et al. (2018) found that infants learning English successfully discriminated it at 4 and 6 months, prior to the process of perceptual attunement to native contrasts.

We included the native and acoustically salient contrast /b/ vs. /p/ and the non-native and acoustically subtle contrast /n/ vs. /ŋ/ with the goal of probing for both inhibitory and facilitative effects of variability. Due to limited prior work investigating impacts of variability on infants' sound discrimination, predictions were necessarily tentative. We were informed by prior work on phonological learning and processing generally, cited above, reporting impacts of variability. However, to our knowledge, no prior studies have examined effects of talker variability on putatively easier vs. harder contrasts. Thus, rather than basing explicit predictions on prior work, we instead selected contrasts that seemed logically most likely to reveal facilitation or interference effects of variability.

We reasoned that a condition where infants show discrimination with a single speaker would be most likely to reveal inhibitory effects of variability (Kuhl & Miller, 1982; Jusczyk, Pisoni, & Mullennix, 1992). As infants should successfully discriminate the native-language, acoustically salient contrast /b/ vs. /p/ in the absence of talker variability (Eimas et al., 1971), the introduction of talker variability during habituation could introduce additional task complexity, impairing infants' detection of a change from /bim/ to /pim/ or vice-versa. Kuhl and Miller (1982)'s findings that pitch discrimination was disrupted by vowel variation, and Jusczyk, Pisoni, and Mullennix's (1992) finding of interference from talker variability in the delay condition, indicate infants' discrimination can be impaired by irrelevant variability. While Kuhl and Miller (1982)

found that infants' discrimination of the native, acoustically salient vowel contrast /i/ vs. /a/ was not disrupted in the presence of pitch-contour variation, the sounds differ in multiple phonetic features: /i/ is a high, front vowel, while /a/ is a low, back vowel. The /b/ vs. /p/ contrast used here, though native and relatively acoustically salient, differ in only one phonetic feature, making them potentially more difficult to discriminate when variability is added (Kuhl & Miller, 1982). However, it must be noted that Jusczyk, Pisoni, and Mullennix (1992) found variability did not impair discrimination of /b/ vs. /d/, which differ in only one phonetic feature, when there was no delay from familiarization to test. It was also possible, therefore, that variability would not impact discrimination of /b/ vs. /p/.

We reasoned that a condition where infants do not show discrimination with a single speaker would be most likely to reveal facilitative effects of variability (analogous to facilitation of early word learning; Rost & McMurray, 2009). While one recent study indicated successful discrimination of the non-native, acoustically subtle contrast /n/ vs. /ŋ/ in infancy (Sundara et al., 2018), another did not (Narayan et al., 2010), so infants may find this contrast difficult to discriminate. In the absence of talker variability, one might predict infants would fail to discriminate /n/ vs. /ŋ/, consistent with one set of prior findings for English-learning infants of this age (Narayan et al., 2010). In such a case, in the multiple-talker condition, exemplars from multiple talkers would offer a broader range of acoustic input on dimensions irrelevant to the contrast, potentially helping infants identify the relevant dimension(s) of contrast and facilitating discrimination (Apfelbaum & McMurray, 2011).

However, one might also predict infants would successfully discriminate /n/ vs. /ŋ/, given a more recent study finding successful discrimination of /n/ vs. /ŋ/ by 4- and 6-month-old English-learning infants (Sundara et al., 2018). Some aspects of our paradigm were more similar to Sundara

et al. (2018), such as a 50% habituation criterion that could be met in any three consecutive trials. Narayan et al. used a 60% habituation criterion that could be met every three trials (e.g., only in trial 9, 12, etc.). If we were to find successful discrimination of /n/ vs. /ŋ/, Experiment 2 could potentially reveal inhibitory effects of variability, as found previously for falling vs. monotone pitch contours in the presence of vowel variability; Kuhl & Miller, 1982). However, lack of discrimination was potentially most likely, because the study design was more similar to Narayan et al.'s design on perhaps the most critical dimension. Like Narayan et al., we used a habituation procedure in which trial lengths were consistent. Sundara et al. used an infant-controlled procedure, where trial lengths were contingent on infant looking. Our trial length (16 seconds) was intermediate between Narayan et al.'s (14 seconds) and Sundara et al.'s maximum trial length (19 seconds).

Experiment 1

Experiment 1 tested discrimination of /b/ vs. /p/ after habituation to a single talker or four talkers. We expected successful discrimination in the single-talker condition. Unsuccessful discrimination in the multiple-talker condition would indicate an interference effect.

Method

Participants

The study was conducted in accordance with the guidelines from the Declaration of Helsinki. Parent/guardian consent was obtained for each child prior to testing. Data were collected with approval from the University of Arizona Institutional Review Board. We included 37 children (21 boys, 16 girls) in analyses, divided between single-talker (n = 18) and multiple-talker

conditions (n = 19). Within each condition, children were habituated to /bIm/ (n = 18; single-talker n = 9; multiple-talker n = 9) or /pIm/ (n = 19; single-talker n = 9; multiple-talker n = 10).

Infants were eligible if gestational age at testing (age adjusted for birth term) was between 7 months, 0 days and 8 months, 0 days. All infants were born at 37 weeks' gestation or more, weighing at least 5 ½ pounds. All infants had heard English at least 70% of the time since birth (for similar language inclusion criteria, see Quam, Knight, & Gerken, 2017; Quam & Swingley, 2010, 2014). Parents reported no history of speech or language issues in their nuclear families. No infants were medicated for ear infection within one week before testing. Eight participants were excluded for fussiness (six), low birth weight (one), or significant foreign-language exposure (one). No infants failed to habituate in 24 trials.

Auditory Stimuli

To generate stimuli, five female native speakers of American English produced /bIm/ and /pIm/ in an infant-directed register. The talkers were previously recorded for a study with 13 female talkers (Quam et al., 2017). The particular subset of five was hand-selected for this study to balance acoustic characteristics between the two habituation conditions and the test phase. One talker was used for the test phase, three for the multiple-talker habituation, and one for both the single-talker and multiple-talker habituation. Talkers were assigned to roles by examining acoustic measurements of tokens (summarized in **Table 1**) and equating the single-talker-habituation talker and test talker to the average of the other three speakers as much as possible.

Table 1: Acoustic Measurements for Each Word Token Used in Experiment 1.

