Published in final edited form as:

Lang Learn Dev. 2015 ; 11(1): 66–79. doi:10.1080/15475441.2014.895249.

The Role of Single Talker Acoustic Variation in Early Word Learning

Marcus E. Galle,

Department of Psychology and Delta Center, University of Iowa

Keith S. Apfelbaum, and

Department of Psychology and Delta Center, University of Iowa

Bob McMurray

Department of Psychology, Department of Communication Sciences and Disorders, and Delta Center, University of Iowa

Abstract

Recent work has demonstrated that the addition of multiple talkers during habituation improves 14-month-olds' performance in the switch task (Rost & McMurray, 2009). While the authors suggest that this boost in performance is due to the increase in acoustic variability (Rost & McMurray, 2010), it is also possible that there is something crucial about the presence of multiple talkers that is driving this performance. To determine whether or not acoustic variability in and of itself is beneficial in early word learning tasks like the switch task, we tested 14-month-old infants in a version of the switch task using acoustically variable auditory stimuli produced by a single speaker. Results show that 14-month-olds are able to learn phonemically similar words within the switch task with increased acoustic variability and without the presence of multiple talkers.

There is an apparent paradox in infant speech perception. Work on discrimination of speech sounds has shown that infants undergo rapid development of phonological abilities, displaying adult-like discrimination by 12 months of age (Werker & Tees, 1984). However, 14-month-old infants have difficulty utilizing these skills in word learning tasks (Stager & Werker, 1997; Pater, Stager, & Werker, 2004; Werker & Fennell, 2004). Thus, whether phonological development appears complete by this age depends largely on the domain from which it is viewed.

Research on word learning during this critical age has frequently used the *switch task* to assess infants' abilities to map similar phonological patterns onto objects. In this task, infants are habituated to presentations of two word/referent pairings, and then tested on correct and incorrect pairings. As identical stimuli are used for habituation and test trials, the critical factor is whether associations between words and referents established during habituation are maintained during test. Successful learning of word/referent pairings during habituation should lead to longer looking to novel, incorrectly paired stimuli. Without

Correspondence should be addressed to Bob McMurray, Department of Psychology, University of Iowa, E11 SSH, Iowa City, IA 52242. Bob-mcmurray@uiowa.edu.

Keith S. Apfelbaum is now at the Department of Psychology, The Ohio State University.

learning, looking time should not differ as the individual stimuli in both trial types are equally familiar.

In this task, 14-month-olds can learn phonologically dissimilar words such as *lif* and *neem*. However, they fail to learn similar-sounding words like *bih* and *dih* despite their ability to discriminate this phonemic contrast (Stager & Werker, 1997). Explanations for these findings focus on the difficulty of word learning, which inhibits access to perceptual abilities (Werker & Yeung, 2005). In support of this hypothesis, research has shown that 14-month-olds succeed in the switch task when tested with familiar word (e.g., *ball* and *doll*), which presumably reduces processing demands (Fennell & Werker, 2003). The difficulty of the testing phase may also mask learning; Yoshida, Fennell, Swingley, and Werker (2009) found that 14-month-olds can learn phonologically dissimilar words when tested in the preferential looking paradigm after a learning phase similar to the standard switch task (see also Swingley & Aslin, 2002, 2007; Ballem & Plunkett, 2005). Whereas the switch task requires infants to determine whether a novel pairing is a correct match, the preferential looking task requires a (presumably) simpler decision of which object is a better match.

An alternative hypothesis is that discrimination skills at 12 months are not indicative of robust phonological categories (Galle & McMurray, in press). Indeed, several studies suggest that speech categorization continues to develop after the first year: the weighting of cues for /s/ and /ʃ/ changes throughout preschool (Nitttrouer, 2001), and /r/ and /w/ categories develop until age 5 (Slawinsky & Fitzgerald, 1998).

In support of this hypothesis, recent work has demonstrated that 14-month-olds can learn phonologically similar words in the switch task when labels are spoken by different talkers during training (Rost & McMurray, 2009, 2010). While the standard switch task uses a small number of auditory exemplars produced by a single talker, Rost and McMurray used auditory stimuli from 18 different talkers (three exemplars of both /buk/ and /puk/ from each talker). On each training trial, infants heard seven different exemplars of either /buk/ or /puk/, randomly chosen from this set. Each talker was heard an equal number of times across trials and for each word, so that infants could not associate a particular talker with either object. Seven new exemplars from the same talkers were used for testing.

If 14-month-olds possess robust phonological categories multitalker training should not provide a large benefit in the switch task, and is more likely to reduce performance. In fact, multitalker variation can adversely affect speech *processing* in adults (Mullennix, Pisoni, & Martin, 1992) and infants (Jusczyk, Pisoni, & Mullennix, 1992), so the presence of multiple talkers in the switch task should increase task difficulty.¹ However, variation across multiple talkers may offer richer data for perceptual learning that can bootstrap infants' incomplete categories. Infants' success under these circumstances suggests failure in word learning tasks is not merely the result of methodological factors, and may reveal incomplete phonological development.

