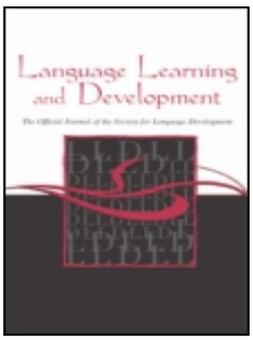

Language Learning and Development

ISSN: 1547-5441 (Print) 1547-3341 (Online) Journal homepage: http://www.tandfonline.com/loi/hlld20

Talker Variation Aids Young Infants' Phonotactic Learning

Amanda Seidl, Kristine H. Onishi & Alejandrina Cristia

To cite this article: Amanda Seidl, Kristine H. Onishi & Alejandrina Cristia (2014) Talker Variation Aids Young Infants' Phonotactic Learning, Language Learning and Development, 10:4, 297-307, DOI: 10.1080/15475441.2013.858575

To link to this article: http://dx.doi.org/10.1080/15475441.2013.858575

Language Learning and Development, 10: 297–307, 2014

ISSN: 1547-5441 print / 1547-3341 online
DOI: 10.1080/15475441.2013.858575

Routledge
Taylor & Francis Group

Talker Variation Aids Young Infants' Phonotactic Learning

Amanda Seidl
Speech, Language, and Hearing Sciences, Purdue University

Kristine H. Onishi
Department of Psychology, McGill University

Alejandrina Cristia
Laboratoire de Sciences Cognitives et Psycholinguistique, CNRS, IEC-ENS, EHESS

We investigated how talker variability impacts novel phonological pattern learning in 4- and 11-month-olds. Both age groups were better able to discriminate between legal and illegal phonotactic strings after exposure to multiple talkers than a single talker. It is argued that these data may be best accounted for by hybrid models that include linguistic representations that abstract away from talker identity, provided that neither a lexicon nor a protolexicon is necessary to allow the separation of linguistic representations and talker identity.

INTRODUCTION

A fundamental question in developmental psycholinguistic research concerns the format of speech representations and how different representations interact. Two extreme positions have been proposed, which can be caricatured as abstractionist-versus exemplar-based. Naturally, there are many variations of these views and many hybrid models combining an exemplar-based stratum upon which abstractions are built (e.g., Cole, Linebaugh, Munson, & McMurray, 2010; Drager, 2011; Johnson, Strand, & D'Imperio, 1999; McMurray & Jongman, 2011; Nygaard, Sommers, & Pisoni, 1994; Pierrehumbert, 2006). Nonetheless, describing the extremes of exemplar and abstractionist models facilitates discussion of the positions and questions at stake.

The abstractionist extreme holds that humans instinctively treat linguistic characteristics, such as phonology, separately from nonlinguistic characteristics, such as talker identity. For example, Bristow et al. (2008) document that different neural networks operate in 2-month-olds' detection of a change in talker identity as opposed to a change in vowel identity. Separate processing of linguistic and talker features has also been invoked in the context of conditioned headturn studies wherein infants are trained to turn their head in response to a vowel change. In some of this work, infants are reported to spontaneously generalize to novel talkers, ignoring both voice and pitch

Correspondence should be addressed to Amanda Seidl, Speech, Language, and Hearing Sciences, Purdue University, 500 Oval Drive, West Lafayette, IN 47906. E-mail: aseidl@purdue.edu

variation (Kuhl, 1983). It follows from this view that talker variation should have little effect on linguistic processing at any age.