Word	Talker Set	Token	Pitch Mean (Hz)	Pitch Max	SD of Pitch Samples	F1	F2	Duration (ms)
<i>/bIm/</i>	<i>Single-Talker Habituation</i>	<i>1</i>	262	314	38	928	2194	857
		<i>2</i>	237	272	25	713	2185	720
		<i>3</i>	241	296	32	651	2124	761
	<i>Single/Multiple</i>	<i>4</i>	250	305	34	704	2112	751
	<i>Multiple-Talker Habituation</i>	<i>1</i>	277	370	71	790	2034	793
		<i>2</i>	192	222	13	1014	2249	825
		<i>3</i>	208	239	18	964	2152	970
	<i>Test</i>	<i>1</i>	222	248	18	871	2178	681
		<i>2</i>	206	241	25	784	2168	607
		<i>3</i>	212	263	34	475	2023	476
		<i>4</i>	214	255	29	781	2143	755
	<i>Single-Talker Habituation</i>	<i>Mean (Standard Deviation)</i>	248 (11)	297 (18)	32 (5)	749 (122)	2154 (42)	772 (59)
	<i>Multiple-Talker Habituation</i>		232 (39)	284 (68)	34 (26)	868 (146)	2137 (89)	835 (95)
	<i>Test</i>		214 (7)	252 (9)	27 (7)	728 (174)	2128 (72)	630 (119)
<i>/pIm/</i>	<i>Single-Talker Habituation</i>	<i>1</i>	252	304	37	640	2178	693
		<i>2</i>	277	348	51	649	2015	653
		<i>3</i>	240	291	33	620	2013	841
	<i>Single/Multiple</i>	<i>4</i>	245	293	34	622	2109	620
	<i>Multiple-Talker Habituation</i>	<i>1</i>	292	396	73	731	2045	646
		<i>2</i>	284	410	94	1029	2301	889
		<i>3</i>	250	348	60	1077	2206	1121
	<i>Test</i>	<i>1</i>	211	267	27	725	2103	695
		<i>2</i>	218	272	36	668	2113	769
		<i>3</i>	212	269	30	585	2113	548
<i>4</i>		217	287	41	894	2130	666	
	<i>Single-Talker Habituation</i>	<i>Mean (SD)</i>	254 (16)	309 (27)	39 (8)	633 (14)	2079 (80)	702 (98)
	<i>Multiple-Talker Habituation</i>		268 (24)	362 (53)	65 (25)	865 (223)	2165 (112)	819 (235)
	<i>Test</i>		215 (4)	274 (9)	34 (6)	718 (131)	2115 (11)	670 (92)

We used four female talkers for the multiple-talker habituation—a relatively small number—out of concern that more talkers might overwhelm such young infants, reducing facilitation effects. Work on early word learning has used 18 talkers, both males and females (Quam et al., 2017; Rost & McMurray, 2009, 2010). However, Seidl, Onishi, and Cristia (2014) found facilitation for phonotactic learning in 4- and 11-month-old infants after familiarization with

just three female talkers (though the stimuli also included many word types in both talker conditions).

Apparatus and Procedure

Infants came to the lab with their parents. In a playroom, they adjusted to the lab environment while the experimenter described the study procedure to parents. When ready, infants and parents were led to a separate, sound-attenuated testing room containing a large screen with a projector, two side speakers, and a video camera for recording looking patterns. Infants sat on parents' laps facing the screen. The experimenter sat in a separate control room and viewed the video of the infant's face on a computer screen.

Audiovisual stimuli were presented using Habit (Cohen, Atkinson, & Chaput, 2004). The habituation phase lasted 24 trials maximum. Each trial began with an attention-getting stimulus that drew children's gaze to the screen: a baby jumping in a crib, with a squeaking pacifier sound (Quam et al., 2017). After infants oriented to the attention-getter, the experimenter pressed a button to start the trial. During the trial, infants viewed a black and red checkerboard while hearing sounds. Each trial was 16 seconds long and contained eight word tokens. The experimenter pressed a second button to mark the start and end of each of looks to the screen. The total looking time for each trial was the sum of all looks to the screen. Habit summed looking times over the first three trials to calculate a baseline level of looking. Then, a moving window computed summed looking times for each set of three consecutive trials until this sum was 50% or less of baseline. At this point, the child was considered to have habituated (Oakes, 2010; Quam et al., 2017) and the test phase began. If children did not habituate by the 24th habituation trial, they still proceeded to the test phase but were excluded from analysis (Oakes, 2010).

Each child was habituated to /bIm/ or /pIm/. Each trial contained four distinct tokens of /bIm/ or /pIm/, repeated twice each for eight tokens per trial. In the single-talker condition, tokens were all spoken by the same talker. In the multiple-talker condition, each token was spoken by a different talker. **Table 1** reports acoustic measurements for each token in each condition. There were eight different within-trial orderings of the tokens, presented in three blocks maximum (for 24 trials maximum). Trial order was randomized within block.

During the test phase, children were presented with two “Same” trials, in which the original word from the habituation phase was presented again, and two “Switch” trials, in which the word was changed from /bIm/ to /pIm/ or vice-versa. Children were randomly assigned to one of four test-trial orders (SWSW, WSWS, SWWS, WSSW, where ‘S’ is a Same trial and ‘W’ a Switch) crossed with the habituation word (/bIm/, /pIm/), for eight possible assignments. The final trial was a post-test, novel trial included to check whether infants were still attending to the task, by confirming that their attention perked up when they heard an entirely new word form: /paez/ for infants familiarized to /bIm/ and /baez/ for infants familiarized to /pIm/. Novel stimuli were pulled from a larger set of /b/- and /p/-initial stimuli recorded for the prior study (Quam et al., 2017), and chosen in particular for being highly distinct from /pIm/-/bIm/ in their nuclei and codas.

Statistical analyses are conducted on the looking times recorded online by the experimenter. However, these looking times were recorded under time pressure. To verify their reliability, we conducted offline coding on 17, or 24%, of the participant videos. Reliability was operationalized as the Pearson’s correlation between trial-by-trial total looking times in the online and offline coding files. The overall correlation was strong, $r = .80$, $p < .001$. All videos had correlation coefficients with moderate-to-large or large effect sizes (M correlation coefficient = .80; range = .44-.96), and all $p < .07$ (according to Cohen’s 1988 guidelines, $r = .3$ is “moderate”

and $r = .5$ is “large” in the context of social and behavioral science). As one video (of the 17 checked) had only a marginally significant correlation, $r = .44$, $p = .061$, we examined discrepancies for the four videos with correlation coefficients below .6 to determine whether any discrepancies were caused by errors in the online coding, focusing on the 13 total trials with the largest discrepancies. Of these, six were caused by issues in the offline coding and seven were caused by issues in the online coding. Both types of discrepancies were usually linked to ambiguity in whether the child was looking at the screen, e.g., due to non-central head position or gaze position near the screen edge. After careful analysis, we determined that none of the issues meaningfully affected the results. As we cannot conduct offline coding on all videos, and the issues did not meaningfully affect the results, we have retained all four participants in analyses.

