PAPER

Speaker variability augments phonological processing in early word learning

Gwyneth C. Rost\(^1\) and Bob McMurray\(^2\)

1. Department of Communication Sciences and Disorders, University of Iowa, USA
2. Department of Psychology, University of Iowa, USA

Abstract

Infants in the early stages of word learning have difficulty learning lexical neighbors (i.e. word pairs that differ by a single phoneme), despite their ability to discriminate the same contrast in a purely auditory task. While prior work has focused on top-down explanations for this failure (e.g. task demands, lexical competition), none has examined if bottom-up acoustic-phonetic factors play a role. We hypothesized that lexical neighbor learning could be improved by incorporating greater acoustic variability in the words being learned, as this may buttress still-developing phonetic categories, and help infants identify the relevant contrastive dimension. Infants were exposed to pictures accompanied by labels spoken by either a single or multiple speakers. At test, infants in the single-speaker condition failed to recognize the difference between the two words, while infants who heard multiple speakers discriminated between them.

Introduction

It is well known that, in the early stages of word learning, infants have difficulty learning phonologically similar words such as \textit{bih} and \textit{dih} (e.g. Stager & Werker, 1997). This stands in contrast to the canonical viewpoint that by this age, infants have good phonological representations that should provide a foundation for early word learning (e.g. Werker & Tees, 1984). Reconciling these findings has been a major endeavor in work on early word recognition and word learning (see Werker & Curtin, 2005, for a review).

The bulk of the evidence for the failure to learn minimal pairs has been provided by the switch task of Stager and Werker (1997). In this task, infants are first habituated to two objects paired with two words, and then are tested on two types of trials. On \textit{same} trials, they are exposed to an object–word pairing identical to that seen during habituation. On \textit{switch} trials they are tested on one object paired with the word it was not paired with during habituation; if infants successfully learned the words, they should dishabituate to this mismatch (see Werker, Cohen, Lloyd, Casasola & Stager, 1998, for a complete description). In essence, this task evaluates whether children have learned a pair of words robustly enough to be surprised by a misnaming. The switch task has demonstrated repeatedly that 14-month-olds are able to learn the pair sufficiently to notice this misnaming when the words differ by multiple phonemes (e.g. \textit{lifl neem}), but they consistently fail to notice it when the words differ by a single phoneme (e.g. \textit{bihl dih}) (Fennell & Werker, 2003; Pater, Stager & Werker, 2004; Thiessen, 2007; Werker et al., 1998; Werker, Fennell, Corcoran & Stager, 2002; see Mills, Prat, Zangl, Stager, Neville & Werker, 2004, for an ERP demonstration). Similar results have also been found in other word-learning tasks (e.g. Swingley & Aslin, 2007), suggesting that this difficulty with minimal pairs does not arise out of the switch task alone (although task does play a role – Ballem & Plunkett, 2005).

At the time, this finding seemed paradoxical. The canonical view of early speech development was that by this age, infants have largely acquired the phonology of their native language. They discriminate between the phonemes of their native language and fail to discriminate irrelevant non-native contrasts (Best, McRoberts, LaFleur & Silver-Isenstat, 1995; Best, McRoberts & Sithole, 1988; Pegg & Werker, 1997; Werker & Tees, 1984); they have formed graded phonemic categories (Kuhl, 1991; Miller & Eimas, 1996; McMurray & Aslin, 2005); and they can segment words from running speech using regularities of their native language (Jusczyk & Aslin, 1995; Jusczyk, Hohne & Baumann, 1999; Mattys, Jusczyk, Luce & Morgan, 1999). Most importantly, 14-month-olds do discriminate the relevant phoneme contrast (/b/ and /d/) when tested in a non-lexical auditory discrimination task: they dishabituate to \textit{bih} when habituated to \textit{dih} (Stager & Werker, 1997). A lexical context alone does not eliminate this ability to discriminate between similar phonemes: 14-month-olds recognize when known and
newly learned words are misproduced (i.e. ball for doll; vaby for baby; fope for vope) (Ballem & Plunkett, 2005; Fennell & Werker, 2003; Swingley & Aslin, 2002). Thus, the failure to acquire minimal pairs in the switch task would not appear to derive from a lack of ability to discriminate the two phonemes or from phonological ability in general.