In contrast, the exemplar-based extreme proposes that initially acoustic experience is represented faithfully, thus conflating linguistic and talker information. It is only through experience that learners come to find which dimensions structure spoken material, including the fact that while acoustic cues may overlap between linguistic and talker dimensions, these different dimensions serve to separate linguistic and talker information (e.g., Hintzman, 1986; Johnson, 1997; just as it is in other areas of categorization, e.g., Kovack-Lesh, Oakes, & McMurray, 2013). This view thus predicts changes in processing with age in terms of the representations involved. Additionally, unlike the extreme abstractionist view, there is a cost to speaker normalization precisely because all information is initially conflated in holistic exemplars. As a consequence, humans of different ages may vary not only in terms of the representations that they use but also in terms of how well they can manage the cost that normalization imposes on speech processing. Exemplar-based views are mainstream in infant literature today because they are supported by three types of evidence. First, it has been shown that infants initially perform more poorly on linguistic tasks when talkers change. For example, a close inspection of the behavioral data noted above, which is commonly cited in support of infants' spontaneous reliance on vowel category over talker changes (Kuhl, 1983), reveals that many more infants fail when talker changes are added to sound categorization tasks, and those who succeed need more training trials (see Experiment 2). Second, the deleterious effects of a change in nonlinguistic dimensions are reduced if infants are exposed to irrelevant variability along this dimension (Singh, 2008), suggesting that variation allows infants to extract the relevant, albeit initially conflated, dimensions. Finally, older infants have been reported to spontaneously generalize to untrained tokens varying along a nonlinguistic dimension in the same task where younger infants fail (Singh, Morgan, & White, 2004), evidence that is consistent with a change in representation as infants learn that linguistic and talker dimensions should be kept separate. Indeed, a mainstream exemplar-based model known as PRIMIR posits that truly abstract linguistic representations await the development of the lexicon, probably well into toddlerhood (Curtin, Byers-Heinlein, & Werker, 2011).

If early representations conflate all information, what are the mechanisms allowing their separation in younger infants' online processing, and their orderly storage in older infants? It is commonly believed that infants and adults alike use variability with respect to an independent label or tag to determine whether a given dimension is relevant or not. For example, Japanese adults' learning of the contrast between 'r' and 'l' was rendered more robust when the stimuli comprised multiple talkers and the target identity of the consonants were also provided to the listeners (Lively, Logan, & Pisoni, 1993). Similarly, talker variation has been found to promote toddlersrsquo; word learning (Rost & McMurray, 2010), possibly because in this study the visual referent remained constant in the face of speaker variability, thus providing a reliable tag as to which exemplars belonged together. This mechanism need not operate only in an explicitly supervised setting. On the contrary, evidence from infants' word segmentation suggests that even constant wordforms can serve this role. Indeed, 7.5-month-olds who heard the same wordform spoken with different affects (Singh, 2008) or voices (Houston & Jusczyk, 2000) more easily generalized to an untrained affect/talker than infants hearing stimuli from a single affect/talker. In other words, the constancy in phonological form across repetitions of a wordform that vary

in other characteristics suffices to prevent young infants' encoding of irrelevant indexical cues, possibly by providing a form of independent tag.

It has been suggested that the initially deleterious effects of talker variation are also reduced as infants accumulate language experience over the course of their typical development, allowing them to eventually separate linguistic and talker information online. Indeed, infants aged 11 months and older no longer require a familiarization with variable indexical/affect cues in order to segment a familiarized word spoken by a new talker or in a new affect (Houston & Jusczyk, 2000; Singh, 2008). These experienced infants have presumably come to know that, in the context of word segmentation, indexical/affect cues are irrelevant.

A bird's eye view reveals a number of open questions in this previous literature. First, whereas researchers working on lexical tasks have long declared the abstractionist extreme dead, there is other evidence that this view uniquely captures. Indeed, how else can we explain that two different brain networks responded to a vowel versus talker mismatch in 2-month-olds (Bristow et al., 2008)? It is likewise difficult to explain why 6-month-olds trained to detect a change in vowel represented in the speech of a single talker spontaneously interpreted this pattern as "a change in vowel category" and generalized it to vowel changes in untrained voices, rather than viewing it as "any acoustic change," in Kuhl (1983). Moreover, separate networks appear logically necessary given the ease with which 7.5-month-olds with brief familiarization with variable input separate linguistic and indexical information in wordform recognition studies. As a matter of fact, the mechanism by which infants are said to downplay the influence of indexical characteristics relates to the variability within this dimension being higher than that found for the other dimension, which presupposes two separate dimensions within which variability can be computed in parallel.