Results and Discussion

Visual inspection of residuals and Shapiro-Wilk tests of normality, conducted separately for each trial type, indicated residuals were normally distributed in all trial types. Mauchly’s Test of Sphericity indicated the sphericity assumption of analysis of variance (ANOVA) was not violated. For the main effect of trial type, Mauchly’s $W = 0.88$, $p = .122$. Thus, we employed parametric tests (ANOVAs and t tests). The novel post-test trial was included in the factor “Trial Type” alongside Same and Switch trials (Rost & McMurray, 2009, 2010; Quam et al., 2017). Subject means for Same and Switch trials were computed across the two trials of each type prior to their inclusion in ANOVAs and t tests.

An ANOVA on raw looking times with the within-subjects factor Trial Type (Same, Switch, Novel) and the between-subjects factor Talker Condition (Single Talker, Multiple Talkers) revealed a main effect of Trial Type, $F(2,70) = 18.26$, $p < .001$, with no main effect or interaction

with Talker Condition. Planned comparisons (paired, two-tailed t tests) indicated that looking times in the Novel trial exceeded looking times in both Same trials, paired $t(36) = 5.28, p < .001$, Cohen's $d = 0.87$, and Switch trials, $t(36) = 4.01, p < .001$, Cohen's $d = 0.66$; see **Table 2** for means. Looking times were also significantly higher in Switch trials than Same trials, $t(36) = 2.39, p = .022$; Cohen's $d = 0.39$.

Table 2: Mean Looking Times (With Standard Deviations) in Experiments 1 and 2.

Trial Type	Exper. 1 Overall	Exper. 1: Single Talker	Exper. 1: Multiple Talkers	Exper. 2 Overall	Exper. 2: Single Talker	Exper. 2: Multiple Talkers
<i>Same</i>	6.4 (2.3)	6.0 (2.1)	6.7 (2.4)	6.8 (2.6)	7.2 (2.8)	6.3 (2.5)
<i>Switch</i>	7.2 (2.5)	6.6 (2.4)	7.8 (2.4)	6.3 (2.8)	6.8 (2.7)	5.8 (2.8)
<i>Novel</i>	8.9 (2.7)	8.5 (3.2)	9.3 (2.2)	7.3 (3.7)	7.4 (3.9)	7.2 (3.7)

An additional ANOVA checked for effects of the additional variables Trained Word (/bIm/, /pIm/), Infant Gender (male, female), and Test-Trial Order (SWSW, WSWS, SWWS, WSSW). These three variables were included as between-subjects predictors in addition to the predictors of interest (Trial Type and Talker Condition). The main effect of Trial Type, $F(2,10) = 15.69, p = .001$, was not meaningfully affected by the inclusion of these other variables, nor were there any significant main effects of or interactions with these variables.

Table 2 and **Figure 1** report mean looking times. Due to *a priori* interest, we report mean looking times overall and separated by talker condition (single talker, multiple talker). Significantly greater looking time in the Novel trial than Same or Switch trials indicates infants were still attending to the task by the end of the experiment. Significantly greater looking in Switch vs. Same trials indicates successful discrimination. No significant effects of variability emerged in the ANOVA. Visual inspection of means indicates that Switch-trial looking times were greater than Same-trial looking times in both talker conditions, but this difference was numerically (though non-significantly) greater in the multiple-talker condition. In an interference effect,

discrimination would have been significantly *worse* in the multiple-talker condition. Thus, results from Experiment 1 are incompatible with an interference effect.

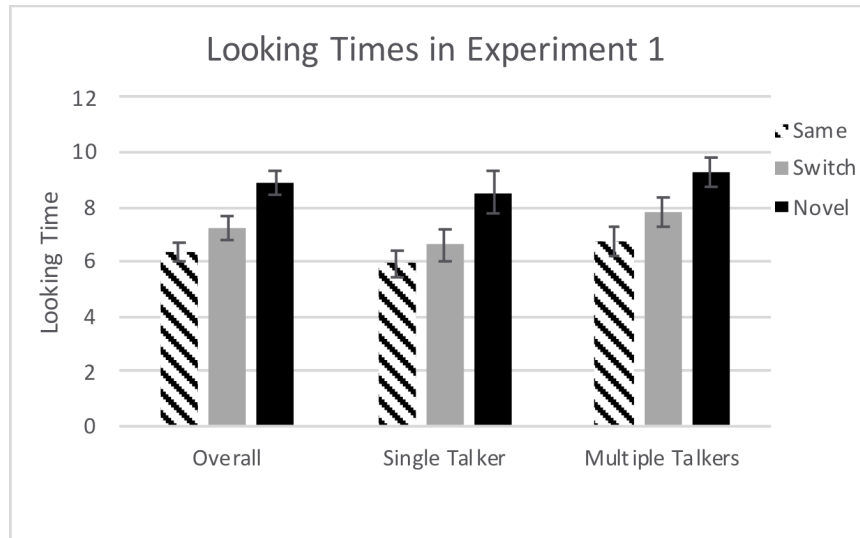


Figure 1: Mean Looking Times (With Standard-Error Bars) in Experiment 1.

Experiment 2

Experiment 2 tested discrimination of /n/ vs. /ŋ/ after habituation to a single talker or four talkers. We predicted, based on one previous study of infants' discrimination of /n/ vs. /ŋ/ (Narayan et al., 2010), that infants might fail to discriminate this contrast in the absence of talker variability. In the presence of talker variability, successful discrimination would be compatible with a facilitation effect.

Method

Participants

Inclusion criteria and consent procedures matched Experiment 1. We included 35 children (17 boys, 18 girls), divided between single-talker ($n = 18$) and multiple-talker conditions ($n = 17$). Within each condition, children were habituated to /nIm/ ($n = 21$; single-talker $n = 10$; multiple-

talker $n = 11$) or /ŋIm/ ($n = 14$; single-talker $n = 8$; multiple-talker $n = 6$). Nineteen participants were excluded for fussiness (11), experimenter error (four), failure to habituate (two), sleepiness (one), and distraction (one; due to an older brother in the room).²

Auditory Stimuli

To generate habituation and test stimuli, five new American-English speakers produced /nIm/ and /ŋIm/ in an infant-directed register. All talkers had training in phonetics, which was necessary for proper pronunciation, as /ŋ/ in onset position does not occur in English. Two were linguistics professors with emphases in phonetics, two were linguistics Ph.D. students with emphases in phonetics, and one was the first author. A phonetics professor (Dr. Natasha Warner) checked tokens of /ŋIm/ to ensure the velar nasal (/ŋ/) was correctly pronounced. One talker was selected for the test phase, three for the multiple-talker habituation, and one for the single-talker and multiple-talker habituation. Talkers were assigned to roles, as in Experiment 1, by comparing acoustic measurements (summarized in **Table 3**).