This has largely led the field to assume that top-down factors are responsible for the failure to learn minimal pairs. Two such hypotheses have been proposed. First, word learning may impose attentional or cognitive demands that prevent the child from accessing all relevant abilities and knowledge (e.g. Werker et al., 1998). Second, it has been proposed that competition with known words (or between pairs of words as they are acquired) can account for such failure (e.g. Swingley & Aslin, 2007).

Werker and colleagues (Werker et al., 1998; Werker & Fennell, 2004) have suggested that the demands of word learning are too great to permit infants to engage their full range of phonetic skills. That is, learning a word requires an array of cognitive and perceptual abilities, including attention, segmentation, memory, inductive thinking, and the detection of referential intent; phoneme discrimination is only one of many components of word learning. This large set of requisite processes taxes infants’ limited cognitive and attentive resources, preventing them from taking full advantage of their phonetic abilities. As these abilities develop, and more general cognitive capacities expand, lexical neighbors cause less difficulty for word learners (Werker et al., 2002).

For children in switch-task experiments, the task itself may add additional demands above those of a discrimination or looking-preference task. The switch task requires that the child encode two acoustic forms, two visual forms, and form connections between those representations. This must be robust enough to withstand incorrect labeling: the switch task does not allow children the luxury of having both choices before them and choosing one (as a looking-preference task would). Rather, they must remember the correct naming, and determine that the visual referent does not match the one they had previously seen. In fact, in a less demanding task, 14-month-olds learn lexical neighbors (Ballem & Plunkett, 2005; Mani & Plunkett, 2008). Thus, by 14 months, the relevant phonological and cognitive processes may have developed sufficiently to perform discrimination and misperception tasks but may not be sufficient for more difficult tasks.

Alternatively, it has been suggested that lexical inhibition and competition processes (e.g. Luce & Pisoni, 1998; McClelland & Elman, 1986) might be responsible for children’s failure to learn lexical neighbors. For example, when children hear a non-word (e.g. tog) that is similar to a known word (dog), the known word is partially active. Over the course of processing, inhibition causes the known word to suppress activation for competitors, making it difficult to represent the non-word. Similarly, when learning two similar words (e.g. bih and dih), these partially learned words compete with each other. Ultimately, each inhibits the other, making it difficult to form unambiguous lexical representations.

In support of the hypothesis that lexical competition hinders infants’ learning of lexical neighbors, Swingley and Aslin (2007) demonstrated that Dutch-speaking children have difficulty learning [sont], a neighbor of hond (‘dog’) and [dal], a neighbor of bal (‘ball’) but not [bis] or [bemp], which are neighborless in the children’s lexicons. Complementing this, Thiessen (2007) showed that these competition processes are not always harmful: some constellations of words can overcome lexical competition. After learning dawgoo and tawgoo, 15-month-olds learning daw did not disambiguate to taw. However, when the presence of a near neighbor offered different acoustic information (daw, dawbow, and tawgoo), infants recognized that taw was a mispronunciation of daw. In this case, the ability to contrast daw and taw was enhanced only by words that shared sounds with the targets but were different from one another. Finally, Werker et al. (2002) report that the size of the lexicon in 17-month-olds is correlated with their ability to succeed at the switch task, although whether the lexicon itself drives the switch-task results or both are mediated by some other cognitive factors is unknown.

Both approaches can account for the available data, but neither is complete. It is not clear why phonetic discrimination (one of the earliest skills acquired, and fundamental to word learning) would not be preserved in the face of capacity limits. Similarly, it is not clear why early phonological representations would not provide enough discriminatory power to overcome lexical competition. While top-down factors contribute a piece of the explanation, co-developing perceptual and phonological abilities may help fill in the gaps (as hypothesized by PRIMIR; Werker & Curtin, 2005).