Since separate coding of vowel and talker identity in 2- to 6-month-olds requires separate processing streams for linguistic and indexical cues, the second pressing question is: How may we integrate these results with those on lexical processing on 7.5- to 14-month-old infants? Given the divergence in methods and ages, we could entertain several hypotheses. For example, one could argue that *lexical* entries conflate indexical and linguistic information, but *phonetic* representations do not.[1] Or perhaps the development of representations is u-shaped, with separate indexical and linguistic encoding both early and late in development. The present study gets at the heart of these possibilities, by assessing both younger and older infants in a domain that is in between phonetic and lexical, that of *phonotactics*.

Phonotactics are regularities in the order and position of sounds and sound classes, constituting an important factor in linguistic development (Jusczyk, 1997). They resemble phonetic knowledge in that they are dependent on accurate definition of sound categories. At the same time, they concern sequences of sounds and are similar to words in that they could require more global processing than individual sounds. Phonotactics can also bring unique evidence to the question of potential developmental trajectories because infants can quickly learn phonotactics in the lab throughout the first year (for a review, see Cristia & Peperkamp, 2012).

[1]Notice that work documenting improvements with talker variation in phonetic learning has conflated talker and token variation (e.g., Lively et al., 1993). To our knowledge, there is no evidence showing a unique contribution of talker variation per se to phonetic learning. The same criticism could be extended to some studies on lexical processing (e.g., Richtsmeier, Gerken, & Ohala, 2011).

In this experiment we used an artificial grammar learning paradigm to assess 4- and 11-month-old infants' ability to extract novel phonotactics in two different talker variation conditions. Half of the infants was familiarized and tested with the same single talker, whereas the other half was familiarized with three different talkers and tested on a novel talker (the same talker used in the single talker condition). We chose a pattern that was structurally challenging in order to avoid ceiling effects: The manner of an initial consonant determined the tenseness of the following vowel. This is not a first-order dependency (sound restricted to syllabic position), but rather a second-order dependency (sound restricted to segmental environment), which is harder for infants to learn (a discussion in Chambers, Onishi, & Fisher, 2011). Among second-order dependencies, onset-vowel (CV) dependencies are likely particularly challenging (Kessler & Treiman, 1997), more so than vowel-coda co-occurrences which are learned by infants at a variety of ages (Seidl, Cristià, Bernard, & Onishi, 2009).[2]

As noted above, a review of previous psycholinguistic work on infant processing of linguistic and indexical properties suggests many different hypotheses. To simplify, we first take the extreme models presented at the outset, and draw predictions from each of them before turning to a third, more nuanced view.

The abstractionist extreme holds that there are two separate networks channeling linguistic and indexical information separately from a very young age and onto adulthood (see, e.g., Bristow et al., 2008, p. 917). Since phonotactics concerns phonemic processing and talker information is irrelevant, then no difference should be found across the two variability conditions, and no interaction with age should be found.

Extreme exemplar models assume that, prior to extensive lexical experience, indexical and linguistic information are conflated in each exemplar, which is faithfully stored in memory. In our single talker condition, no extrapolation to a novel talker was necessary. Therefore, it follows that both 4- and 11-month-olds should be able to recognize the sound pattern extracted from the familiarization talker in the novel wordforms presented at test in the voice at that same talker (provided that the rule is not overly challenging). In contrast, infants in our multitalker condition had to extrapolate to a novel talker at test. While this might prove too challenging for the 4-month-olds tested here, it is clear that 11-month-olds should be just as good in the single as in the multiple talker conditions, because through their (proto)lexical experience they have learned to ignore indexical cues in phonological processing (Houston & Jusczyk, 2000).

Other work has documented beneficial effects of variation in toddlers (Rost & McMurray, 2010), children (Gómez, 2002), and adults (Bradlow, Nygaard, & Pisoni, 1999; Creel, Aslin, & Tanenhaus, 2008; Holt & Lotto, 2006; Iverson, Hazan, & Bannister, 2005; Jamieson & Morosan, 1989; Lively et al., 1993; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002). The mechanism behind improved performance holds that, in all these cases, 'tags' that remain constant inform learners about which exemplars belong together, and which should be stored separately. Such tags are evident in cases of word learning (where a constant visual referent is present) and in phonetic classification with feedback (where the correct orthographic forms are