² It is interesting that rates of fussiness (11) and failure to habituate (two) in Experiment 2 were higher than Experiment 1 (six excluded for fussiness and zero failed to habituate). In a prior study (Quam, Knight, & Gerken, 2017), training that was more complex, due to pairing talker gender with words, led to more fussiness (23 children excluded of 59 tested, or 39%) than a training context that was simpler, containing talker variability that varied randomly (six children excluded of 24 tested, or 25%; see also Gerken, Wilson, & Lewis, 2005). In Experiment 2, both of the children who failed to habituate and 8/11 of the children excluded for fussiness were tested in the multiple-talker condition. These children were not over-represented in the training with the word /ŋIm/, suggesting it was not the non-native phoneme in particular that increased task difficulty. We suspect task complexity was increased by an additive effect of words containing two nasal consonants (n+m or ŋ+m) spoken by multiple talkers.

Table 3: Acoustic Measurements for Each Word Token Used in Experiment 2.

Word	Talker Set	Token	Pitch Mean (Hz)	Pitch Max	SD of Pitch Samples	F1	F2	Duration (ms)
<i>/nIm/</i>	<i>Single-Talker Habituation</i>	<i>1</i>	199	255	33	470	1732	1010
		<i>2</i>	197	259	35	486	1772	965
		<i>3</i>	210	294	48	517	1840	1054
	<i>Single/Multiple</i>	<i>4</i>	197	258	34	510	1716	1084
	<i>Multiple-Talker Habituation</i>	<i>1</i>	231	273	22	564	1964	880
		<i>2</i>	215	269	35	507	1574	813
		<i>3</i>	209	302	48	521	1882	1157
	<i>Test</i>	<i>1</i>	189	270	37	565	1800	691
		<i>2</i>	194	256	33	550	1759	640
		<i>3</i>	202	305	49	527	1813	666
		<i>4</i>	188	261	37	563	1824	640
	<i>Single-Talker Habituation</i>	<i>Mean (SD)</i>	201 (6)	267 (18)	38 (7)	496 (22)	1765 (55)	1028 (52)
	<i>Multiple-Talker Habituation</i>		213 (14)	276 (19)	35 (11)	526 (26)	1784 (174)	984 (163)
	<i>Test</i>		193 (6)	273 (22)	39 (7)	551 (17)	1799 (28)	659 (24)
<i>/yIm/</i>	<i>Single-Talker Habituation</i>	<i>1</i>	201	243	26	485	1781	996
		<i>2</i>	199	238	24	546	1825	1045
		<i>3</i>	199	246	33	445	1834	849
	<i>Single/Multiple</i>	<i>4</i>	207	265	34	457	1777	849
	<i>Multiple-Talker Habituation</i>	<i>1</i>	294	388	66	734	1846	1225
		<i>2</i>	206	251	28	610	1740	1226
		<i>3</i>	225	362	68	523	1793	976
	<i>Test</i>	<i>1</i>	217	325	60	517	1766	817
		<i>2</i>	201	259	31	653	1869	709
		<i>3</i>	209	298	46	616	1847	769
		<i>4</i>	195	247	29	512	1698	649
	<i>Single-Talker Habituation</i>	<i>Mean (SD)</i>	202 (4)	248 (12)	29 (5)	438 (45)	1804 (29)	935 (101)
	<i>Multiple-Talker Habituation</i>		233 (42)	317 (69)	49 (21)	581 (120)	1789 (44)	1069 (188)
	<i>Test</i>		206 (10)	282 (36)	42 (14)	575 (71)	1795 (78)	736 (73)

Apparatus and Procedure

All procedures matched Experiment 1.

Results and Discussion

Visual inspection of residuals and Shapiro-Wilk tests of normality, conducted separately for each trial type, revealed that residuals in the following conditions were not normally distributed: Switch, $W = 0.93$, $p = .027$, and Novel trials, also $W = 0.93$, $p = .027$. Upon visual

inspection, both trial types exhibited right-tailed distributions (positive skew). However, log transformation of looking times would not be appropriate, as residuals in Same trials were normally distributed. To avoid introducing bias by normalizing data inappropriately, we instead conducted planned comparisons using both parametric (t tests) and nonparametric tests (exact Fisher-Pitman permutation tests; Legendre & Legendre, 1998; see Quam et al., 2017, for similar use of these tests). The exact Fisher-Pitman permutation test involves first calculating the mean difference between groups, then scrambling the assignment of data-points to groups and recomputing the mean difference between groups for every possible permutation of the data. The p value indicates the fraction of permutations in which the difference between the group means exceeded the true difference between groups.

We first conducted ANOVAs, which are fairly robust to moderate non-normality (Glass, Peckham, & Sanders, 1972; Harwell, Rubinstein, Hayes, & Olds, 1992). An ANOVA on raw looking times with factors Trial Type (Same, Switch, Novel) and Talker Condition (Single Talker, Multiple Talkers) revealed no significant main effects or interactions. **Table 3** and **Figure 2** report mean looking times. An additional ANOVA checking for effects of additional variables Trained Word (/bIm/, /pIm/), Infant Gender (male, female), and Test-Trial Order (SWSW, WSWS, SWWS, WSSW) revealed no significant main effects of or interactions with these variables.

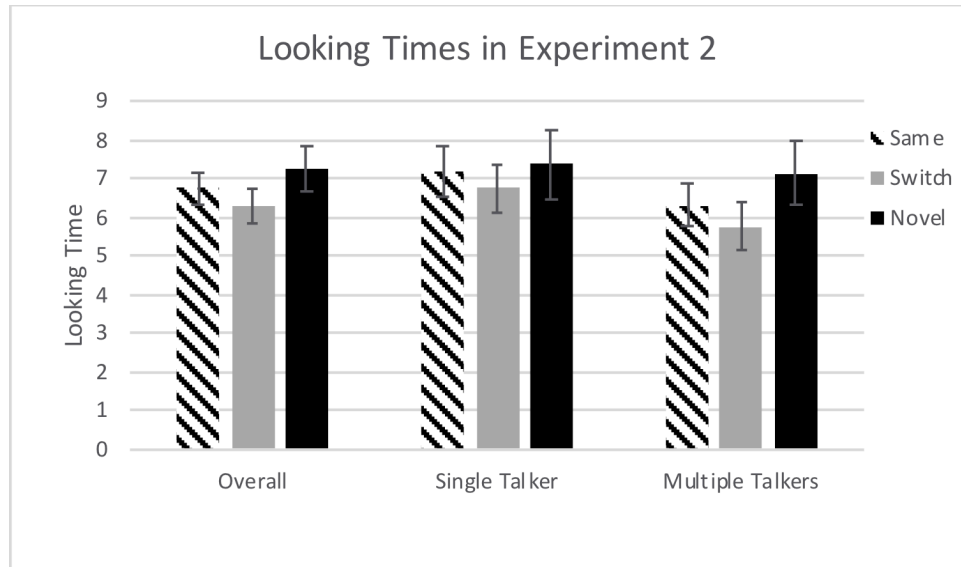


Figure 2: Mean Looking Times (With Standard-Error Bars) in Experiment 2.