¹An intriguing alternative is that talker variability helps when the stimuli are phonetically similar, as in Rost and McMurray (2009) but hinders when it is more variable. This currently remains untested as work with adults such as Mullennix and Pisoni (1990) had not examined phonetically similar items.

In a subsequent study, Rost and McMurray (2010) hypothesized that increased acoustic variability in multitalker training highlights *relevant* acoustic differences for distinguishing similar word forms (or inhibits irrelevant acoustic differences, like talker characteristics). Their original (2009) stimulus set contained variation in both phonologically relevant (e.g., voice onset time [VOT] that distinguishes b/p) and irrelevant cues (e.g., talker). To separate these effects, they trained infants in one of two conditions. In one condition, words varied only in VOT along a bimodal distribution (variation around the prototypical voiced and voiceless values), but the rest of the words were acoustically identical (e.g., no variation in talker, prosody). In the other condition, VOTs were held constant (at the voiced and voiceless prototypes) but the rest of the word varied on irrelevant factors (largely talker, but also prosody). Only infants trained on variation in irrelevant cues succeeded in learning the words. Yet if 14-month-olds indeed have robust phonological categories, talker variability should have inhibited word learning.

Rost and McMurray argue that the variability of acoustic cues indicates whether they are relevant for word learning, with less variable cues being more relevant for word learning than variable cues. Apfelbaum and McMurray (2011) offer a simple associative model to illustrate this hypothesis (Figure 1). Their model suggests that in single talker training aspects of the talker's voice become associated with the objects. At test, when a mismatching word is presented in the same voice, the association between the object and the voice make rejecting the pairing difficult (Figure 1A). In contrast, when the talker varies, no single voice is strongly associated with the object, so associations with relevant cues dominate (Figure 1B). Critically, as a general associative principle, this model suggests that it is not *talker* variability per se that helps; rather high variability in any cue prevents association of those cues with the objects, and allows associations with other, more relevant, cues to control behavior.

This is one instantiation of accounts stressing the role of incomplete phonological development. However, there are alternatives accounts which make no claims about phonological development but posit specifically *talker* variability as the crucial factor. For example, two studies (Fais et al., 2012; Fennell & Waxman, 2010) suggest that the onset of referential awareness or social pragmatics (infants' realization that this is a word learning situation) is critical to success in the switch task. In support of this, Fennell and Waxman show that when 14-month-olds are briefly presented with known words and objects prior to standard training in the switch task, they can succeed in learning and discriminating minimal pairs. They argue that this brief exposure alerts children to the referential nature of the task, enabling them to succeed. While performance in this task (and early word learning more broadly, see Hollich et al., 2000) is clearly a product of both top-down and bottom-up factors, they also suggest that prior results on talker variability may stem largely from such top down factors. That is, the presence of multiple talkers may create social expectations that the context calls for word learning. It may also offer converging evidence to support the word/object mapping. When a name that has been used with an object by multiple people, it is much more likely to be the correct name, than one used by only a single talker. Thus, single-talker variation should not be sufficient to engage this hypothesis.

An alternative that also requires specifically talker variability may be that infants may need to develop low-level talker compensation processes in order to succeed in the single-talker switch task. That is, 14-month-olds may be adept at phonetic categorization but poor at compensating for novel talkers (the one heard in the lab), and work on segmentation and memory supports this (Hollich, 2006; Houston & Jusczyk, 2003; but see Van Heugten & Johnson, 2012). This may be because, all though even younger infants can discriminate different talkers (Johnson, Westrek, Nazzi, & Cutler, 2011) they have not yet learned to identify talkers (Creel & Jiménez, 2012). Critically under many accounts, it is talker identification that is essential for normalization (McMurray & Jongman, 2011; Nygaard, Sommers, & Pisoni, 1994). Relatedly, infants may not have sufficient experience with multiple talkers to know how different acoustic cues vary across talkers. Talker variation during training may help infants develop these skills.

In contrast, our cue-weighting account predicts that sufficient variability during training in any phonemically irrelevant cue should help infants learn in this paradigm, whereas social and talker-compensation accounts predict that single-talker training will not lead to enhanced performance even when acoustic variability is as high as multitalker training. We examined this by manipulating acoustic variability during training with a single talker. If increased variability within a single talker during training leads to success, then specifically talker variation may be unnecessary.

This hypothesis is bolstered by evidence that 7.5-month-old infants have difficulty segmenting words from running speech that differ in pitch from familiarized words (Singh, White, & Morgan, 2004) but succeed when pitch is variable during training (Singh, 2008). Intriguingly, 9-month-old infants succeed under both conditions (Singh, 2008). Given that acoustic variability during training helps in Singh's task (at a much younger ages), it seems likely it would also help in the switch task. However, the relationship between Singh's work and the switch task is less obvious than one might think. Minimally, Singh's work suggests that by 14 months variability is not necessary (as 9-month-olds do not require it), while Rost and McMurray's studies suggest otherwise. Thus, there may be something different about these two situations. We see at least three possibilities that raise doubts about the extensibility of Singh's results to this situation.