To date, there have been few investigations of the role of bottom-up factors in the acquisition of lexical neighbors. One exception to this is Nazzi (2005), who demonstrated that children have more difficulty learning neighbors that differ by vowel than by initial or medial consonant. However, participants were 20 months old, an age at which infants succeed at consonantal pairs in the switch task (Werker et al., 2002), and Nazzi’s task was substantially different from standard switch designs. Thus, while revealing new dimensions of difficulty (vowels), it cannot explain 14-month-olds’ failure to learn minimal pairs differing by consonants. It does nonetheless support the idea that perceptual or phonological processes relevant for word learning are still developing at 20 months.

If this is the case, it is possible that the phonetic representations available to 14-month-olds are sufficient for relatively simple tasks tapping discrimination (e.g. Werker & Tees, 1984), or misproduction (e.g. Swingley & Aslin, 2002), or more supportive word-learning tasks (e.g. Ballem & Plunkett, 2005). However, these same phonological representations may not be sufficient to overcome lexical competition or task demands.
Bottom-up input in early word learning

In considering the contrast between phonological discrimination and the acquisition of lexical neighbors, it is important to consider the nature of phonetic categories. If phonetic categories were represented as boundaries, studies demonstrating discrimination of phonemes would provide firm evidence that such categories were well developed by 14 months and should support lexical contrast. However, work on adult speech perception suggests that phonological categories are represented as either graded prototypes (e.g. Miller, 1997, 2001; see also Kuhl, 1991; McMurray, Aslin, Tanenhaus, Spivey & Subik, 2008) or clusters of exemplars (Goldinger, 1998), not as boundary-defined categories. Likewise, infants’ representations show similar properties of attunement to prototype structure in both consonants and vowels (Kuhl, 1991; Miller & Eimas, 1996; McMurray & Aslin, 2005). Thus, categories are described by their prototypical (or most frequent) values and by the range of acceptable variation.

This representation is well suited to the task of identifying positive exemplars of a category, a process that would be sufficient for simple discrimination in habituation-type tasks. However, identifying negative exemplars is much more difficult. Because category membership falls off in a graded manner as a stimulus departs from the prototype, there is no clear line over which a given token is clearly not a category member, yet the ability to make this judgment is essential for performance in the switch task and in learning tasks like Swingley and Aslin’s (2007) in which infants must notice that a non-word (e.g. tog) is not a member of a known word category (dog). Thus, the phonetic representations available to listeners may make word learning more difficult than discrimination.

The fact that such a representation does not easily support minimal-pair learning is consistent with the task demands framework of Werker and colleagues (Werker et al., 1998; Werker & Curtin, 2005). This is particularly true given the aforementioned evidence that when the task is structured in a way that maximizes the effectiveness of such categories, sensitivity to phonemic contrast is found (Ballem & Plunkett, 2005; Mani & Plunkett, 2008; see also Mani & Plunkett, 2007; Swingley & Aslin, 2002). However, this approach differs in a few key ways.

First, the task demands do not come from external capacity limits, nor do they force infants to ignore phonetic information. Rather, the limitation arises out of the nature of early phonetic representations, and needs of word learning. Second, it suggests that in addition to manipulating top-down factors (e.g. task) to understand the failure to learn lexical neighbors, we may also need to examine the structure, use and acquisition of phonetic categories themselves.