The lack of greater looking times in Switch vs. Same trials indicates children did not discriminate /nIm/ vs. /ŋIm/. This result is not consistent with a facilitation effect of talker variability. However, infants also failed to significantly detect the Novel trial (i.e., the change from /nIm/ to /ŋaez/ or /ŋIm/ to /naez/; though looking times were numerically higher in Novel vs. Same or Switch trials; **Figure 2** and **Table 3**). Thus, one possible explanation for children's discrimination failure is that not all children were still attending to the experiment by the test phase. Habituation stimuli in Experiment 2, which contained (for some children, non-native) nasal onset consonants may have been more complex and therefore more taxing to attend to. This could explain the higher number of exclusions due to fussiness or failure to habituate in Experiment 2 vs. 1 (see Footnote 2). If this account is correct, children who *did* detect the Novel stimulus should successfully discriminate /nIm/ vs. /ŋIm/.

To investigate this possibility, an ANOVA included the additional factor Novelty Detection (infants who detected the novel stimulus, $n = 20$; vs. did not, $n = 15$). Infants were considered to have detected the Novel stimulus if they looked longer in the Novel trial vs. the mean

of Same and Switch trials. Because Novelty Detection was defined by looking times in Novel trials, these trials had to be excluded from the dependent variable, meaning the factor Trial Type had two levels (Same, Switch). Talker Condition was again included as a factor. The ANOVA did not reveal significant main effects. There was a significant interaction of Trial Type with Novelty Detection, $F(1,31) = 5.64, p = .024$. Follow-up comparisons indicated it was driven by significantly *lower* looking in Switch ($M = 6.1, SD = 3.1$) vs. Same trials ($M = 7.4, SD = 2.9$) for children who detected the Novel stimulus, $t(20) = -2.91, p = .009$, Fisher-Pitman $p = .016$, Cohen's $d = 0.65$. (Children who did not detect the Novel stimulus showed a non-significant tendency for longer looking in Switch, $M = 6.4, SD = 2.3$, vs. Same trials, $M = 6.0, SD = 2.1$, Cohen's $d = .20$). The Switch paradigm makes a clear directional prediction, so longer looking times in Same than Switch trials is not a meaningful looking pattern. As novelty detection was not linked with successful discrimination, failure to discriminate /nIm/ vs. /ŋIm/ cannot be explained by failure to stay focused on the task during test.

Why did more children not detect the Novel stimulus, as they had in Experiment 1? Because the /n/-/ŋ/ contrast is non-native in onset position and acoustically subtle, infants may have been less able to detect the change from /nIm/ to /ŋaez/ (or /ŋIm/ to /naez/) than from /bIm/ to /paez/ (or /pIm/ to /baez/). Children in Experiment 2 did not discriminate /nIm/ vs. /ŋIm/, so they were not discriminating /n/ vs. /ŋ/ in onset position. Thus, the Novel stimulus was likely less noticeably novel to infants in Experiment 2. It seems surprising that children would not detect the dramatic change in the nucleus and coda (from /-Im/ to /-aez/). However, this finding may be compatible with evidence that differences later in the word are less noticeable to infants than differences in onsets (Jusczyk, Goodman, & Baumann, 1999; Zamuner, 2006; Von Holzen, Nishibayashi, & Nazzi, 2018; see also Creel & Dahan, 2010; but see Swingley, 2009).

Another possible explanation for failure to discriminate /nIm/ vs. /ŋIm/ is that, despite children having numerically met the habituation criterion, some children could have habituated by chance (Oakes, 2010). This would mean they had not finished processing the habituation word and therefore should be less likely to detect a change. Our maximum number of habituation trials (24) was large compared to the 13-15 trials that has been suggested to minimize risk of habituating by chance (Dannemiller, 1984; Oakes, 2010). In our sample, $n = 10$ of $N = 72$ infants habituated in 16 or more trials, and eight of these were in Experiment 2. However, an additional Experiment 2 ANOVA excluding these eight infants still showed no evidence of discrimination. The ANOVA was modeled on the one reported previously that included factors Trial Type, Talker Condition, and Novelty Detection. It revealed the same interaction of Trial Type and Novelty Detection, $F(1,23) = 4.48, p = .045$. The group that detected the Novel stimulus still showed shorter looking times in Switch ($M = 5.9, SD = 3.0$) than Same trials ($M = 7.1, SD = 2.6$), $t(13) = -2.53, p = .025$, Fisher-Pitman $p = .048$, Cohen's $d = .68$. Thus, it seems unlikely that the lack of discrimination found in Experiment 2 could be driven by more children habituating by chance.

While it does not appear that infants habituating by chance could explain lack of discrimination in Experiment 2, patterns of habituation across the two experiments could shed light on infants' processing of stimuli. To that end, we conducted univariate ANOVAs on number of habituation trials and total habituation looking time across the two experiments, with Experiment (1, 2) and Talker Condition (Single Talker, Multiple Talkers) as predictors. Number of habituation trials did not significantly vary by Experiment or Talker Condition. However, an exploratory analysis employing Levene's test for equality of variances indicated variance was significantly greater in Experiment 2 (range: 7-23 trials) than 1 (range: 6-16 trials), $F = 6.76, p = .011$. The univariate ANOVA on total habituation looking indicated it was significantly greater in

Experiment 2 ($M = 106$ seconds, $SD = 47$) than 1 ($M = 87$ seconds, $SD = 31$), $F(1,68) = 4.17$, $p = .045$, Cohen's $d = .49$. Levene's test indicated variance was marginally greater in Experiment 2 (range: 26-235 seconds) than 1 (range: 33-169 seconds), $F = 3.70$, $p = .059$.

To determine whether longer and more variable habituation profiles were driven by habituation to the non-native onset consonant (/ŋ/), we conducted additional univariate ANOVAs on number of habituation trials and total habituation looking, with predictors Talker Condition (single, multiple) and Habituation Word (/nim/, /ŋim/). No significant effects emerged from either ANOVA. Thus, it appears phonotactic complexity of nasals in both onset and coda positions—more than the non-native onset consonant—drove longer and more variable habituation trajectories.

Children's failure to discriminate /nIm/ vs. /ŋIm/ does not seem to be attributable either to failure to remain on task or habituating by chance. Infants' failure to discriminate /n/ vs. /ŋ/ when habituated to a single talker was a predictable result. However, failure to discriminate when habituated to four talkers was not consistent with a facilitation effect. In the General Discussion below, we integrate findings from both experiments and consider their implications for theories of phonological learning and for future work.