First, it could be a matter of difficulty: the word learning tasks engaged here use minimal pairs (*buk/puk*) for novel words, while Singh et al.'s (2004) study asked infants to discriminate highly differentiable words like *bike*, *hat*, *tree* and *pear*. If variability is involved in fine-tuning perceptual categories, it seems likely that sensitivity to fine-grained differences could develop later, and hence need variability at later ages. Thus, if single-talker variability leads to success in the switch task, this could reinforce the role of perceptual development.

Second, and perhaps more importantly, the fundamental task or goals engaged by these experiments is different. Speech segmentation requires participants to overcome prosodic variability to correctly recognize surface forms in running speech; a sense of familiarity with the word should suffice to demonstrate learning. For segmentation variability may help the more invariant, repeated, word-form "pop-out" from the context. In contrast, word learning

requires listeners to learn those surface forms and also form associations with visual referents. Here, variability plays a different role, as the goal is for one set of cues to emerge as stronger than others for the specific purpose of mapping to a referent. Thus, variability could help infants develop the sense of familiarity needed for segmentation, but play a different role in learning word/referent associations. This component of learned associations with visual referents is absent from segmentation tasks. Thus, infants may need more variability than a single talker can provide to develop the auditory/phonological representations needed to map words to referents, and indeed the standard switch task (Stager & Werker, 1997) employs some variability yet infants fail nonetheless. These factors make word learning a different, more complex task than segmentation, and it is not clear that the results of the Singh, White, and Morgan (2008) studies will necessarily generalize. Although 14-month-olds no longer require acoustic variability to succeed in segmentation tasks with more differentiable words, they may need it to succeed in the switch task (for perhaps different reasons).

Finally, perhaps the most important difference is the role that single-talker variability may play in disentangling theoretical accounts. As we have described, work on minimal pair word learning in young infants has raised important debates about whether the critical development that enables such learning is conceptual/social (Fais et al., 2012; Fennell & Waxman, 2010) or perceptual. Indeed, given the widespread assumption that perceptual development is in place by the onset of word learning, this becomes particularly important. People have also debated whether associative learning principles (as in the Apfelbaum & McMurray, 2011 account) play a role in such learning or whether it should be seen as primarily a conceptual problem (Waxman & Gelman, 2009; see Namy, 2012, for a review). Single-talker variability may play a role in documenting whether such high level accounts can explain the results of Rost and McMurray (2009, 2010) or whether there is a unique bottom-up component to this development, which would push back the window of perceptual development and support these associative models. In contrast, segmentation is widely and uncontroversially seen as a perceptual learning problem in which perceptual factors like acoustic variability are likely play a clear role. Thus, given the differences in the nature of the segmentation and word learning problems as well as the theoretical debates surrounding them evidence from segmentation may not be sufficient to impact theoretical debates about word learning. Single-talker variability may thus play a useful role in understanding the contribution of bottom-up perceptual development to word learning at a somewhat later time period than is often assumed.

The critical question therefore is not whether 14-month-olds benefit from acoustic variability in word learning tasks, but why this kind of variability is beneficial. Whereas Rost and McMurray (2010) argue for the same bottom-up processes hypothesized by Singh and colleagues for segmentation, others have suggested top-down mechanisms (Fennell & Waxman, 2010). An investigation of the processes at work during these tasks is important for assessing the continuity of development across age groups. Differences in likely mechanisms, task difficulty, and underlying theory necessitate investigating whether Singh and colleagues' findings hold in word learning contexts.

The present study examined the effect of increased acoustic variability within a single talker. While prior instantiations of the single-talker switch task have included some variation between trials (Pater et al., 2004; Stager & Werker, 1997; Werker, Fennell, Corcoran, & Stager, 2002), typically, only 7–10 exemplars were used throughout an experiment. The present study increased acoustic variability by using a single talker who attempted to maximize variability over several hundred tokens. This also contrasts with the single talker experiments in Rost and McMurray (2010) (in which VOT was manipulated) in that Rost and McMurray included no variation in prosodic or other cues within their single talker. If acoustic variability alone is sufficient, 14-month-olds should succeed despite exposure to only one talker. Failure would indicate the necessity of talker variation (Fennell & Waxman, 2010).

METHODS

Participants

Twenty-two infants (10 boys and 12 girls) between 13 and 15 months of age ($M = 13.7$ months $SD = 0.64$) participated. Six were excluded from analysis (4 for fussiness, 1 for parental interference, and one for failure to dishabituate during the novel trial), leaving 16 participants. Participants were screened for recent ear infections and came from monolingual English households. Infants received a t-shirt for participation.

Auditory Stimuli

The words used in this study were /buk/ and /puk/. Prior work in our laboratory has shown that 14-month-olds cannot learn these words in the no-variability switch task or in versions with only variability in relevant cues like VOT, but they can in the multitalker paradigm (Rost & McMurray, 2009, 2010). Auditory stimuli were recorded by a phonetically trained female talker using a Kay CSL model 4150 digital signal processor and a Shure WH3 microphone in a quiet room. To maximize acoustic variability naturalistically, the talker was instructed to use an infant directed register and vary the overall pitch (normal/high/low), pitch contour (rising/flat/falling) and length (normal/short/long) of each utterance (e.g., one /buk/ had a duration of 250 ms and a flat intonation around 175 Hz, while another had a duration of 200 ms and an intonation that rose from 150 to 210 Hz; see Table 1 for detailed measurements). This resulted in 119 utterances of both /buk/ and /puk/. The utterances were normalized for amplitude.