[2] As in previous phonotactic learning studies, the pattern used was not phonetically grounded (Onishi, Chambers, & Fisher, 2002; Seidl & Buckley, 2005), which allows better control over infants' exposure to the pattern prior to the study. Phonetically grounded patterns typically occur as statistical trends even in languages where a specific pattern is not phonologized. Nonetheless, patterns similar to the ones in this paper are not linguistically impossible; many languages show alternations with tenseness (Kiparsky, 1995; Vaux, 1996) as well as consonant-vowel interactions (Vaux, 1996).

provided); they are arguably also present in the form of a more abstract wordform shape in segmentation studies, perhaps in terms of the sounds that are virtually identical across talkers. In the present case, however, the tag was akin to the pattern itself; in other words, only discovering the pattern would allow infants to tell which sounds co-occurred. Notwithstanding this conceptual leap, an inductive generalization from this previous body of work would predict that infants in the present study should be better in the multiple than in the single talker conditions.

EXPERIMENT

During the familiarization phase, 4- and 11-month-old infants heard 24 CV syllables instantiating specific phonotactics and produced either by three talkers (Multi condition) or by a single talker (Single condition). During the test phase, all infants were tested with new syllables, spoken by the talker used in the Single condition (who had not been presented during the Multi familiarization). Half of the test syllables was legal (followed the phonotactics present in the familiarization), and half was illegal (violated the phonotactics from familiarization). A statistically significant preference for one of these types during test provides evidence for learning.

Method

Participants. Thirty-six monolingual English-learning 11-month-old (23 male; mean = 10.92; range = 10.43-11.5) and 36 4-month-old (26 male; mean = 4.48; range = 3.98-5.03) infants participated in the study. Sixteen additional infants were tested but not included in the final sample because they cried during the experiment (12), had orientation times to legal or illegal syllables that were more than 2 standard deviations from the mean orientation times (3), or because of equipment error (1). Parental consent was obtained for all participants. Participants were given a book or toy for their participation in the study.

Design and stimuli. During familiarization, syllables were CV pseudowords exhibiting a dependency between the C and V. The 'stop-tense' rule contained syllables with stop onsets followed by tense vowels and fricative onsets followed by lax vowels (e.g., /bi/ /zɛ/). The 'stop-lax' rule showed with the reverse pattern: stops followed by lax and fricatives followed by tense vowels (e.g., /bɛ/ /zi/). The initial C was /b,k,p,t,g,d/ (stops) or /s,v,ʃ,ʒ,f,z/ (fricatives). Vowels were one of /I,ɛ,ʊ/ (lax) or /i,e,u/ (tense). By combining the Cs and Vs, there were 36 non-overlapping pseudowords for the stop-tense rule (18 stop-tense, 18 fricative-lax) and 36 for the stop-lax rule. For each rule the 36 pseudowords were divided into three sublists. For each infant, two sublists were presented during familiarization, and the remaining sublist served as legal test items. Illegal test items were taken from the third sublist of the counterbalanced group who were familiarized with the opposite rule. In other words, half of the infants at each age was assigned to the stop-tense rule and half was assigned to the stop-lax rule, and within each rule and age one third of the infants were assigned to each familiarization list.

Syllables were recorded by four female native English speakers in non-words of the shape CVd.da for lax vowel and CV.da for tense vowels, so that the phonotactics of English were respected. The target CV was then excised in Praat at an upward zero crossing. Twelve naïve

English adults listened to and labelled the CV stimuli. A perception study with adults revealed that, on average, the syllables were correctly labelled 87% of the time (range for individual stimuli 33–100%). Half the infants in each age group were randomly assigned to the Multi condition and the other half randomly assigned to the Single condition. Infants in the Single condition heard all pseudowords spoken by a single talker, and were tested with pseudowords from that same talker. Infants in the Multi condition heard familiarization pseudowords spoken by three different talkers, and were tested on a fourth talker (the one used in the Single condition). The tokens used in the single talker condition were rated correctly by adults as often as those used in the multi talker condition [$t(71) = 1.25$, $p > .2$].

During familiarization, five repetitions of the 24 familiarization pseudowords were presented in a randomized order, with 600 ms pauses in between pesudowords and lasted approximately two minutes. At test, there were eight test trials; four consisted of different orderings of the 12 stop-tense pseudowords, and four of different orderings of the 12 stop-lax trials. Thus, each test trial was legal for half the infants and illegal for half. The order of presentation of the eight test trials was randomized by the computer program used to run the study.