Use and acquisition of phonetic categories

While phonetic categories defined by prototypes and variation may not be optimal for learning lexical neighbors, this type of representation may be functionally useful for understanding language, as actual language use requires flexible boundaries. Mispronunciation is normal; infant-directed speech contains production parameters that are more variable than, and different from, adult-directed speech (e.g. England, 2005); and social or dialectic factors can create systematic pronunciation variability. It may therefore be adaptive for infants to map mispronunciations like vaby to targets like baby in the absence of evidence that vaby is being used to refer to a new word – especially if infants cannot be confident in their developing phoneme categories. The benefits of such tolerance to mispronunciation have been demonstrated in adult listeners (McMurray, Tanenhaus & Aslin, 2009). Moreover, 14- and 15-month-olds are willing to identify mismatching words (e.g. vaby/baby) as available known referents even though they are sensitive to the mismatch (Mani & Plunkett, 2007; Swingley & Aslin, 2002).

Thus, while the demands of word recognition require infants to restrict word forms to a small, usable representation, infants are also under pressure to be flexible in what counts as a word. Achieving categories that can support this type of flexibility may take considerably longer than one year. In fact, it is known that phonological development continues well into childhood (Edwards, Beckman & Munson, 2004; Munson, Swenson & Mathei, 2005).

Moreover, the nature of phonological representations seen in childhood would seem to support exactly this sort of flexibility. For example, Slawinsky and Fitzgerald (1998) demonstrated that the perceptual boundaries for approximants (/w/ vs. /l/) sharpen considerably between 5 and 9 years of age. Importantly, the shallow slope seen in 5-year-olds would permit more ambiguity in what counts as a category member. Likewise, children are still learning to use multiple cues and context for singular contrasts well into childhood (Hicks & Ohde, 2005; Mayo & Turk, 2004; Morrongiello, Robson, Best & Clifton, 1984; Nittouer, 2002; Ohde & Haley, 1997). Finally, the dramatic vocabulary growth of later childhood may have a significant effect on the development of phoneme perception, as there is evidence that perceptual processes are altered by the structure of the lexicon (McClelland, Mirman & Holt, 2006; Magnuson, McMurray, Tanenhaus & Aslin, 2003; Newman, Sawusch & Luce, 2005). Thus, acquiring a flexible phonemic representation based on prototypes or exemplars may not be complete by 14 months.

As a result of this, an understanding of the mechanisms by which such categories emerge may yield a way to augment them and drive word learning during this period. More importantly for the present purposes, it provides a test of the importance of phonological factors in predicting performance in word-learning tasks.

One important account of the development of speech categories is the distributional learning hypothesis (Maye, Werker & Gerken, 2002; see also McMurray, Aslin & Toscano, in press; Maye, Weiss & Aslin, 2008). Under this view, speech cues in the environment form statistical
clusters (e.g. Lisker & Abramson, 1964) and simple statistical learning mechanisms extract these clusters from the underlying categories. The listener develops phoneme categories by calculating the mean and allowable variance for any given phoneme. If the variance of a phoneme category (i.e. the range of acceptable tokens) is not well estimated, the listener might be unsure if two tokens are members of a single, wide, phoneme category, or are representative of two different phonemes.

Alternatively, exemplar views of speech perception (e.g. Goldinger, 1998) posit that phonological categories emerge out of large sets of accumulated exemplars of individual words. As under statistical accounts, the range of exemplars that have been recorded is crucial: only by gathering a sufficiently broad sample can accurate categories emerge (see Pierrehumbert, 2003, for a critical review). Thus, under both views of phonological categorization, a single exemplar of each word – or even a small set of exemplars – may be inadequate to create a sufficiently robust category for the word.

**Multiple exemplars**

Current versions of the switch task use either a single instantiation of the auditory token or a small range of very similar ones (i.e. spoken by the same speaker in the same context). Thus, infants are placed in exactly the situation in which both approaches predict difficulty in forming functional categories. Importantly, if infants come to the lab with undeveloped or partially developed categories, this might present an obstacle to acquiring sufficient information during habituation. Even if phonological categories are largely developed, a lack of variability may hinder the maintenance of these categories: a single exemplar repeated over and over would warp the representativeness of a set of exemplars, and could have biasing effects on the phoneme being estimated. Importantly, under both statistical and exemplar models, variability would help infants augment or maintain their still-developing phonological categories in a way that assists with making lexical contrast, and would allow 14-month-olds to learn lexical neighbors.