To assess the amount of variability in our stimuli, we measured 10 acoustic cues that are potentially informative for talker identity (see Table 2) for the present stimuli, the Rost and McMurray (2009) multitalker stimuli, and a new set of control stimuli.² Control stimuli consisted of seven repetitions of both /buk/ and /puk/ recorded by the two naïve female talkers. The talkers were told to use infant directed speech but not instructed to vary their utterances.

²The Rost and McMurray (2009) control experiment, unlike prior single-talker switch task experiments, used a single exemplar of each word. As a result there was no variation to measure. Thus we recorded new control items to simulate how standard versions of the switch task would work.

As expected, several acoustic cues are less variable in the current stimuli than in Rost and McMurray (2009): mean pitch, pitch excursion, pitch direction, and spectral tilt (Table 2). These are classically considered important cues to talker identity. Interestingly, several other cues were *more* variable in the current stimuli (duration, harmonic to noise ratio, spectral mean, spectral variance, spectral skew, and spectral kurtosis); these likely varied more in our stimuli as a result of instructing the talker to vary speaking styles. Levene's test for equality of variances revealed significant differences for every acoustic measurement ($p < 0.05$). Not surprisingly, the current stimuli were more variable on nearly every acoustic dimension we measured than were the control stimuli (with the exception of pitch excursion and spectral skew for speaker A, and mean pitch for speaker B, although differences were slight). However, only the differences in duration and harmonic to noise ratio were significant for both talker A and B, with the addition of spectral variance, kurtosis and skew for talker B ($p < 0.05$). This is most likely due to the small number of available data points ($n = 14$).

Nonetheless, the stimuli used in the current study were more variable than the kind of stimuli used in many versions of the switch task, and nearly as variable as the Rost and McMurray (2009) stimuli, though along different dimensions.

Visual Stimuli

Visual stimuli consisted of three images of uncommon objects on a black background (the same images used in Rost & McMurray, 2009, 2010): a pink ball covered with soft points, a yellow paddle-shaped object with a hole in the middle, and a blue semi-transparent toy resembling a large three sided jumping jack.

Apparatus

Experimental sessions were conducted in a curtained-off portion of a quiet, dimly lit room. Visual stimuli were presented on a 42" flat-screen monitor and auditory stimuli were presented using a pair of speakers located on either side of the monitor. A small infrared camera below the TV allowed the experimenter to code looking behavior online. Presentation of stimuli and computation of looking time was automated by the HABIT program (Cohen, Atkinson, & Chaput, 2004).

Procedure

Participants sat on their caregiver's lap approximately 24-inch from the monitor. Caretakers were instructed to look straight ahead, and both the caregiver and the experimenter listened to music over headphones to mask the auditory stimuli. The experiment consisted of two phases: habituation and test. During habituation trials, infants saw a still image of an object and heard seven exemplars of the corresponding auditory stimulus (/buk/ or /puk/) presented at two second intervals. The order of auditory stimulus presentation within trials was randomized for each individual. The associations between words and objects, and thus the pairings used during the habituation phase, were counterbalanced between subjects.

The habituation phase lasted 30 trials (with 15 /buk/ trials and 15 /puk/ trials) for a maximum of seven minutes or until the participant's looking time over a four trial window was 50% or less than the first four trials (a minimum of eight trials, or 1 minute 52 seconds).

Only looks to the center of the TV screen were counted toward total looking time, but each trial continued for all seven exemplars of the auditory stimulus.

A large number of auditory exemplars were available for each word (119). As a result 105 unique auditory stimuli were available for the habituation phase (15 trials of 7 stimuli), with the remaining 14 stimuli held out for the test phase. Given the maximum of 30 habituation trials, infants never heard the same token twice across habituation and test.

After habituation, the test phase began. The first two test trials were *same* and *switch* trials and the third was the *novel* trial. In the *same* trial, participants saw one object from habituation, paired with the matching word (Object-A Word-A). During the *switch* trial that same object was paired with the other, mismatching word (Object-A Word-B). Which word was heard, which object appeared on *same* and *switch* trials and the order of *same* and *switch* trials were counterbalanced across subjects. *Novel* trials were always third to measure overall attention to the experiment. Here, infants saw an object not seen previously, paired with a trained word (e.g., Object-C Word-A). The word used for the novel trial was counterbalanced. As with the habituation phase, the auditory stimuli used for the *same* and *switch* trials were never repeated during test.

RESULTS

Results resemble Rost and McMurray (2009, 2010), with longer looking time to *switch* than *same* trials, and longer looks still to *novel* trials (Figure 1). Data were analyzed in a mixed-design ANOVA with looking time as the dependent measure. The primary factor of interest was condition (*same*, *switch*, and *novel*), which was within-subjects. We also examined three between-subjects factors: the order of the *same* and *switch* trials; the object seen at test (yellow or pink); and the word/object pairings used during habituation (pink object → *buk* + yellow object → *puk* vs. pink object → *puk*, yellow object → *buk*).

