

Young Infants' Word Comprehension Given An Unfamiliar Talker or Altered Pronunciations

Elika Bergelson
Duke University

Daniel Swingley
University of Pennsylvania

To understand spoken words, listeners must appropriately interpret co-occurring talker characteristics and speech sound content. This ability was tested in 6- to 14-month-olds by measuring their looking to named food and body part images. In the *new talker* condition ($n = 90$), pictures were named by an unfamiliar voice; in the *mispronunciation* condition ($n = 98$), infants' mothers "mispronounced" the words (e.g., *nazz* for nose). Six- to 7-month-olds fixated target images above chance across conditions, understanding novel talkers, and mothers' phonologically deviant speech equally. Eleven- to 14-month-olds also understood new talkers, but performed poorly with mispronounced speech, indicating sensitivity to phonological deviation. Between these ages, performance was mixed. These findings highlight the changing roles of acoustic and phonetic variability in early word comprehension, as infants learn which variations alter meaning.

Recognizing spoken words requires interpreting highly variable auditory signals. A stranger on the phone, a kindergartener hollering from outside, and a parent whispering all sound quite different, even if they use the same words. Yet as competent language users, we recognize what such talkers say.

This is not achieved by ignoring phonetic dimensions that signal talker identity. For instance, the critical acoustic resonances (F1/F2) of a man's /o/ vowel (*boat*) can be identical to those in a woman's /u/ vowel (*boot*), yet adults readily identify both words by both sexes (Hillenbrand, Getty, Clark, & Wheeler, 1995). Such findings imply that listeners partition phonetic variation appropriately between talker characteristics and speech sounds. Listeners retrieve both talker and message. Here, we examine how infants learn this partitioning by testing a large, racially and socioeconomically diverse sample of infants.

Infants' Speech and Word-Form Representations

Young infants begin learning words while still learning the phonetic categories (consonants and

vowels) of their language (words: Bergelson & Swingley, 2012; Tincoff & Jusczyk, 1999, 2012; phonetic categories: Kuhl et al., 2006; Polka & Werker, 1994; Werker & Tees, 1984, and others). Lacking robust speech sound representations, infants may recognize words based on *acoustic* differences from prior tokens, rather than *phonological* ones (e.g., Singh, Morgan, & White, 2004). If so, talker and phonological changes might each hinder comprehension similarly.

Infants begin to show speech sound categorization specific to their native language during the 1st year. That they do so using synthesized speech or unfamiliar talkers implies that infants generalize across talkers under some conditions (Johnson, Seidl, & Tyler, 2014; Kuhl, 1979; Kuhl et al., 2006; Polka & Werker, 1994; van Heugten & Johnson, 2012; Werker & Tees, 1984, inter alia). Eleven-month-olds also know the phonological forms of at least some words, as shown when they differentiate spoken lists of common words and slight phonological "mispronunciations" of those words (Hallé & de Boysson-Bardies, 1996; Swingley, 2005; Vihman, Nakai, DePaolis, & Hallé, 2004). Five-month-olds tested on changes to the initial sounds in their own name paint a mixed picture: French infants detected a sound change to their name but only for vowel-onset names; English infants did not (Bouchon,

The authors would like to acknowledge members of the Penn Infant Language Center, in particular Elizabeth Crutchley, as well as the infants and families involved in this research. This work was funded by NIH R01-HD049681 to Daniel Swingley and by NSF-IGERT/GRFP, NIH T32-DC000035, and DP5-OD019812-01 to Elika Bergelson.

Correspondence concerning this article should be addressed to Elika Bergelson, Psychology & Neuroscience, Box 90086, Duke University, Durham, NC 27708. Electronic mail may be sent to elika.bergelson@duke.edu.

© 2017 The Authors

Child Development © 2017 Society for Research in Child Development, Inc. All rights reserved. 0009-3920/2017/xxxx-xxxx

DOI: 10.1111/cdev.12888

2 Bergelson and Swingley

Floccia, Fux, Adda-Decker, & Nazzi, 2015; Delle Luche, Floccia, Granjon, & Nazzi, 2017). These studies used unfamiliar talkers, implying that when infants successfully discriminated, they transferred their phonological knowledge into expectations for new voices. However, none of these studies measured whether there is a *cost* for this cross-voice transfer.

Here, we investigated the effect of indexical and phonetic manipulations on young infants' word comprehension. Over 48 trials, infants heard sentences naming one element of a visual display. Word comprehension was operationalized as gaze to the named object. Infants as young as 6 months visually orient to named pictures when words are spoken by the infant's mother (Bergelson & Swingley, 2012, hereafter B&S12). However, previous works leaves unclear whether young infants understand words spoken by other talkers (yes: Tincoff & Jusczyk, 1999, 2012; no: Parise & Csibra, 2012).