Studies of visual category learning in infancy are consistent with this. These studies have shown that infants trained on a single exemplar or a low variability set will discriminate individual exemplars (i.e. they do not assume a category), while infants trained on multiple exemplars will discriminate tokens that don’t belong to the trained category (Younger & Cohen, 1986). Additionally, the amount of variability or similarity in the training set affects which items infants will assign to the category (Oakes, Coppage & Dingel, 1997; Quinn, Eimas & Rosenkrantz, 1993).

Multiple-exemplar tasks may also facilitate the acquisition of auditory word-form categories. Indeed, Singh (2008) demonstrates that variation in affect can help 7.5-month-olds segment words from running speech. In adults, training new speech categories via multi-talker input has been shown to improve both discrimination and generalization of novel phoneme categories (Lively, Logan & Pisoni, 1993; McClelland, Fiez & McCandliss, 2002). Multi-talker training is therefore a theoretically and ecologically valid way to support auditory encoding of lexical categories.

Multiple-exemplar training also increases the top-down task demands on the child in at least two overlapping ways. First, during habituation and testing, it requires them to normalize quickly for different speakers, pitches, and speaking rate, something that imposes a significant delay on normal word recognition (Mullennix, Pisoni & Martin, 1989; Ryalls & Pisoni, 1997). Second, learning a word in this situation requires them to maintain more items in memory, and they must be stored in more detail. In addition, a manipulation of acoustic variability does not directly manipulate lexical structure or competition (assuming that the variable tokens are not phonetically ambiguous). Thus, success at a multiple-exemplar version of the switch task would be difficult to account for with explanations based on top-down factors such as capacity limits or lexical mechanisms.

Though multiple-exemplar habituation can control for top-down factors, it does not rule out their broader role in learning. In fact, the hypothesis that bottom-up factors can augment learning in the moment is compatible with both attentional-demands and lexical-competition hypotheses for failure in previous instantiations of this experiment. As we have argued, 14-month-olds’ failure to learn minimal pairs lies at the nexus of perceptual/phonological factors and the top-down demands of word learning situations like the switch task.

The role of acoustic/phonetic variability was tested in two switch-task experiments. The first used single-exemplar habituation and test similar to those reported by Werker and colleagues. The second employed multiple-exemplar habituation and test, in which 54 auditory exemplars from 18 speakers were used to train and test the words.

**Experiment 1**

Experiment 1 used the switch task developed by Werker et al. (1998; Stager & Werker, 1997), with only four minor changes to the design. First, we added a completely novel object to the end of the test sequence. Because the expected result was a null effect, dishabituation to this object would provide evidence that infants were learning something during habituation. Second, single-color real objects were used for the novel objects, rather than fabricated multi-color objects. Third, we used photographs of the object, rather than moving film, because Fennell and Werker (2003) reported that still photos yielded the same results. Fourth, the words *buke* (/buk/) and *pook* (/puk/) were used rather than Stager and Werker’s (1997) original *bih* and *dih*. Because *bih* and *dih* violate
the phonology of English (/l/ is not permissible in word-final position), this change was expected to create a more natural, easier learning situation, and it has been shown that infants also fail to learn lexical neighbors with more word-like referents (Pater et al., 2004).1

Method

Participants

Thirty-three monolingual English-learning 14-month-olds (between 13:9 and 14:29) participated in Experiment 1. The infants were recruited via mail and phone from a database of county birth records, and were considered eligible if they were normally developing and without history of ear infection according to parental report. Data from 17 were excluded because the infants did not complete the experiment due to fussiness (12), because they did not habituate to the training set (three), or because they were learning a language other than English (two). Sixteen children, nine boys and seven girls, formed the resulting experimental set.