General Discussion

The present study did not find significant impacts of talker variability on infants' sound discrimination. Experiment 1 tested discrimination of the native contrast /b/ vs. /p/. Children overall discriminated words and detected Novel stimuli. No effects of talker variability emerged in ANOVAs. However, the difference between Switch-trial and Same-trial looking times was numerically greater in the multiple-talker condition. Thus, the results from Experiment 1 are incompatible with an interference effect (as found in some conditions by Kuhl & Miller, 1982 and

Jusczyk, Pisoni, and Mullennix, 1992). Instead, they are compatible with null effects found with younger infants by Kuhl and Miller (1982) for vowel discrimination in the presence of pitch-contour variability, and by Jusczyk, Pisoni, and Mullennix (1992) for discrimination of consonants differing by one phonetic feature in the presence of talker variability (when no delay was introduced before test).

Experiment 2 tested discrimination of the non-native onset contrast /n/ vs. /ŋ/. Children overall failed to discriminate words. No effects of talker variability emerged, inconsistent with a facilitation effect of variability on a non-native, acoustically subtle contrast. Children also failed to detect the Novel stimulus (/naez/ for infants habituated to /ŋIm/; and /ŋaetz/ for /nIm/). Given children's inability to discriminate /n/ vs. /ŋ/, and the importance of onsets for word differentiation (Jusczyk, Goodman, & Baumann, 1999; Zamuner, 2006; Von Holzen, Nishibayashi, & Nazzi, 2018), it seems likely that the novel stimulus was more difficult to detect in Experiment 2 than 1.

One factor that could have impacted discrimination of /n/ vs. /ŋ/ was the introduction of a novel talker in the test phase. Using a single, novel talker equated the test phases between the two talker conditions (Quam, Knight, & Gerken, 2017; Gonzales, Gerken, & Gómez, 2018; Potter & Saffran, 2017). In both conditions, children had to generalize from the habituation talker(s) to the test talker. The multiple-talker group had to generalize from multiple talkers to a single talker. The single-talker group was highly familiar with a particular talker and therefore might be more likely to notice the test talker's novelty, which could have disrupted word recognition. However, all talkers were female. Talker changes impact processing more at this age when talker gender changes than when it does not (Houston & Jusczyk, 2000; see also Bergelson & Swingley, 2018).

Using a novel test talker unintentionally led test tokens to be shorter on average than habituation tokens. Recording tokens naturally introduced variation in durations. We attempted to

equate the experiments as much as possible. In both, test tokens were shorter on average than habituation tokens. However, this difference was numerically larger in Experiment 2. To compare experiments, we conducted a univariate ANOVA on durations with Experiment (1, 2) and Phase (habituation, test) as factors.³ The ANOVA revealed a significant main effect of Experiment, $F(1,44) = 12.89, p = .001$, reflecting overall longer durations in Experiment 2 ($M = 902$ ms., $SD = 186$ ms.) than 1 ($M = 749$ ms., $SD = 133$ ms.), a significant main effect of Phase, $F(1,44) = 41.5, p < .001$, reflecting shorter durations in test ($M = 674$ ms., $SD = 86$ ms.) than habituation ($M = 901$ ms., $SD = 163$ ms.), and a significant interaction of Experiment and Phase, $F(1,44) = 4.98, p = .031$, indicating the difference between habituation- and test-token durations was more pronounced in Experiment 2, $t(22) = 6.11, p < .001$, than 1, $t(22) = 2.99, p = .007$. Thus, test tokens were significantly shorter in both experiments, but this difference was more pronounced in Experiment 2, which could potentially have made it more difficult for infants to differentiate Same vs. Switch trials.

When the experiments are viewed together, one possibility that emerges is that talker variability may not impact sound discrimination at 7.5 months. The ANOVAs indicated no significant effects of talker variability in either experiment. Facilitative effects of variability on infants' phonotactic learning (Seidl, Onishi, & Cristia, 2014) and word learning (Rost & McMurray, 2009) and on adults' sound-category learning in some studies (e.g., Lively, Logan, & Pisoni, 1993) suggested we might find facilitation for /n/ vs. /ŋ/. It is possible that English-learning children's lack of exposure to syllable-initial /ŋ/ could have reduced facilitative effects of variability on /n/-/ŋ/. Future work could determine whether Filipino-learning 7.5-month-old babies might benefit more from variability. However, it should be noted that at this age, infants have not

³ One token of each word type was used in both single-talker and multiple-talker habituation sets, so it was included twice in the model input to reflect its frequency in trial orders.

yet undergone the perceptual reorganization that reduces discrimination of non-native consonants, so discrimination is still language-universal (e.g., Kuhl et al., 2006; Narayan et al., 2010). Infants also fail to show strong language-specific phonotactic processing until 9 months (Jusczyk et al., 1993).

The null effects found here for variability may reflect differences between sound discrimination and other phonological tasks, and thus may have implications for whether theoretical accounts of other phonological-learning tasks can be generalized to sound-category learning. In particular, while infant word-learning studies have demonstrated facilitative effects of variability (e.g., Rost & McMurray, 2009), in Apfelbaum and McMurray's (2011) model, variability operates on cue weights linked to visual objects. Sounds in the laboratory task used here are not paired with visual referents. Given prior evidence of impacts of variability on adult L2 speech-sound identification (e.g., Lively, Logan, & Pisoni, 1993; Antoniou & Wong, 2016) and infant sound discrimination (Kuhl & Miller, 1982; Jusczyk, Pisoni, & Mullennix, 1992), however, perhaps a different mechanism is needed to account for effects of variability in tasks that do not involve sound-object associations.

In real language learning, in contrast to laboratory tasks, infants hear sounds in words that often have visual referents. In naturalistic environments, sound discrimination and word learning could therefore be more similar—in terms of the role of visual referents—than in this particular, extensively used laboratory task, and variability could play a stronger facilitative role. In one laboratory study, Yeung and Werker (2009) found that 9-month-old English-learning infants only discriminated the non-native Hindi dental-retroflex contrast ([ɖa] vs. [ɗa]) after seeing distinct visual objects paired with tokens from each category. Talker variability could have a bigger impact on sound discrimination in a task that incorporates visual referents, like the one employed by

Yeung and Werker (2009). Of course, in real language input, infants hear abstract words that do not refer to concrete objects, but they do not comprise a large proportion of early vocabularies (Bates, Dale, & Thal, 1995). In addition, though parents frequently produce words when visual referents are not present, children may weight highly informative instances of words more highly than these less-informative instances (Medina, Snedeker, Trueswell, & Gleitman, 2011).

Another feature of our experimental design that could have limited impacts of variability is that the multiple-talker condition included only four female talkers in habituation, producing one word token each. However, each talker produced the syllable with different acoustic characteristics, which could theoretically help children rule out irrelevant acoustic dimensions and zero in on the contrastive dimension(s). To informally assess whether the four tokens in the multiple-talker habituation were more variable than in the single-talker habituation, **Tables 1 and 3** report standard deviations of acoustic measurements of habituation stimuli for each talker condition and word type. In every case, the multiple-talker set was numerically more variable than the single-talker set.