Among older children, Hallé and de Boysson-Bardies' (1996) "intentional mispronunciation" has been used with word comprehension to evaluate children's knowledge of phonological form (Mani & Huettig, 2012; Mani & Plunkett, 2007, 2010; Swingley & Aslin, 2000; Yoshida, Fennell, Swingley, & Werker, 2009). These studies generally find that after age 1, children look at named pictures less when labels are mispronounced (e.g., *cur* for *car*). This shows that their phonetic knowledge, as realized by an unfamiliar talker, is sufficiently detailed to match canonical forms better than slightly deviant forms. However, these experiments have not compared variation due to talker identity and variation due to phonological changes.

Present Research

Here we examine how well infants understand words for common nouns when they are said by an experimenter (*new talker condition*) or when a familiar talker (their parent) deliberately mispronounces them (*mispronunciation condition*). The new talker condition tests the degree to which infants' lexical representations are talker independent. The mispronunciation condition tests infants' word-form precision, that is, the degree to which single phoneme deviations disrupt comprehension. We make use of B&S12's data as a comparison case for examining newly collected data that alters talker or phonetic characteristics.

In our mispronunciation condition, we manipulated vowels because they are generally more tightly linked with talker variation than consonants

(e.g., imagine identifying someone from her [s] vs. her [a]). Furthermore, infants' attention to vowels has been clearly demonstrated by 6 months, both in word-form recognition tasks (Bouchon et al., 2015; cf. Delle Luche et al., 2017) and native sound category learning (Kuhl et al., 2006; Polka & Werker, 1994).

Mature listeners recognize words spoken by unfamiliar talkers very well. Indeed, adults generally understand new talkers (Pierrehumbert, 2016), though subtle hindrances can be detected (Walker & Hay, 2011). By contrast, intentional mispronunciations strongly impede word recognition in adults (Swingley, 2009).

How word recognition is tuned over the course of the 1st year is not clear, given the inchoate nature of the early phonology and prior demonstrations that nonphonetic modifications of word forms can prevent recognition (Johnson, Westrek, Nazzi, & Cutler, 2011; Schmale, Cristià, Seidl, & Johnson, 2010; Singh et al., 2004). In principle, there are four possible patterns, as follows:

1. *Phonemes over talkers* (adult pattern): Even 6-month-olds might already weigh phonological variation over talker familiarity, recognizing new talker words more readily than mispronounced words.
2. *Overly strict variability weighting*: Infants might start out with strict matching criteria, failing to understand unfamiliar talkers or deviant pronunciations, only understanding words said in the exact manner most familiar to them.
3. *Overly loose variability weighting*: Infants might begin with relatively loose criteria over both sorts of variation, succeeding in both conditions.
4. *Talkers over phonemes*: Infants might show a clearly unadult pattern, weighing talker familiarity over phonological variation. This would cause degraded performance for new talkers but not for mispronounced words.

We employed a between-subjects design with 6- to 14-month-olds. For each condition we asked two questions: (a) whether infants understood words with the introduced sound manipulations (talker or phonetic change) and (b) whether infants' comprehension was worse than when the words were pronounced correctly by their mother (B&S12).

We expected that by 14 months, infants would find words spoken with the wrong vowels hard to understand relative to words spoken correctly by an unfamiliar talker (Swingley & Aslin, 2002). We

did not have a prediction about when this pattern would emerge in early development. Because our goal was to characterize a possible developmental change, we used identical procedures and materials in a large sample over a wide age range.

Method

Participants

Subjects were one hundred eighty-eight 6- to 14-month-olds. We modeled age as a continuous predictor but also split infants into three age groups for ease of comparison with other studies. Our youngest group was 6- to 7-month-olds, matching the youngest group examined by B&S12; our eldest group was 11- to 14-month-olds, whom we had reason to believe would be affected by phonological mispronunciations based on previous studies (e.g., Swingley & Aslin, 2002; Vihman et al., 2004). Although B&S12 examined 6- to 9-month-olds as a group, they also considered 6- to 7- and 8- to 9-month-olds separately. We use this narrower youngest group here because of increasing interest in very early lexical development (e.g., Bouchon et al., 2015).

The final sample consisted of fifty-four 6- to 7-month-olds, fifty-six 8- to 10-month-olds, and seventy-eight 11- to 14-month-olds, split approximately evenly across the new talker and mispronunciation conditions (Table 2). Infants were recruited from the Philadelphia area by mail, e-mail, phone, and in person. All children were carried full-term (estimated as > 34 weeks from conception) and heard 75% or more English at home. None had chronic ear infections; 43% were girls. An additional 12 infants were tested but were excluded from the final sample for the following reasons: technical problems ($n = 3$), hearing < 75% English at home ($n = 6$), premature birth status ($n = 3$), age outside of tested range ($n = 3$). Finally, 79 infants were excluded for not contributing data to at least half of item pairs tested (46 in the new talker condition, 33 in the mispronunciation condition). For this last criterion, we first removed trials in which infants (a) failed to look at both displayed images, (b) failed to look at either image for more than one third of the window of interest, and (c) were crying or screaming during the target sentence. We removed trials in which (a) parents failed to say the target word as intended, (b) parents pointed to a specific part of the screen while peeking under the visor, or (c) experimenters failed to log a key press when the target was said. Such events tended to co-occur in trials where infants were fussy.