We found a significant main effect of test condition ($F(2,16)=26.7, p<.01$).³ Simple main effects showed that looking time was greater in *switch* than *same* trials (*switch*: $M=7.7, SD=3.2$; *same*: $M=5.8, SD=2.2$; $t(15)=2.5, p<.05$), and greater in *novel* ($M=11.4, SD=2.1$) than *switch* trials ($t(15)=4.5, p<.01$). Cohen's effect size ($d=.63$) for the same/switch difference was similar to Rost and McMurray's (2009) results ($d=.60$) suggesting there is no additional benefit to multiple talkers over and above the variability present here. There was no main effect for test-object ($F(1,8)=2.9, p=.12$) nor pairing ($F<1$), and these did not interact with each other or with condition (all $p>.2$). There was a marginally significant effect of test order ($F(1,8)=4.4, p=.069$); infants who experienced same trials first had longer overall listening times ($M=9.2, SD=2.2$) than those who received switch trials first ($M=7.6, SD=1.2$). However, this did not interact with any other factor (all $p>.12$). Thus, infants showed evidence of learning when trained with multiple exemplars from the same talker.

³Looking times are generally right-tailed, violating the normality assumptions of ANOVA, however an ANOVA run on log-transformed looking times revealed comparable effects.

How Many Talkers Were Present?

Although the stimuli in the current study were produced by a single talker, they may have been perceived as *multitalker* utterances. To assess this, 15 adult listeners judged whether groups of seven /buk/ or /puk/ utterances were produced by a single talker or multiple talkers. There were eight test trials (4 buk/4 puk) consisting of seven exemplars from Experiment 1 (female voice). There were also 24 single-speaker trials selected from three new speakers (one female, two male) whose stimuli were recorded similarly to the control stimuli described above. Finally, there were 24 multiple talker trials containing a mixture of all four talkers.

Across the 15 participants 90% of the multiple talker trials were correctly identified as having multiple talkers; 99% of the control trials were correctly identified as a single talker, and 79% of the test trials were classified as single talker. Thus, although acoustic variability influenced the participants' ability to classify the experimental stimuli, they were predominantly classified as single talker utterances.

Given the results of this adult identification task, along with evidence that much older children (between 3 and 6 years old) are poor at identifying talkers from the same gender (Creel & Jiménez, 2012), it seems highly likely that the infants in this study perceived the stimuli as belonging to a single talker. However, even though older children have difficulty identifying individual speakers, infants as young as 7-months-old appear adept at discriminating similar sounding talkers (Johnson et al., 2011). Thus it remains a possibility that the stimuli used here were perceived by our participants as multiple talkers. While we cannot completely rule out this possibility, it does suggest that for accounts in which multiple-talkers serve as a cue for the referential nature of a situation, there must be a clear theory of how infants detect the presence of multiple talkers (and this too may develop).

GENERAL DISCUSSION

The success of 14-month-olds in the present study indicates an important role for general acoustic variability beyond multiple talkers. This finding affirms Rost and McMurray's (2009, 2010) conclusions and extends them to single talker versions of the switch task. This evidence resonates with what we know about the standard infant language environment. Caregivers are the primary source of speech input for infants. Even infants who are regularly exposed to more talkers presumably are not regularly exposed to the number of distinct talkers present in Rost and McMurray (2009, 2010). Instead, infants typically receive somewhat variable tokens from only a handful of talkers. How much acoustic variability is necessary for phonological development is still an open question. Perhaps the acoustic variability produced by a single talker is enough over time to learn which cues are important for word learning, and which are not. Indeed, the increased variability in infant directed speech likely enhances the natural variability that occurs across utterances even in adult directed speech (Bortfeld & Morgan, 2010). Our single-talker results suggest this may be useful for development.

Fennell and Waxman (2010) suggest that the presence of referential cues associated with multiple talkers may account for learning in the Rost and McMurray studies (2009, 2010).

While multiple talkers may provide converging evidence for the name of a particular object, as Fennell and Waxman suggest, the present findings refute the notion that multiple talkers are *necessary* for successful performance in this task. These results do not imply, however, that acoustic variability is the only relevant factor. For example, studies have demonstrated the effect of task difficulty on 14-month-old performance in early word learning tasks (Ballem & Plunkett, 2005; Yoshida et al., 2009), that the size of an individual's lexicon (Werker et al., 2002) predicts his/her performance, and of course, that the lack of social cueing within such tasks may contribute to the difficulties observed (Fennell & Waxman, 2010). We are not arguing that these factors play no role. Rather, the fact that perceptual categories are not yet complete make tasks such as the switch task difficult, and this in turn allows us to see the importance of these other factors. Thus, reducing task difficulty makes it easier for infants to succeed from the limited basis of their still developing speech perception skills. Crucially however, the switch task is hard because infants are still mastering the requisite perceptual/phonological skills at 14 months.