Some prior findings indicate four tokens from four female talkers could be sufficient variability. For example, Gerken and Knight (2015) found that 11-month-old infants generalized a phonological rule from only four examples. Seidl, Onishi, and Cristia (2014) found facilitation for infants' phonotactic learning after familiarization with just three female talkers, though 24 (pseudo)word types were included in both single-talker and multi-talker conditions. While variation in types could not have contributed directly to the facilitative effect of variability, it could have interacted with talker variability to boost learning. Facilitation might be more likely to emerge with a larger set of male and female talkers producing a larger set of tokens. Work on infants' word learning has used 18 male and female talkers (Quam et al., 2017; Rost & McMurray, 2009, 2010).

Another possible explanation for why we did not find impacts of talker variability is that the helpful aspects of talker variability for aiding formation and representation of sound-based categories (Singh, 2008) somehow interacted with the increase in task complexity introduced by the variability (e.g., Quam et al., 2017), resulting in a null effect at the group level. If some infants experienced facilitation from variability and others experienced interference, we would expect greater variance in the Switch vs. Same difference score across infants in the multiple-talker vs. single-talker condition. This is not the case for Experiment 1, where the standard deviation in Switch minus Same looking scores is actually lower in the multiple-talker ($SD = 2.0$) vs. single-talker condition ($SD = 2.4$). However, it could be the case in Experiment 2, where the standard deviation is numerically higher in the multiple-talker ($SD = 2.6$) vs. single-talker condition ($SD = 1.9$).

To verify that null effects of talker variability were not driven by an underpowered design, we conducted a power analysis based on a prior study by Quam, Knight, and Gerken (2017) that used similar stimuli and procedures. Briefly, Quam et al. replicated Rost and McMurray's (2009) word-learning effect when /buk/ and /puk/ were spoken by 18 talkers. However, when nine female talkers said one word and nine male talkers said the other, infants did not learn words. Quam et al. speculated that pairing talker genders and words introduced an additional correlated cue, which increased the task complexity, impairing learning (see also Gerken, Dawson, Chatila, & Tenenbaum, 2015).

Quam et al. found a significant Experiment by Trial Type interaction, reflecting an effect of talker distribution on word learning. We asked what overall sample size in each experiment (across two between-subjects groups measured in three trial types) would be necessary to reach 80% power to detect a significance level of $p < .05$. We reconstructed the partial η^2 for the

Experiment by Trial Type interaction, which was 0.07, indicating a medium-to-large effect size (Cohen, 1988). We entered the partial η^2 into G*Power (Faul, Erdfelder, Lang, & Buchner, 2007; 2009) to calculate Cohen's f , which was 0.28. To estimate the correlation across repeated measures, we used the mean of the three Pearson correlations between trial types, which was .37. As sphericity was not violated, we used a nonsphericity correction ϵ of 1. Results indicated the interaction between Experiment (i.e., talker condition) and Trial Type would be expected to reach 80% power with a total sample size of 30, indicating our total sample sizes of 37 in Experiment 1 and 35 in Experiment 2 were adequate.

The lack of robust effects of talker variability on discrimination in this study do not preclude the possibility that acoustic variability might impact discrimination at different ages and/or for different sound contrasts (Kuhl & Miller, 1982; Jusczyk, Pisoni, & Mullennix, 1992). Perhaps /b/ vs. /p/ is too well discriminated at this age, while /n/ vs. /ŋ/ is too difficult. Discrimination of the nasal contrast appears to have been especially difficult in the context of the /-Im/ coda, which contained another nasal. Future work could explore whether a contrast of intermediate difficulty would benefit more strongly from variability. It is possible that at the extremes of the continuum from ease to difficulty, variability does not impact processing, but in the middle, it would exert facilitation effects.

Perhaps older infants learning /n/-/ŋ/ as a native contrast might benefit more. One promising future direction is to test 8- to 10-month-old Filipino-learning babies on stimuli with /n/ and /ŋ/ onsets. The present results are consistent with prior findings that infants prior to 8 months cannot discriminate this contrast in a non-infant-controlled habituation paradigm (Narayan et al., 2010; but see Sundara et al., 2018), while we know infants older than 10 months learning Filipino can (Narayan et al., 2010). In the non-infant-controlled habituation paradigm, Filipino-learning

infants in the intermediate age range might fail to discriminate without variability but succeed with variability.

Conclusion

The present study found no significant impacts of talker variability on 7.5-month-olds' sound discrimination. This suggests that perhaps facilitative effects of variability on early word learning and phonotactic learning do not extend to sound discrimination. However, future work should probe for effects of variability under slightly different experimental conditions. Manipulating several experimental-design features could potentially enhance effects of variability. These include testing discrimination of different contrasts at different ages (in particular, we suggest testing 8- to 10-month-old Filipino-learning babies on /n/-/ŋ/); introducing more talkers, both male and female; and including visual referents (Yeung & Werker, 2009). We urge caution in introducing all these features simultaneously, however, as they could additively increase the task difficulty and thus increase the attrition rate due to fussiness and failure to habituate (Quam et al., 2017; see Footnote 2).

Acknowledgements

We thank the parents and infants who graciously participated in this study. We also thank members of the Tweety Language Development Lab at the University of Arizona, including Joleen Kuzdas and Leah Mann, who recruited and tested participants. Thanks to members of the linguistics department at the University of Arizona who served as voices for Experiment 2: Natasha Warner, Jessamyn Schertz, Diane Ohala, and Maureen Hoffman (Dr. Warner and Dr. Schertz also provided phonetics consultations). Several members of the Center for Research in Language at UC San

Diego served as voices for Experiment 1. We are grateful to members of the Child Language Learning Center at Portland State University, including Aminah Kariye, Eliza Minculescu, Rachel Atkinson, Anna Zhen, Josie Johnson, Abigail Tolomei, Liz Bort, and Molly Franz, who conducted reliability coding or otherwise assisted with manuscript preparation. Finally, we thank Rebecca Gómez and members of the Tigger Child Cognition Lab, particularly Courtney Meola and Elizabeth Salvagio Campbell for support with participant recruitment. Funding was provided by NIH grant F32HD065382 to CQ and NSF grant 0950601 to LAG. The authors declare no conflicts of interest with regard to the funding sources for this study.