Parents gave written informed consent on behalf of their infant. Maternal education and race information is available in the Supporting Information. Data were collected between December 2012 and June 2015.

Materials

Visual materials were identical to those in B&S12. There were 48 test trials, split into two interspersed trial types, along with colorful 2 s attention getters between every eight test trials. On *paired picture* trials ($n = 32$), infants saw two images, one food-related item and one body part, against a gray background. Each image was 16.9×12.7 cm, placed on the left or right side of the screen. On *scene* trials ($n = 16$), infants saw a scene of food-related items (one of two different tabletops with four food items on each) or a person (an image of a whole body, or just the face). These images varied slightly in height and width, and were presented in the center of the screen. Whereas the paired picture trials are standard in the field (Fernald, Pinto, Swingley, Weinberg, & McRoberts, 1998), the scene trials were a novel addition in B&S12; they were retained in the current study so that the only change from the previous experiment would be the talker and pronunciation changes.

In the new talker condition, auditory materials were identical to B&S12 except that the (female) experimenter produced the utterances instead of the infant's mother (Figure 1). In the mispronunciation condition, 6.1% of infants heard the words from their father rather than their mother (in B&S12, it was 7.5%; we use "mother" hereafter). In the mispronunciation condition, each noun was deliberately mispronounced by the mother with a change in the stressed vowel. The vowel shift was maximally different from the original vowel, did not create a real word known to infants, and was not similar to the vowel in any competitor words displayed simultaneously (Table 1).

Apparatus and Procedure

After completing paperwork, which included the long-form words and gestures version of the MacArthur Communicative Development Inventory (Dale & Fenson, 1996) and an optional demographic questionnaire, parents and infants were seated in front of a computer display (a 34.7×26.0 cm LCD 75dpi screen) in a dimly lit room. In both conditions (as in B&S12), the talker wore headphones over which she heard a sentence prompt, which she

4 Bergelson and Swingley

Table 1
Word Stimuli

Original	Mispronunciation [IPA Transcription]
Apple	opal [opəl]
Banana	banoona [bənunə]
Bottle	biddle [bɪdl]
Cookie	khaki [kaki]
Juice	jouse [dʒaʊs]
Milk	mulk [mʌlk]
Spoon	spoon [spun]
Yogurt	yaygurt [jeˈgɔrt]
Ear	or [ɔr]
Eyes	A's [eɪz]
Face	fouse [faʊs]
Foot/feet	foat [fot]
Hair	har [hɑr]
Hand/hands	hund/s [hʌnd/z]
Leg/legs	loog/s [luɡ/z]
Mouth	mith [mɪθ]
Nose	nazz [næz]

Note. "Original" indicates the word in B&S12, and the new talker condition. "Mispronunciation" reflects the pronunciation of that same word in the mispronunciation condition, with International Phonetic Alphabet (IPA) notation in brackets.

repeated to the child. Mothers wore an opaque visor; they could see their child but not the display screen. In the new talker condition, an experimenter wore the headphones and repeated the sentence aloud to the child, from behind the experimenter's computer; infants remained in mothers' laps (Figure 1).

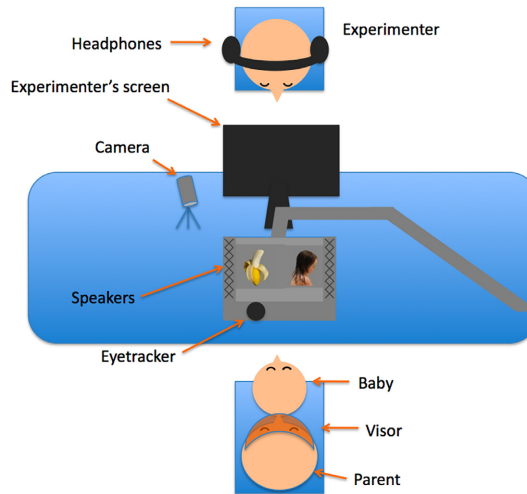


Figure 1. Experimental setup, new talker condition. Infants sat in mothers' laps; an experimenter sat behind a second computer. The experimenter heard sentence prompts over headphones that she repeated to the child. Mothers wore a visor; they could not see the displayed images. The setup for the mispronunciation condition and B&S12 was identical, except the mother wore the headphones.