Procedure and apparatus. Infants were tested using the Headturn Preference Procedure (Jusczyk & Aslin, 1995). The infant was seated on a caregiver's lap in a small room with lights on the front and side walls and an audiospeaker behind each side light. Caregivers and experimenters wore Peltor aviation headphones and listened to loud masking music during the experiment. During the experiment each trial began with the front light flashing to attract the infant's attention. After the infant oriented toward it, the light was extinguished and one of the two side lights began flashing. Orientation time was recorded when the infant maintained orientation within 30 degrees of the flashing light after an initial 90-degree headturn toward it. Total orientation time did not include time orienting away, although during orientations away shorter than two consecutive seconds, the sounds and flashing continued. Familiarization sounds were presented continuously and simultaneously from both side speakers. They were initiated by the first orientation toward the first flashing side light and terminated after all familiarization lists had been presented. Thus, during familiarization, lights but not sounds were contingent on the infantrsquo;s head orientation (e.g., Seidl & Buckley, 2005). Test sounds were presented from the single side speaker behind whichever side light was flashing on that trial. Each of the eight test trials was initiated as in familiarization and terminated when (a) all the 12 pesudowords had been presented three times, or (b) the infant oriented away from the light for more than two consecutive seconds. During test, both lights and sounds were contingent on the infant's head orientation.

RESULTS

We ran a repeated-measures analysis of variance (ANOVA) on average orientation times with Trial type (Legal, Illegal) as a repeated measure, and Age (4, 11), Condition (Single, Multi), Rule (stop-tense, stop-lax), and List (1, 2, 3) as between-subjects factors. The ANOVA revealed main effects of Trial type [$F(1,53) = 7.84$, $p = .007$], due to overall larger orientation times to Illegal trials; and Age [$F(1,53) = 14.28$, $p < .0004$], due to overall longer orientation times by the 4-month-olds. The factor of interest, Condition, significantly interacted with Trial type [$F(1,53) =$

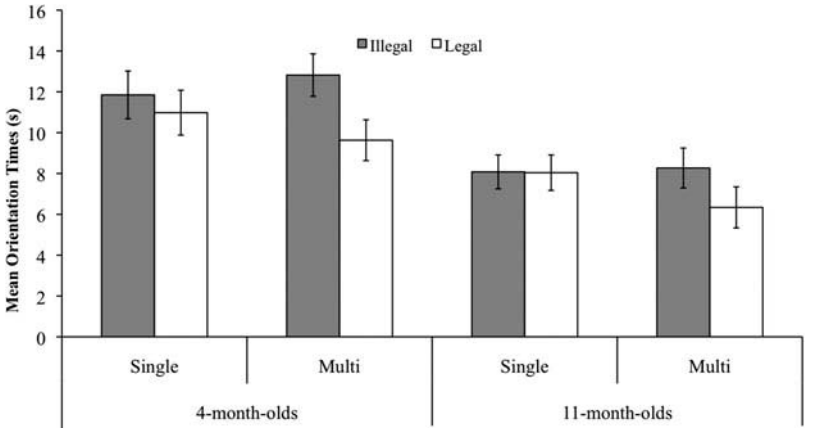

FIGURE 1 Mean orientation times and standard errors in orientation times to legal and illegal test pseudowords by Condition (Single, Multi) and Age (4-, 11-month-olds).

3.84, $p = .05$] due to a facilitation effect for multiple talkers (Figure 1). No other main effects or interactions were significant [Fs < .1.6, ps > .21].

To explore the interaction between Condition and Trial type, we ran the same ANOVA within each Condition. In the Single condition, there was a main effect of Age [$F(1,27) = 4.98$, $p = .034$]; a marginal effect of Rule [$F(1,27) = 3.99$, $p = .06$]; and no other main effects or interactions (Fs < 1.5, ps > .22). In the Multi condition, there was a main effect of Age [$F(1,27) = 13.47$, $p = .001$]; and Trial type [$F(1,27) = 10.54$, $p = .003$], with longer orientation times to illegal items; and no other main effects or interactions [Fs < 1.07, ps > .31]. Thus, the interaction Condition x Trial type was due to a facilitation when exposed to multiple talkers, as shown in Figure 1. Specifically, in the Multi condition, 27 out of 36 infants looked longer to Illegal than Legal trials ($p < .002$, sign test), whereas in the Single condition only 20 out of 36 showed this pattern ($p = .302$).