References

- Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*, 1105–1138.
- Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher, P. & B. MacWhinney (Eds.), *The Handbook of Child Language* (pp. 95-151). Oxford: Basil Blackwell.
- Bergelson, E., & Swingle, D. (2018). Young infants' word comprehension given an unfamiliar talker or altered pronunciations. *Child Development*, *89*, 1567-1576.
- 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.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

- Cohen, L. B., Atkinson, D. J., & Chaput, H. H. (2004). *Habit X: A new program for obtaining and organizing data in infant perception and cognition studies [computer software]*.
- Creel, S. C., & Dahan, D. (2010). The effect of the temporal structure of spoken words on paired-associate learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*, 110–122.
- Dannemiller, J. L. (1984). Infant habituation criteria: I. A Monte Carlo study of the 50 % decrement criterion. *Infant Behavior and Development*, *7*, 147–166.
- Davis, A. (2015). *The interaction of language proficiency and talker variability in word learning* (Doctoral dissertation, University of Arizona, Tucson, Arizona, USA). Retrieved from <https://repository.arizona.edu/handle/10150/556484>.
- Eimas, P. D., Siqueland, E. R., Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science*, *171*, 303–306.
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149-1160.
- Fennell, C. T., & Waxman, S. R. (2010). What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*, *81*, 1376–1383.
- Fennell, C. T., & Werker, J. F. (2003). Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*, *46*, 245–264.

- Gerken, L.A., Dawson, C., Chatila, R., & Tenenbaum, J. (2015). Surprise! Infants consider possible bases of generalization for a single input example. *Developmental Science, 18*, 80–89.
- Gerken, L.A., Wilson, R., & Lewis, W. (2005). Infants can use distributional cues to form syntactic categories. *Journal of Child Language, 32*, 249-268.
- Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance. *Review of Educational Research, 42*, 237–288.
- Gonzales, K., Gerken, L.A., & Gómez, R. L. (2018). How who is talking matters as much as what they say to infant language learners. *Cognitive Psychology, 106*, 1-20.
- Gonzalez-Gomez, N., Poltrock, S., & Nazzi, T. (2013). A “bat” is easier to learn than a “tab”: Effects of relative phonotactic frequency on infant word learning. *PLoS ONE, 8*, e59601.
- Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. *Journal of Educational Statistics, 17*, 315–339.
- Höhle, B., Fritzsche, T., Meß, K., Phillip, M., & Gafos, A. (2020). Only the right noise? Effects of phonetic and visual input variability on 14-month-olds’ minimal pair word learning. *Developmental Science, e12950*. <https://doi.org/10.1111/desc.12950>
- Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1570–1582.

- Houston, D. M., & Jusczyk, P. W. (2003). Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 1143–1154.
- Jusczyk, P. W., Friederici, A. D., Wessels, J. M., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language*, *32*, 402-420.
- Jusczyk, P. W., Goodman, M.B., & Baumann, A. (1999). Nine-month-olds' attention to sound similarities in syllables. *Journal of Memory and Language*, *40*, 62-82.
- Jusczyk, P.W., Pisoni, D.B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*, *43*, 253-291.
- Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PLoS ONE*, *7*, e36399.
- Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*, 263-285.
- Kuhl, P. K., & Miller, J. D. (1982). Discrimination of auditory target dimensions in the presence or absence of variation in a second dimension by infants. *Perception & Psychophysics*, *31*, 279-292.
- Kuhl, P. K., 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.
- Legendre, P., & Legendre, L. F. J. (1998). *Numerical Ecology* (2nd English).

- Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America*, *94*, 1242–1255.
- Maye, J., Werker, J. F., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, *82*, B101–B111.
- Mazuka, R., Hasegawa, M., & Tsuji, S. (2014). Development of non-native vowel discrimination: Improvement without exposure. *Developmental Psychobiology*, *5*, 192–209.
- Medina, T. N., Snedeker, J., Trueswell, J. C., & Gleitman, L. R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 9014-9019.
- Narayan, C. R. (2008). The acoustic–perceptual salience of nasal place contrasts. *Journal of Phonetics*, *36*, 191–217.
- Narayan, C. R., Werker, J. F., & Beddor, P. S. (2010). The interaction between acoustic salience and language experience in developmental speech perception: Evidence from nasal place discrimination. *Developmental Science*, *13*, 407–420.
- Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognitive Development*, *11*, 255–268.
- Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*, *130*, 461-472.

- Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 421–435.
- Potter, C. E., & Saffran, J. R. (2017). Exposure to multiple accents supports infants' understanding of novel accents. *Cognition*, *166*, 67-72.
- 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.
- Quam, C., & Swingley, D. (2014). Processing of lexical-stress cues by young children. *Journal of Experimental Child Psychology*, *123*, 73-89.
- Quam, C., Knight, S., & Gerken, L. (2017). The distribution of talker variability impacts infants' word learning. *Journal of the Association for Laboratory Phonology*, *8*, 1–27.
- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*, 339–349.
- Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*, 608–635.
- Sadakata, M., & McQueen, J. M. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America*, *134*, 1324-1335.
- Sadakata, M., & McQueen, J. M. (2014). Individual aptitude in Mandarin lexical tone perception predicts effectiveness of high-variability training. *Frontiers in Psychology*, *5*, 1318.
- Seidl, A., Onishi, K. H., & Cristia, A. (2014). Talker variation aids young infants' phonotactic learning. *Language Learning and Development*, *10*, 297–307.

- Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*, 833–870.
- Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*, *51*, 173–189.
- Singh, L., White, K. S., & Morgan, J. L. (2008). Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition. *Language Learning and Development*, *4*, 157-178.
- Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, *34*, 1244-1286.
- Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, *388*, 381–382.
- Sundara, M., Ngon, C., Skoruppa, K., Feldman, N. H., Onario, G. M., Morgan, J. L., & Peperkamp, S. (2018). Young infants' discrimination of subtle phonetic contrasts. *Cognition*, *178*, 57-66.
- Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, *60*, 252-269.
- Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*, 850-855.
- van der Feest, S. V. H., & Johnson, E. K. (2016). Input-driven differences in toddlers' perception of a disappearing phonological contrast. *Language Acquisition*, *23*, 89–111.
- Von Holzen, K., Nishibayashi, L.-L., Nazzi, T. (2018). Consonant and vowel processing in word form segmentation: An infant ERP study. *Brain Sciences*, *8*, 24.

- Weatherhead, D., & White, K. S. (2016). He says potato, she says potahto: Young infants track talker-specific accents. *Language Learning and Development, 12*, 92-103.
- Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*, 197-234.
- Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development, 7*, 49-63.
- Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*, 1-30.
- Yeung, H. H., & Werker, J. F. (2009). Learning words' sounds before learning how words sound: 9-month-olds use distinct objects as cues to categorize speech information. *Cognition, 113*, 234-243.
- Yoshida, K. A., Fennell, C. T., Swingley, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-sounding words. *Developmental Science, 12*, 412-418.
- Zamuner, T. S. (2006). Sensitivity to word-final phonotactics in 9- and 16-month-old infants. *Infancy, 10*, 77-95.