In the mispronunciation condition, mothers deliberately mispronounced the target items, repeating what they heard in their headphones. Before the experiment mothers listened to the target words (with no carrier phrase) over closed ear headphones in the laboratory's waiting room. While listening, they saw a list written out in a phonetically transparent way (see Table 1) along with a "sounds like" aid, for example, for the mispronunciation of "nose" they saw "nazz, rhymes with jazz." Mothers did not see the original word (to help prevent accidental correct pronunciations during testing). Occasionally, during the experiment mothers mispronounced the mispronunciation. Based on experimenter notes and review of the session's video footage, we removed trials in which mother's mispronunciation (a) lacked the initial consonant (e.g., "ouse" for "fouse" [face]), (b) resulted in a phonetic change larger than the intended change (e.g., "hines" for "hunds" [hands]), or (c) turned the word into a word infants might know (e.g., "more" for "oar" [ear]). This resulted in 45 trials being removed across 30 babies. In the new talker condition, the experimenter was a native English-speaking female researcher; she did not mispronounce target words. The experimental setup and structure were identical to that in B&S12, except the auditory materials.

Results

We analyze results from each trial type (paired picture and scene trials) using the same approach described in B&S12. Fixation results on paired picture trials underwent a difference score analysis, which considered the proportion of time (over the 367–3,500 ms window of interest) infants looked at an image when it was the target versus when it was the distracter. This measures whether infants recognized words in each pair, controlling for infants’ picture preferences. For scene trials, the outcome measure was the proportion of target looking, corrected for baseline looking via subtraction, over the 367–3,500 ms window of interest. The present criteria for subject and trial inclusion (see “Participants”) were retroactively applied to B&S12, so that the samples compared below were the result of the same selection process; 57 infants from B&S12 are included for cross-study analyses here (Table 2).

Given added analytic complexities in the analysis of the scene trials, which were retained to maximally equate the present study design with B&S12, we present those results in Supporting Information, focusing here on the more standard paired picture trials (Fernald, Zangl, Portillo, & Marchman, 2008; Fernald et al., 1998). All raw anonymized data and an R script showing how statistics and analyses were generated are available on the first author’s website.

Analysis of Covariance for Age and Study

New Talker and Mispronunciation Conditions

A two-way analysis of covariance testing for an interaction between (centered) age as a continuous predictor and condition (new talker, mispronunciation) accounted for a significant proportion of

variance, adjusted $R^2 = .036$, $F(3, 184) = 3.3$, $p = .022$. There was no main effect of age, a marginal effect of condition, new talker > mispronunciation, $F(3, 184) = 3.47$, $p = .064$, partial $\eta^2 = .019$, and a significant interaction: infants in the new talker condition performed better as age increased, $F(3, 184) = 4.13$, $p = .043$, partial $\eta^2 = .022$. Inclusion of an interaction term was justified by model comparison ($\Delta Ss = .063$; $p = .042$ by chi-square test). Thus, there were developmental differences in how children responded to talker changes and mispronunciations. We also found that age and vocabulary were strongly and significantly correlated across all vocabulary measures (see Supporting Information for details).

Comparison With B&S12

We next combined our new results with the original B&S12 data and performed the same analysis, now with three levels of “condition” (original B&S12, new talker, mispronunciation). The overall model accounted for a significant proportion of variance, adjusted $R^2 = .056$, $F(3, 184) = 3.89$, $p = .0021$. Here, a Condition \times Age interaction was not justified by model comparison ($p = .124$) nor was such an interaction significant when included (.126 by chi-square test). We found a main effect of condition, $F(5, 239) = 5.87$, $p = .0033$, partial $\eta^2 = .047$, and a marginal effect of (centered) age, $F(5, 239) = 3.53$, $p = .062$, partial $\eta^2 = .012$.

Analysis by Age Group

We next considered each age group separately (see Figures 2–4). First, we analyzed subject means and item pair means in the current experiment

Table 2
Performance by Condition and Age Group

Age group (months)	Unit	No. of positive/no. of total subjects, item pairs			Kruskal–Wallis χ^2 (p value)
		Original (%)	Mispronunciation (%)	New talker (%)	
6–7	Subjects	13/17 (76)	20/30 (67)	15/24 (63)	2.0 (.36)
	Item pairs	7/8	6/8	7/8	1.1 (.58)
8–10	Subjects	15/18 (83)	16/32 (50)	11/24 (46)	5.1 (.076)
	Item pairs	5/8	4/8	3/8	2.2 (.33)
11–14	Subjects	16/22 (73)	20/36 (56)	32/42 (76)	6.3 (.041)
	Item pairs	8/8	6/8	8/8	7.0 (.031)

Note. “Original” refers to B&S12, where mothers pronounced words correctly. Columns 2–4 indicate how many participants (or item pairs) had positive means, over the total number of participants (item pairs.) Column 5 indicates the Kruskal–Wallis test outcome, testing for condition effects. Only the oldest age group (bottom row) showed evidence for condition differences in this test.