Together, this suggests that a critical developmental problem involves sorting out which cues are relevant for lexical access, and which are not. Schmale, Hollich, and Seidl (2011) have shown that children as old as 2 have difficulty generalizing newly learned words in one accent to another. This sensitivity to irrelevant acoustic cues during word learning speaks to an ongoing aspect of phonological development even later than ages typically studied in the switch task. In fact, remnants of this can even be seen in adulthood: Eisner and McQueen (2005) have shown that perceptual retuning of speech categories can extend only to talkers heard during training, and studies such as Creel, Aslin and Tanenhaus (2008) and McLennan and Luce (2005) suggest that indexical cues can activate specific words—words heard more frequently in one voice show greater activation when presented in the same voice. While such findings have often been claimed to support an episodic characterization to the lexicon, they may also be remnants of a developmental system that associates all cues with words and must learn to sort which are relevant using principles like variability. While no one would argue adults have incomplete phonological categories, this research suggests that such information is never completely irrelevant even in adulthood. It highlights the continuity of these processes across development and supports the associative model espoused here by suggesting a gradual process of sorting out the relationship between indexical and phonetic cues over phonological development.

Sensitivity to indexical cues by infants, toddlers and adults then may reflect the same underlying associative system at different stages of development. Of course, there is likely more to this than simply learning the relevance (or irrelevance) of indexical cues. Adult listeners also engage active processes that compensate for specific talker characteristics (e.g., McMurray & Jongman, 2011; Johnson, Strand, & D'Imperio, 1999), and it is currently unknown how this develops or how it relates to the much coarser process of determining what cues are relevant for word learning that we have emphasized.

The current study demonstrates that variability can influence performance in early word learning tasks via bottom-up information, but top-down information may play an important role as well. Thiessen (2007) demonstrated that infants benefit from exposure to two-syllable novel words in which one-syllable test words are embedded (*dawbow/tawgoo* when learning

daw/taw). Importantly, there is no benefit when test words are embedded in similar two-syllable words (*dawgoo/tawgoo*). Subsequent work has shown that this benefit generalizes to new vowel contexts only when familiarized to several different embedded vowel contexts (Thiessen, 2010). Thus, in both top-down and bottom-up approaches acoustic variability along dimensions that are not directly related to the phoneme contrast in question is critical. Critically it may be top-down information that enables infants to use the bottom-up acoustic variability—it is the presence of the common visual referent across all of the acoustic variability that allows the invariant cues to become noticed (Apfelbaum & McMurray, 2011; Yeung & Werker, 2009).

More generally, this adds to a growing body of evidence for the benefit of variability during learning in a variety of domains including second language learning (Lively et al., 1993), artificial grammar learning (Gomez, 2002), education (Richland, Bjork, Finely, & Linn, 2005), abstract skill learning (Helsdingen, Van Gog, & Van Merriënboer, 2011), and reading (Apfelbaum, Hazeltine, & McMurray, 2013). For example, varying the speed and distance of a target during training in motor-movement tasks improves posttraining performance (Catalano & Kleiner, 1984; Kerr & Booth, 1978). This suggests that variability in components of a task that are irrelevant to the ultimate outcome (e.g., talker voice) may be a broad principle of learning that applies in many domains.

This work clearly demonstrates that infants can succeed in the switch task without multiple talkers, and that their success with multiple talkers is likely due to acoustic variability in irrelevant acoustic cues. This is consistent with the view that during the second year of life speech perception is still developing, at least as it is applied to word learning. This also fits with numerous other switch task studies which have shown that performance is enhanced when the task difficulty is lessened, but importantly argues that this boost in performance is due not to the difficulty of the switch task, but to the immaturity of 14-month-olds phonological development.

Acknowledgments

The authors would like to thank Dan McEchron and the undergraduates of the MACLab for assistance recruiting babies; Ashley Farris-Trimble for recording stimuli; and Gwyneth Rost for helpful discussions.

FUNDING

This research is supported by NIH grant DC008089 to Bob McMurray.

References

- Apfelbaum KS, Hazeltine RE, McMurray B. Statistical learning in reading: Variability in irrelevant letters helps children learn phonics skills. *Developmental Psychology*. 2013; 49(7):1348–1365. [PubMed: 22924367]
- Apfelbaum KS, McMurray B. Using variability to guide dimensional weighting: Associative mechanisms in early word learning. *Cognitive Science*. 2011; 35(6):1105–1138. [PubMed: 21609356]
- Ballem KD, Plunkett K. Phonological specificity in children at 1;2. *Journal of Child Language*. 2005; 32(1):159–173. [PubMed: 15779881]
- Bortfeld H, Morgan JL. Is early word-form processing stress-full? How natural variability supports recognition. *Cognitive Psychology*. 2010; 60(4):241–266. [PubMed: 20159653]