DISCUSSION

This experiment is the first to assess whether talker variation affects 4- and 11-month-old infants' learning of novel phonotactic patterns. Our results align with previous work reporting that older infants show better learning with exposure to talker variation (Rost & McMurray, 2010), but go further to suggest that this is also the case for much younger infants. The lack of a difference in the benefits reaped from variable talkers between two diverse ages may be surprising in the context of at least one mainstream exemplar-based model of infant speech perception proposing that only lexical entries allow truly abstract representations (Curtin et al., 2011). Our results also sit uncomfortably with extreme abstractionist views, where non-overlapping brain networks encode talker and phonemic identity (Bristow et al., 2008), since an improvement with talker variation is quite difficult to account for if these two networks run completely in parallel. The integration

with our results with such comprehensive and widespread views of early language acquisition brings to the fore questions that must be explored in future research.

In general, our results appear to fit in with exemplar-based models at large as follows. Since the speech is highly variable in the multitalker condition, the information that is consistent across forms (the phonotactic regularities) comes to be more heavily weighted. Phonotactic regularities are thus highlighted and become easier to learn. In contrast, the single-talker condition contains many more phonetic similarities due to the fact that forms in this condition are all produced by the same speaker.[3] Thus, the weighting of the similarities in phonotactics may be relatively less salient, rendering the phonotactic pattern more challenging or less noticeable for the infant. This line of reasoning could easily accommodate the poor generalization that has been found in word segmentation tasks, where young infants only succeed when the voice and affect are matched across familiarization and test. In typical word segmentation tasks, infants hear a handful of tokens of two word types, and speech sound discrimination studies show even less variability, often relying on a single acoustic token. Repetitive familiarization leads learners to encode minute details, instead of allowing them to select a more abstract level of encoding by making such details uninformative (Holt & Lotto, 2006). While such "abstraction" is not included in pure exemplar-based models (e.g., Goldinger, 1996), many more recent hybridized models include such possibilities (e.g., Pierrehumbert, 2006). Indeed, recent data from infants also supports this idea. For example, van Heugten and Johnson (2012) document that even 7.5-month-olds succeed in cross-gender word recognition if the familiarization; consists of passages, which likely make the target word more acoustically variable. (Note that this ability to segment across genders is not seen in similar work when words in isolation are used during familiarization, Houston & Jusczyk, 2000.) Similarly, in our phonotactic pattern, infants heard 24 different words, each repeated a few times, which spanned nearly every manner-tenseness combination of three types of obstruents and three types of vowels.

While thus far the story is simple, a key feature of some mainstream exemplarist models of infant perception is that lexical development drives abstraction. Indeed, lexical development is thought to explain the fact that 9-month-olds fail to segment words across a familiar and an unfamiliar accent (Schmale, Cristià, Seidl, & Johnson, 2010) and across different affects (Singh et al., 2004), whereas 12-month-olds succeed; and lexical development is again invoked as toddlers come to better learn (Schmale, Hollich, & Seidl, 2011) and recognize (Mulak, Best, Tyler, Kitamura, & Irwin, 2013) words across varied accents. Such statements led us to predict differences between 4- and 11-month-olds because 4-month-olds most certainly do not have the same amount of experience with wordforms or real word-meaning associations that 11-month-olds have (there is little work on the lexicon of 4-month-olds, but we can imagine how limited it is by comparing the hesitant performance of slightly older 6-month-olds with that of toddlers in Bergelson and Swingley, 2012).