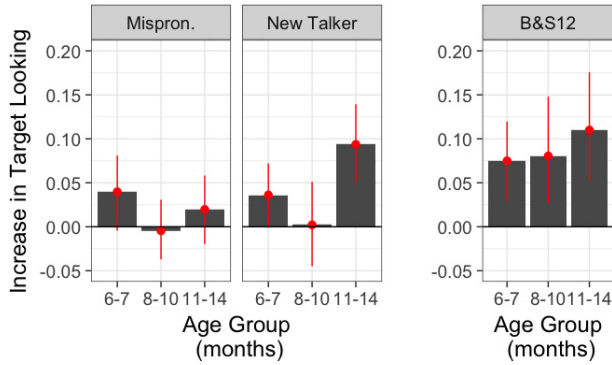


Figure 2. Increase in target looking by age group and condition. The panels show average performance across infants by age group in the mispronunciation condition, new talker condition, and B&S12, left to right. The y-axis indicates infants' subject mean difference scores in the 367–3,500 window of interest across item pairs; error bars are 95% nonparametric bootstrapped CIs.

against chance using two-tailed Wilcoxon tests. Next, combining with the B&S12 data, we ran Kruskal–Wallis tests to see whether there were differences among the three conditions. Finally, we examined all pairwise comparisons among conditions using two-tailed Wilcoxon tests (two sample or paired).

6- to 7-Month-Olds

Across new talker and mispronunciation conditions, the 6- to 7-month-olds looked at the named image at above-chance rates. Of 54 infants, 35 attained positive mean difference scores ($M = .038$, $SD = .11$, $p < .05$ by binomial and Wilcoxon test;

patterns of significance were equivalent over item pairs). See Table 2. Although most infants showed positive performance for most pairs (Table 2), subject and item pair means in each condition taken individually did not differ from chance (new talker: $M_{6-7Ss} = .036$, $SD = .098$; mispronunciation: $M_{6-7Ss} = .04$, $SD = .12$; $ps > .05$).

Although difference scores over the whole sample were not normally distributed, the 6- to 7-month subject means did not differ from a normal distribution by Shapiro test (each $p > .05$). Given this, and our directional hypothesis that infants would either look more at the target image, or show no preference for either image, we also conducted one-tailed t tests on looking scores in each condition. These tests show above-chance performance for the 6- to 7-month-olds in each condition (each $M = .04$, $p < .05$).

Considering all three conditions, Kruskal–Wallis test and pairwise comparisons over subject means and item pair means indicated no significant differences across any pair of conditions (all $ps > .16$ by Wilcoxon test). This pattern suggests that 6- to 7-month-olds recognized words equally whether said by their mother or another woman, or whether pronounced correctly or with a changed vowel. Performance at this age was modest but consistent, likely reflecting the somewhat fragile nature of very early word comprehension.

8- to 10-Month-Olds

The 8- to 10-month-olds showed surprisingly poor performance in the current study. Collapsing

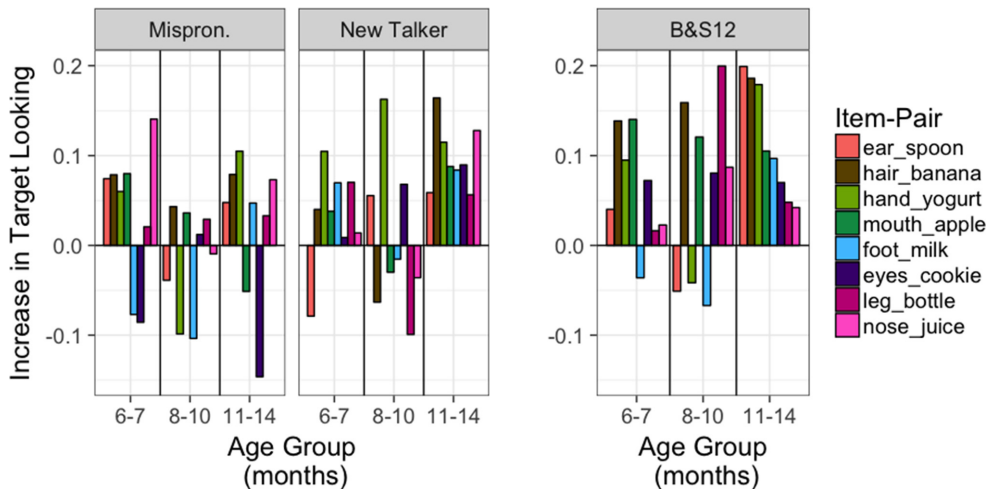


Figure 3. Increase in target looking across item pairs, age groups, and conditions. The panels show average performance in each age group for each item pair in the mispronunciation condition, new talker condition, and B&S12, left to right. The y-axis indicates mean difference scores for a given age group and item pair in the window of interest (367–3,500 ms after target onset). Bar color indicates target word; words are ordered on the plots as in the legend.