Stimuli

Experimental stimuli consisted of three digital photographs and two sound files. The photographs were of single-color novel objects (Figure 1) photographed against a black background. Two of the pictures were designated as habituation stimuli, the third was presented only at test as a novel control stimulus. Pilot testing of the photographs had revealed that none of the three objects was inherently more preferable than the others to infants of a similar age. The sound files contained one exemplar each of /buk/ and /puk/ recorded by an adult female native-English speaker who said the words in an infant-directed register. Each word was copied and spliced to itself so that the resulting sound file contained seven presentations of the same exemplar at 2-second intervals. The picture and sound presentations were synchronized so that each picture appeared for 14 seconds, while the word was presented seven times.

Apparatus

The testing booth was a curtained-off portion of a laboratory room. A comfortable chair faced a flat-screen monitor with stereo sound speakers mounted on either side of the monitor. A small infrared camera was situated below the screen, which allowed the experimenter, who could neither see nor hear the stimulus, to code looking time online. Interexperimenter reliability was above 90%. The HABIT computer program (Cohen, Atkinson & Chaput, 2004) controlled item presentation and data capture.

Procedure

After informed consent was obtained, infants and parents were shown the testing room and apparatus where infants were seated on the caregiver’s lap and testing began. Both the experimenter and the caregiver listened to music over headphones at a level loud enough to mask the auditory stimuli.

Each habituation and testing trial lasted for 14 seconds and consisted of a single still picture paired with seven

---

1 While there was some concern that buke may be too similar to the English word, book, we wanted to restrict the set to Stop-Vowel-Stop sequences that contrast in voicing, and the English lexicon contains very few such pairs for which both are non-words.
repetitions of a single auditory stimulus (at 2-second intervals). During habituation, looking time to the visual stimulus was measured and infants were considered habituated when their looking time over a four-trial window was 50% of their looking on the first four trials. This is the more stringent of the two habituation criteria that have been used by Werker et al. (Werker et al., 2002; Fennell & Werker, 2003; see Stager & Werker, 1997; Werker et al., 1998; and Pater et al., 2004, for a less stringent criterion), and is consistent with Casasola and Cohen’s (2000) parameters for switch-task habituation. The order of habituation trials was pseudo-randomized such that no word–object pairing appeared more than twice consecutively.

After reaching habituation, infants were tested on three trials. Same trials consisted of one of the habituation objects paired with the same word. Switch trials consisted of the same object (as the same trial) paired with the opposite word. Finally, novel trials consisted of an object the infants had not yet seen paired with one of the two words. The order of test trials was counter-balanced, with half of the children receiving the same trial first followed by the switch and half receiving the switch trial first followed by the same. In order to maintain continuity with prior work, the novel trial was always presented third. The object–word pairing was counter-balanced across participants.

Results and discussion

Results are shown in Figure 2 and largely replicate prior work (e.g. Stager & Werker, 1997). Data were analyzed using a mixed-design ANOVA with two factors. First, test condition (same, switch, and novel) was a within-subject factor and the primary variable of interest. Second, as a consequence of the switch task, infants were tested on only one of the two words (/buk/ or /puk/). It was therefore important to determine if the effect of test condition was similar across both test words. Thus, test word was included as a between-subjects factor.

We found a significant main effect of condition ($F(2, 28) = 11.375, p < .01$). There was no main effect of test word ($F < 1$) or interaction between test word and condition ($F < 1$), indicating that infants’ responses were not driven by a preference for one of the words. Planned comparisons determined that the effect of condition was driven by the novel object trial. Infants did not show a significant difference in looking time between the same ($M = 6.11$ sec, $SD = 3.24$) and switch ($M = 6.31$ sec, $SD = 3.78$) trials ($F < 1$), suggesting that they failed to learn this pair of novel words. However, infants did look to the novel object ($M = 10.28$ sec, $SD = 3.13$) trial significantly longer than same and switch trials ($F(1, 14) = 22.9, p < .01$).