- Catalano JF, Kleiner BM. Distant transfer in coincident timing as a function of variability of practice. *Perceptual and Motor Skills*. 1984; 58(3):851–856.
- Cohen, LB.; Atkinson, DJ.; Chaput, HH. Habit X: A new program for obtaining and organizing data in infant perception and cognition studies. Austin, TX: University of Texas; 2004. (Version 1.0)
- Creel S, Aslin RN, Tanenhaus MK. Heeding the voice of experience: The role of talker variation in lexical access. *Cognition*. 2008; 106(2):633–664. [PubMed: 17507006]
- Creel SC, Jiménez SR. Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*. 2012; 113:487–509. [PubMed: 22958962]
- Eisner F, McQueen JM. The specificity of perceptual learning in speech processing. *Perception & Psychophysics*. 2005; 67:224–238. [PubMed: 15971687]
- Fais L, Werker Janet F, Cass B, Leibowich J, Barbosa AV, Vatikiotis-Bateson E. Here's looking at you, baby: What gaze and movement reveal about minimal pair word-object association at 14 months. *Laboratory Phonology*. 2012; 3(1):91–124.
- Fennell CT, Waxman SR. What paradox? Referential cues allow for infant use of phonetic detail in word learning. *Child Development*. 2010; 81(5):1376–1383. [PubMed: 20840228]
- Fennell CT, Werker JF. Early word learners' ability to access phonetic detail in well-known words. *Language and Speech*. 2003; 46(2–3):245–264. [PubMed: 14748446]
- Galle ME, McMurray B. The development of voicing categories: A quantitative review of 40 years of infant research. *Psychonomic Bulletin and Review*. in press.
- Gómez R. Variability and detection of invariant structure. *Psychological Science*. 2002; 13:431–436. [PubMed: 12219809]
- Heldsdingen AS, Van Gog T, van Merriënboer JJ. The effects of practice schedule on learning a complex judgment task. *Learning and Instruction*. 2011; 21(1):126–136.
- Hollich GJ. Combining techniques to reveal emergent effects in infants' segmentation, word learning, and grammar. *Language & Speech*. 2006; 49(1):3–19. [PubMed: 16922060]
- Hollich GJ, Hirsh-Pasek K, Golinkoff RM, Brand RJ, Brown E, Chung HL, ... Bloom L. Breaking the language barrier: An emergentist coalition model for the origins of word learning. *Monographs of the Society for Research in Child Development*. 2000; 65(3):i–135. [PubMed: 12467096]
- Houston DM, Jusczyk PW. Infants' long-term memory for the sound patterns of words and voices. *Journal of Experimental Psychology: Human Perception and Performance*. 2003; 29(6):1143–1154. [PubMed: 14640835]
- Johnson EK, Westrek E, Nazzi T, Cutler A. Infant ability to tell voices apart rests on language experience. *Developmental Science*. 2011; 14:1002–1011. DOI: 10.1111/j.1467-7687.2011.01052.x [PubMed: 21884316]
- Johnson KC, Strand EA, D'Imperio M. Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics*. 1999; 24:359–384.
- Jusczyk PW, Pisoni DB, Mullennix J. Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition*. 1992; 43:253–291. [PubMed: 1643815]
- Kerr R, Booth B. Specific and varied practice of motor skill. *Perceptual and Motor Skills*. 1978; 46:395–401. [PubMed: 662537]
- Lively SE, Logan JS, Pisoni DB. 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 Acoustic Society of America*. 1993; 94:1242–1255.
- McLennan C, Luce PA. Examining the time course of indexical specificity effects in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*. 2005; 31(2):306–321.
- McMurray B, Jongman A. What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review*. 2011; 118(2):219–246. [PubMed: 21417542]
- Mullennix JW, Pisoni DB. Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*. 1990; 47:379–390. [PubMed: 2345691]
- Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. *Journal of the Acoustic Society of America*. 1989; 85(1):365–378.