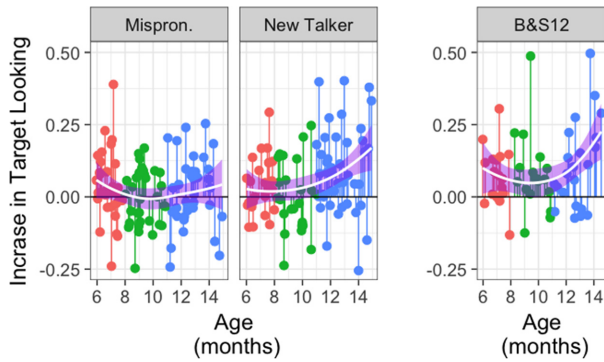


Figure 4. Subject means. The panels show average performance for each infant in the mispronunciation condition, new talker condition, and B&S12, left to right. The *y*-axis indicates infants' subject mean difference scores in the window of interest (367–3,500 ms after target onset) across item pairs. Lollipop color indicates age group. The white line and purple confidence band indicate a smoothed loess (local estimator) fit over the data (span = 2).

conditions, 27 of the 56 infants attained positive mean difference scores ($M = -.0017$, $SD = .11$, $p > .05$). After pooling with the B&S12 data, a Kruskal–Wallis test revealed no significant effect of condition over subjects or item pairs. Over subjects, two-sample Wilcoxon tests indicated poorer performance in each new condition than in the original study (B&S12 > new talker: estimated difference = .072, $p = .048$; B&S12 > mispronunciation: estimated difference = .068, $p = .039$) and no difference between the two new conditions (new talker: $M_{8-10\text{Ss}} = .0022$, $SD = .12$; mispronunciation: $M_{8-10\text{Ss}} = -.0045$, $SD = .098$; $p\text{s} > .05$). Over item pairs, performance between the original study and mispronunciation condition differed significantly by Wilcoxon test (B&S12 > mispronunciation, estimated difference = .077, $p = .016$), but not the original and new talker or new talker and mispronunciation conditions ($p > .05$). Infants in this age group struggled in the present study; we return to this in the Discussion.

11- to 14-Month-Olds

Finally, the 11- to 14-month-olds showed the overall pattern across conditions that would be expected from mature listeners. Collapsing over the two new conditions, 52 of the 78 infants attained positive mean difference scores ($M = .06$, $SD = .14$, $p \leq .001$). This strong performance was driven by the infants in the new talker condition: 8/8 pair means and 32/42 subject means were positive in

this condition; performance in this condition alone was significant by Wilcoxon and binomial tests (new talker $M_{11-14\text{Ss}} = .094$, $SD = .15$, $p \leq .001$). In contrast, 6/8 pair means and only 20/36 subject means were positive in the mispronunciation condition (mispronunciation $M_{11-14\text{Ss}} = .02$, $SD = .12$, $p = .315$). Combining with the B&S12 data, a Kruskal–Wallis test indicated a significant difference across the three conditions (Kruskal–Wallis $\chi^2 = 7$, $p = .031$). Infants' performance in B&S12 and the new talker condition were each significantly higher than in the mispronunciation condition, across both subjects and item pairs (B&S12 > mispronunciation: estimated difference = .079; $p = .042$) by Wilcoxon test; new talker > mispronunciation: estimated difference = .069; $p = .024$); B&S12 and new talker did not differ significantly ($p = .95$); see Figures 2–4).

Discussion

At 6–7 months infants understood words equally when said by their own mother correctly, when said incorrectly, and when said by a new talker. In these young infants, performance was above chance, but modest, across all conditions. In contrast, by 11–14 months, infants' comprehension was impaired for phoneme changes but robust across talker changes.

Studies of infant memory have shown that minor changes to relevant or irrelevant features of a training situation can interfere with recall (e.g., Rovee-Collier, Schechter, & Shyi, 1992). Indeed, in discrimination studies infants have failed to recognize trained syllables given nonphonetic alterations (e.g., Singh et al., 2004; cf. van Heugten & Johnson, 2012). Before we tested younger infants, it was equally plausible that indexical and phonetic alterations would affect word comprehension differently. Our results suggest that this is not the case.

Unexpectedly, infants' performance at 8–10 months was at chance in both conditions. In B&S12, 8- to 10-month-olds performed as well as the 6- to 7-month-olds had. Statistical comparisons between the present conditions and B&S12 presented a mixed picture (see Table 2), so we cannot be certain that there is truly a difference between the original and new results, that is, we may be simply failing to replicate a weak effect at this age. Alternatively, 8- to 10-month-olds' poor performance in the current study may reflect a genuine linguistic change aligned with phonetic reorganization. As infants converge on their language's speech sounds, they may recognize that broad acoustic similarity is not an

appropriate criterion and instead use overly stringent matching (Werker & Curtin, 2005). This account is consistent with recent EEG (electroencephalography) results that found 9-month-olds only responded to a labeling mismatch when the speaker was their mother rather than an experimenter (Parise & Csibra, 2012). This possibility is intriguing, though it is unclear why a change in phonetic representation would lead infants to be more sensitive to talker variation rather than less. Future work could examine this by studying the same infants longitudinally using the present manipulations.