The mainstream infant exemplarist model, PRIMIR (Curtin et al., 2011), proposes that any effect could be modulated by task demands. Therefore, one could argue that perhaps the facilitation observed at 4 months is not due (as in 11-month-olds) by lexical development, but rather to

[3]One anonymous reviewer points out that this speech is perhaps more variable than studies involving word segmentation, since the same words are not repeated. Results indicate that this type of "type" variability may not be sufficient to elicit the learnability boosts seen with multiple talkers.

this same task being easier with multiple talkers at this age. However, this argument is unconvincing, as it is still expected that task demands will be greater for 4-month-olds in the multi- than single-talker condition. Indeed, in addition to the difficulty of finding the relevant dimensions (lacking the separable dimensions afforded by a larger lexicon), these younger infants have more limited memory and selective attention resources, and must thus find it challenging to focus on the relevant dimension. Thus, our results suggest that not only is a lexicon unnecessary to form separable talker and linguistic dimensions, but also that this task does not greatly tax memory or selective attention. This is also suggested in Houston and Jusczyk (2003) in which even 7.5-month-old infants were able to recall a briefly presented voice after a three-day delay suggesting an impressive ability to recall indexical information.[4] Future work could explore the role of memory and selective attention more directly, focusing particularly on 4-month-olds, for whom lexical development is already discounted as a source of information for separating talker and linguistic information.

As mentioned in the Introduction, abstractionist models assume that even 2-month-olds are able to treat talker and linguistic dimensions as separable. Thus, the lack of age effects in our data pose no particular problems for these models. These models, however, predict a lack of an effect of talker variation which was not supported by the present data, since we did find an effect of talker variation. By and large, abstractionist models are less popular in infant research because they cannot explain interactions, which are frequently reported; that is, they fail to predict that talker changes between familiarization and test hinder discrimination and word segmentation performance, and that talker variability during familiarization facilitates lexical learning – and, our data suggests, phonotactic learning. The only way in which abstractionist models could account for multi-talker boosts of performance is if they assume that a variable familiarization is more interesting, and increases in performance are the result of increases in arousal. Arousal is a relevant factor for both the NLM-e framework (Kuhl et al., 2008), which attempts to integrate social and cognitive factors to language acquisition, and PRIMIR (Curtin et al., 2011), which underlines the effects of task characteristics on infant performance. Might have varying the talker during familiarization rendered our task more socially interesting, hence lending to increased arousal? In fact, there is little support in our data for such an interpretation, as an increase in arousal should have caused higher overall levels of orientation time at either familiarization or test in the multi than the single condition which was not found in this work[5] or in similar work (see Rost & McMurray, 2010 for discussion). Nonetheless, future work may investigate this hypothesis more directly by incorporating measures of arousal, such as heart-rate and measures derived from electro-encephalography (Richards, 2001).

In sum, we document a boost in phonotactic learning via exposure to multiple talkers, and this in both 4- and 11-month-olds. These data sit uncomfortably with both extreme exemplarist and abstractionist views of language acquisition, and pose an interesting challenge for hybrid models of infant speech perception.

[4]One anonymous reviewer points out that the recent Apfelbaum and McMurray (2011) model would also have difficulty accounting for the lack of an age effect here, while Toscano & McMurray's (2010) model, in which variability along one dimension can impact the weighting of other less variable dimensions, may be broad enough to explain these results.

[5]An ANOVA with orientation times in familiarization as within and Age group and Talker as between subjects variables yielded no main effects or interactions ($Fs < .703$, $ps > .402$).

ACKNOWLEDGMENTS

We thank all the parents and infants who participated in these studies as well as Yuanyuan Wang and Carrie Wade for help in running participants. We would also like to thank Matt Goldrick, Marieke van Heugten, and two anonymous reviewers for comments on earlier drafts.

FUNDING

This work was supported by funds from NSF 0843959 to the authors.

REFERENCES

Apfelbaum, K. S., & McMurray, B. (2011). Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*, *35*, 1105–1138.

Bergelson, E., & Swingley, D. (2012). At 6 to 9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences*, *109*, 3253–3258.

Bradlow, A., Nygaard, L., & Pisoni, D. (1999). Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics*, *6*, 206–219.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., et al. (2008). Hearing faces: How the infant brain matches the face it sees with the speech it hears. *Journal of Cognitive Neuroscience*, *21*, 905–921.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2011). Representations for phonotactic learning in infancy. *Language Learning and Development*, *7*, 287–308.