- Namy L. Getting specific: Early general mechanisms give rise to domain-specific expertise in word learning. *Language Learning and Development*. 2012; 8(1):57–60.
- Nittrouer S. Challenging the notion of innate phonetic boundaries. *Journal of the Acoustic Society of America*. 2001; 110(3):1598–1605.
- Nygaard L, Sommers M, Pisoni DB. Speech perception as a talker contingent process. *Psychological Science*. 1994; 5:42–46. [PubMed: 21526138]
- Pater J, Stager CL, Werker JF. The lexical acquisition of phonological contrasts. *Language*. 2004; 80:361–379.
- Richland, LE.; Bjork, RA.; Finley, JR.; Linn, MC. Linking cognitive science to education: Generation and interleaving effects. In: Bara, BG.; Barsalou, L.; Bucciarelli, M., editors. *Proceedings of the Twenty-Seventh Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum; 2005.
- Rost GC, McMurray B. Speaker variability augments phonological processing in early word learning. *Developmental Science*. 2009; 12(2):339–349. [PubMed: 19143806]
- Rost GC, McMurray B. Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy*. 2010; 15(6):608–635.
- Schmale R, Hollich G, Seidl A. Contending with foreign accent in early word learning. *Journal of Child Language*. 2011; 38:1096–1108. [PubMed: 21310097]
- Singh L. Influences of high and low variability on infant word recognition. *Cognition*. 2008; 106:833–870. [PubMed: 17586482]
- Singh L, Morgan JL, White KS. Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language*. 2004; 51:173–189.
- Singh L, White KS, Morgan JL. 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*. 2008; 4(2):157–178.
- Slawinski E, Fitzgerald LK. Perceptual development of the categorization of the /r-w/ contrast in normal children. *Journal of Phonetics*. 1998; 26(1):27–43.
- Stager CL, Werker JF. Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*. 1997; 388(6640):381–382. [PubMed: 9237755]
- Swingle D, Aslin RN. Lexical neighborhoods and the word-form representations of 14-month-olds. *Psychological Science*. 2002; 13(5):480–484. [PubMed: 12219818]
- Swingle D, Aslin RN. Lexical competition in young children's word learning. *Cognitive Psychology*. 2007; 54(2):99–132. [PubMed: 17054932]
- Thiessen ED. The effect of distributional information on children's use of phonemic contrasts. *Journal of Memory and Language*. 2007; 56(1):16–34.
- Thiessen, ED. Variability in lexical form facilitates children's generalization of phonemic contrasts. Paper presented at the 17th Biennial International Conference on Infant Studies; Baltimore, MD. 2010.
- Van Heugten M, Johnson EK. Learning to contend with accents in infancy: Benefits of brief speaker exposure. *Journal of Experimental Psychology*. 2014; 143(1):340–350. DOI: 10.1037/a0032192 [PubMed: 23506084]
- Waxman SR, Gelman S. Early word-learning entails reference, not merely associations. *Trends in Cognitive Sciences*. 2009; 13:258–263. [PubMed: 19447670]
- Werker, JF.; Fennell, CT. Listening to sounds versus listening to words: Early steps in word learning. In: Hall, G.; Waxman, S., editors. *Weaving a lexicon*. Cambridge, MA: MIT Press; 2004. p. 79–109.
- Werker JF, Fennell CT, Corcoran KM, Stager CL. Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*. 2002; 3(1):1–30.
- Werker JF, Tees RC. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*. 1984; 7:49–63.
- Werker JF, Yeung H. Infant speech perception bootstraps word learning. *Trends in Cognitive Sciences*. 2005; 9:519–527. [PubMed: 16202639]

- Yeung HH, Werker JF. Learning words' sounds before learning how words sound: 9-Month-olds use distinct objects as cues to categorize speech information. *Cognition*. 2009; 113:234–243. [PubMed: 19765698]
- Yoshida KA, Fennell CT, Swingley D, Werker JF. Fourteen-month-old infants learn similar-sounding words. *Developmental Science*. 2009; 12:412–418. [PubMed: 19371365]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

FIGURE 1.

A depiction of the connectionist network used by Apfelbaum and McMurray (2011) to model the effect of acoustic variability on performance in the switch task. (A) The model trained on one talker. Thick arrows represent strong connections between both VOT and F0, and the visual items. (B) The model trained on multiple talkers. Thick arrows between VOT and the visual items represent strong connections between those two layers. Dotted lines between F0 and the visual items represent weak connections between those layers. This figure references Apfelbaum and McMurray (2011).

FIGURE 2.

Average looking time to each of the three test trials. Same trials represent presentations of the correct word/referent pairings, switch trials represent presentations of incorrect word/referent pairings, and novel trials represent presentations of new previously unseen referents.

TABLE 1

Means (Standard Deviations in Parentheses) for Pitch, Length and VOT for Present Stimuli

Word form	Pitch(Hz)	Length(ms)	VOT(ms)
/buk/	<i>M</i> = 201.49 (38.67)	<i>M</i> = 512.55 (83.80)	<i>M</i> = 12.25 (4.15)
/puk/	<i>M</i> = 221.23 (43.18)	<i>M</i> = 985.25 (621.72)	<i>M</i> = 47.57 (16.13)

TABLE 2

Standard Deviations of Acoustic Measurements for Two Sets of Control Stimuli, Stimuli from the Current Study, and Those Used in Rost and McMurray (2009) (Numbers with asterisks were significantly different from the present stimuli according to Levene's test for equality of variance. Numbers in parentheses represent stimuli from Experiment 3 of Rost & McMurray, 2010)

Acoustic Dimensions	Control A	Control B	Present Stimuli	Rost and McMurray
Mean Pitch (Hz)	28.76	75.06	37.73	77.54*
Duration (ms)	41.55*	42.30*	203.65	112.03*(101.86*)
Pitch Excursion (Hz)	70.76	28.10	64.48	137.22*
Pitch Direction (Hz/ms)	3.49	3.85	6.67	9.57*
Harmonic/noise ratio (dB)	1.79*	1.22*	3.95	3.07*
Spectral Mean (Hz)	175.75	100.59	198.13	151.19*
Spectral Variance (SD)	65.60	35.66*	90.54	59.64*
Spectral Skew	265.50	67.2*	251.34	153.91*
Spectral Kurtosis	67.65	28.53*	88.50	44.10*
Spectral Tilt (dB/kHz)	<0.001	<0.001	0.0020	0.0076*

Note. This table references Rost and McMurray (2009, 2010).