Our 6- to 7-month-old results suggest that whether infants readily abstract away from a talker's idiosyncratic voice characteristics or not, their speech-to-lexicon matching ability is quite (even overly) flexible, at least under the present referential conditions. More extreme manipulations in either domain might render a different pattern of results.

Not until 11–14 months did infants show the adult-like pattern: unhindered comprehension of a new talker (concordantly with previous studies: Mani & Plunkett, 2010; Swingley & Aslin, 2000, 2002) and poor performance with mispronounced words. This suggests that infants must learn which sources of variation are linguistically meaningful. Perhaps this occurs as infants gain experience with multiple talkers (e.g., Rost & McMurray, 2009, 2010). Given that the talkers infants hear use similar words, infants could simplify their representations by discovering the phonetic function that relates, say, their mother's lexicon with their father's. Such a discovery could be mediated by meaning (mommy's and daddy's *apple* both refer to the fruit) or just the word forms alone (Seidl, Onishi, & Cristià, 2014). Simultaneous manipulation of both indexical and phonetic properties could elucidate this.

Young infants' understanding of words produced by an unfamiliar talker could mean that infants have learned to generalize over talkers. Other infants drawn from a similar population heard, on average, 60% of common word instances from their mother (Bergelson & Aslin, under review), so if infants had talker-specific representations, we would expect infants to perform better in B&S12 than here. Alternatively, together with infants' performance on mispronounced vowel words, the present results may indicate that infants' matching criteria are sufficiently lax to match a wide range of pronunciations to their early representations.

Taken together, our results suggest that speech sound reorganization occurs in tandem with lexical growth over the first postnatal year. Although

infants might reveal a performance cost with more thorough individual testing, here infants were beholden to highly familiar tokens during early word comprehension. Infants' lexical representations allow them to connect novel word tokens to the conceptual categories they denote at the same (modest) levels demonstrated for correctly pronounced maternal speech. Between 6 and 14 months, they reorganize their speech sound interpretation to continue accepting words spoken with new voices but to exclude realizations that deviate from the conventional form in their phonology.

References

- Bergelson, E., & Aslin, R. (under review). *Early (noun) relations: Cross-word and home-lab linkages in 6-month-olds*.
- Bergelson, E., & Swingley, D. (2012). At 6–9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences of the United States of America*, 109, 3253–3258. <https://doi.org/10.1073/pnas.1113380109>
- Bouchon, C., Floccia, C., Fux, T., Adda-Decker, M., & Nazzi, T. (2015). Call me Alix, not Elix: Vowels are more important than consonants in own-name recognition at 5 months. *Developmental Science*, 18, 587–598. <https://doi.org/10.1111/desc.12242>
- Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, & Computers*, 28, 125–127. <https://doi.org/10.3758/BF03203646>
- Delle Luche, C., Floccia, C., Granjon, L., & Nazzi, T. (2017). Infants' first words are not phonetically specified: Own name recognition in British English-learning 5-month-olds. *Infancy*, 22, 362–388. <https://doi.org/10.1111/inf.12151>
- Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. *Psychological Science*, 9, 228–231. <https://doi.org/10.1111/1467-9280.00044>
- Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. *Developmental Psycholinguistics: On-Line Methods in Children's Language Processing*, 2008, 97–135. <https://doi.org/10.1017/CBO9781107415324.004>
- Hallé, P. A., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. *Infant Behavior and Development*, 19, 463–481. [https://doi.org/10.1016/S0163-6383\(96\)90007-7](https://doi.org/10.1016/S0163-6383(96)90007-7)
- Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, 97, 3099–3111. <https://doi.org/10.1121/1.411872>