Cole, J. S., Linebaugh, G., Munson, C., & McMurray, B. (2010). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics*, *38*, 167–184.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*, *106*, 633–664.

Cristia, A., & Peperkamp, S. (2012). Generalizing without encoding specifics: Infants infer phonotactic patterns on sound classes. In *Proceedings of the 36th Annual Boston University Conference on Language Development* (pp. 126–138). Somerville, MA: Cascadilla.

Curtin, S., Byers-Heinlein, K., & Werker, J. F. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, *39*, 492–504.

Drager, K. (2011). Speaker age and vowel perception. *Language and Speech*, *54*, 99–121.

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

Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science*, *13*, 431–436.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *The Journal of the Acoustical Society of America*, *119*, 3059–3070.

Houston, D., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, *26* (5), 1570–1582.

Houston, D., & 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* (6), 1143–1154.

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

Jamieson, D. G., & Morosan, D. E. (1989). Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology*, *43*, 88–96.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. K. Johnson & J. W. Mullenix (Eds.), *Talker variability in speech processing* (pp. 145–166). San Diego, CA: Academic Press.

Johnson, K., Strand, E. A., & D'Imperio, M. (1999). Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*, *24*, 359–384.

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Jusczyk, P. W., & Aslin, R. (1995). Infants' detection of sound patterns of words in fluent speech. *Cognitive Psychology*, *29*, 1–23.

Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. *Journal of Memory and Language*, *37*, 295–311.

Kiparsky, P. (1995). The phonological basis of sound change. J. Goldsmith (Ed.), *Handbook of phonological theory* (pp. 640–670). Oxford, England: Blackwell.

Kovack-Lesh, K. A., Oakes, L. M., & McMurray, B. (2013). Contributions of attentional style and previous experience to 4-month-old infants' categorization. *Infancy*, *17*, 324–338.

Kuhl, P. K. (1983). Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development*, *6*, 263–285.

Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*, *363*, 979–1000.

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. *Journal of the Acoustical Society of America*, *94*, 1242–1255.

McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the r/l contrast to Japanese adults: Predictions of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, and Behavioral Neuroscience*, *2*, 89–108.

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*, *118*, 219–246.

Mulak, K. E., Best, C. T., Tyler, M. D., Kitamura, C., & Irwin, J. R. (2013). Development of phonological constancy: 19-month-olds, but not 15-month-olds, identify words in a non-native regional accent. *Child Development, 84,* 2064–2078.

Nygaard, L., Sommers, M., & Pisoni, D. B. (1994). Speech perception as a talker contingent process. *Psychological Science*, *5*, 42–46.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition*, *83*, B13–B23.

Pierrehumbert, J. (2006). The next toolkit. *Journal of Phonetics*, *34*, 516–530.

Richards, J. E. (2001). Attention in young infants: A developmental psychophysiological perspective. C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 321–338). New York, NY: MIT Press.

Richtsmeier, P. T., Gerken, L., & Ohala, D. K. (2011). Contributions of phonetic token variability and word-type frequency to phonological representations. *Journal of Child Language*, *38*, 951–978.

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.

Schmale, R., Cristià, A., Seidl, A., & Johnson, E. K. (2010). Infants' word segmentation across dialects. *Infancy*, *15*, 650–662.

Schmale, R., Hollich, G., & Seidl, A. (2011). Contending with Foreign Accent in Early Words Learning. *Journal of Child Language, 38,* 1096–1108.

Seidl, A., & Buckley, E. (2005). On the learning of arbitrary phonological rules. *Language Learning and Development*, *1*, 289–316.

Seidl, A., Cristià, A., Bernard, A., & Onishi, K. H. (2009). Allophones and phonemes in infantsrsquo; learning of sound patterns. *Language Learning and Development*, *5*, 191–202.

Singh, L. (2008). Influences of high and low variability on infant word recognition. *Cognition*, *106*, 833–870.

Singh, L., Morgan, J., & 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.

Toscano, J., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, *34*, 434–464.

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.

Vaux, B. (1996). The status of ATR in feature geometry. *Linguistic Inquiry*, *26*, 175–182.