- Johnson, E. K., Seidl, A., & Tyler, M. D. (2014). The edge factor in early word segmentation: Utterance-level prosody enables word form extraction by 6-month-olds. *PLoS ONE*. Retrieved from <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0083546>
- Johnson, E. K., Westrek, E., Nazzi, T., & Cutler, A. (2011). Infant ability to tell voices apart rests on language experience. *Developmental Science*, *14*, 1002–1011. <https://doi.org/10.1111/j.1467-7687.2011.01052.x>
- Kuhl, P. K. (1979). Speech perception in early infancy: Perceptual constancy for perceptually dissimilar vowel categories. *The Journal of the Acoustical Society of America*, *66*, 1168–1679.
- 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*. <https://doi.org/10.1111/j.1467-7687.2006.00468.x>
- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 843–847. <https://doi.org/10.1037/a0029284>
- Mani, N., & Plunkett, K. (2007). Phonological specificity of vowels and consonants in early lexical representations. *Journal of Memory and Language*, *57*, 252–272. <https://doi.org/10.1016/j.jml.2007.03.005>
- Mani, N., & Plunkett, K. (2010). Twelve-month-olds know their cups from their keps and tups. *Infancy*, *15*, 445–470. <https://doi.org/10.1111/j.1532-7078.2009.00027.x>
- Parise, E., & Csibra, G. (2012). Electrophysiological evidence for the understanding of maternal speech by 9-month-old infants. *Psychological Science*, *23*, 728–733. <https://doi.org/10.1177/0956797612438734>
- Pierrehumbert, J. B. (2016). Phonological representation: Beyond abstract versus episodic. *Annual Review of Linguistics*, *2*, 33–52. <https://doi.org/10.1146/annurev-linguist-030514-125050>
- 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. <https://doi.org/10.1037/0096-1523.20.2.421>
- Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental Science*, *12*, 339–349. <https://doi.org/10.1111/j.1467-7687.2008.00786.x>
- Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*, *15*, 608–635. <https://doi.org/10.1111/j.1532-7078.2010.00033.x>
- Rovee-Collier, C., Schechter, A., & Shyi, G. C. (1992). Perceptual identification of contextual attributes and infant memory retrieval. *Developmental*. Retrieved from <http://psycnet.apa.org/journals/dev/28/2/307/>
- Schmale, R., Cristià, A., Seidl, A., & Johnson, E. K. (2010). Developmental changes in infants' ability to cope with dialect variation in word recognition. *Infancy*, *15*, 650–662. <https://doi.org/10.1111/j.1532-7078.2010.00032.x>
- Seidl, A., Onishi, K. H., & Cristià, A. (2014). Talker variation aids young infants' phonotactic learning. *Language Learning and Development*, *10*, 297–307. <https://doi.org/10.1080/15475441.2013.858575>
- 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. <https://doi.org/10.1016/j.jml.2004.04.004>
- Swingle, D. (2005). 11-month-olds' knowledge of how familiar words sound. *Developmental Science*, *8*, 432–443. <https://doi.org/10.1111/j.1467-7687.2005.00432.x>
- Swingle, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language*, *60*, 252. <https://doi.org/10.1016/j.jml.2008.11.003>
- Swingle, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition*, *76*, 147–166. [https://doi.org/10.1016/S0010-0277\(00\)00081-0](https://doi.org/10.1016/S0010-0277(00)00081-0)
- Swingle, D., & Aslin, R. N. (2002). Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*, *13*, 480–484. <https://doi.org/10.1111/1467-9280.00485>
- Tincoff, R., & Jusczyk, P. W. (1999). Some beginnings of word comprehension in 6-month-olds. *Psychological Science*, *10*, 172–175. <https://doi.org/10.1111/1467-9280.00127>
- Tincoff, R., & Jusczyk, P. W. (2012). Six-month-olds comprehend words that refer to parts of the body. *Infancy*, *17*, 432–444. <https://doi.org/10.1111/j.1532-7078.2011.00084.x>
- van Heugten, M., & Johnson, E. K. (2012). Infants exposed to fluent natural speech succeed at cross-gender word recognition. *Journal of Speech, Language, and Hearing Research*, *55*, 554–560. [https://doi.org/10.1044/1092-4388\(2011/10-0347\)](https://doi.org/10.1044/1092-4388(2011/10-0347))
- Vihman, M. M., Nakai, S., DePaolis, R. A., & Hallé, P. A. (2004). The role of accentual pattern in early lexical representation. *Journal of Memory and Language*, *50*, 336–353. <https://doi.org/10.1016/j.jml.2003.11.004>
- Walker, A., & Hay, J. (2011). Congruence between “word age” and “voice age” facilitates lexical access. *Laboratory Phonology*. Retrieved from <http://www.degruyter.com/view/j/lp.2011.2.issue-1/labphon.2011.007/labphon.2011.007.xml>
- Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development*, *1*, 197–234. https://doi.org/10.1207/s15473341lld0102_4
- 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. [https://doi.org/10.1016/S0163-6383\(84\)80022-3](https://doi.org/10.1016/S0163-6383(84)80022-3)
- Yoshida, K. A., Fennell, C. T., Swingle, D., & Werker, J. F. (2009). Fourteen-month-old infants learn similar-

sounding words. *Developmental Science*, 12, 412–418.
<https://doi.org/10.1111/j.1467-7687.2008.00789.x>

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's website:

Figure S1. Age by Children's Depression Inventory Productive Vocabulary

Figure S2. Subject Means by Children's Depression Inventory Productive Vocabulary

Figure S3. Proportion of Target Looking by Age Group