Infants responded to this task as expected: they failed to distinguish between lexical neighbors that had been paired with visual stimuli. However, the fact that infants dishabituated to a new visual token suggests that they do learn during the task; specifically, they encode visual information. While it has been surmised that the phonological similarity in the training set does not block learning of the visual form, this has not been explicitly tested; the use of the novel object with the familiar word confirmed that visual representation was established during the habituation stage.

We were concerned about our selection of /buk/ as one of our word pairs, as the similar sounding word, ‘book’, is well-known by this age (Dale & Fenson, 1996). However, our replication of Stager and Werker’s (1997) failure to learn minimal pairs suggests that the presence of the similar known word did not play a role here; this is confirmed by the absence of statistical interaction between condition and word. Moreover, because Experiment 2 used the same pairs, any evidence for learning found in that experiment must be due to the experimental manipulation: acoustic variability.

Having replicated infants’ failure to learn phonologically similar words in Experiment 1, we next assessed our hypothesis by using multiple-exemplar habituation.

Experiment 2

The second experiment tested our primary hypothesis that variability during habituation helps listeners form more robust lexical categories and that variability during test helps infants respond appropriately to word–object mismatch. This variability was achieved through the use of multiple auditory exemplars from a variety of talkers and speech registers. Thus, the exemplar set contained variability in indexical, prosodic, and phonetic (e.g. voice-onset time) dimensions. It was hypothesized that this would help the children form more robust auditory word-form representations that would allow them to notice the switch.
Method

Participants

Recruitment and inclusion criteria were the same as in Experiment 1. Twenty-two 14-month-olds (13;05–15;0) participated in Experiment 2. Data from six were excluded because the children failed to habituate (three), were unable to complete the experiment due to fussiness (two), or had a history of recurrent ear infections (one). The remaining 16 children, ten boys and six girls, formed the experimental group.

Stimuli

Eighteen adult native-English speakers recorded a series of /buk/s and /puk/s in an infant-directed register. Three tokens of /buk/ and /puk/ were spliced from each of the recordings, and the final set of prepared exemplars was then normalized for amplitude. This resulted in a set of 54 exemplars of each word. This set included three tokens of /buk/ and /puk/ from the speaker used in Experiment 1 (one of the three tokens was the token used in Experiment 1).

In both the training and test phases, each 14-second trial contained seven different exemplars of the word from seven different speakers at 2-second intervals, and each child was trained and tested on an individually randomized set of tokens. These seven exemplars were pseudo-randomly selected in advance. This was constrained such that no voice or exemplar was repeated during any given presentation and such that one exemplar of each speaker saying each word was held out for test. Each child therefore heard 36 different exemplars of each word (in all 18 voices) over the course of the habituation, and seven previously unheard exemplars of each word during the testing phase (though the speakers were familiar).

Apparatus and procedure

The experimental set-up and procedures were identical to Experiment 1.

Results and discussion

Data were collected and analyzed in the same manner as in Experiment 1, using a mixed-design ANOVA in which test condition (same, switch, and novel) was the within-subject factor, and test word (/buk/ or /puk/) was the between-subjects factor. Results are shown in Figure 3. There was a main effect of test trial ($F(2, 28) = 7.8, p = .002$). Again, there was no main effect of test word ($F(1, 14) = 1.35, p = .26$). There was a marginal interaction between testing condition and looking response time ($F(2, 28) = 3.19, p = .06$), driven entirely by responses to the novel object. Thus, learning was not affected by infants’ preference for either word or prior experience with book.

Planned comparisons on the effect of test condition showed a different pattern from Experiment 1. Infants looked at the switch trial ($M = 6.96$ sec, $SD = 3.54$) significantly longer than the same ($M = 4.95$ sec, $SD = 3.01$) trial ($F(1, 14) = 7.1, p = .018$), and this did not interact with test word ($F < 1$). Additionally, infants looked at the novel object trial ($M = 8.69$ sec, $SD = 4.04$) significantly longer than the same and switch trials ($F(1, 14) = 8.2, p = .013$).

Infants in the multiple-exemplar switch task succeeded at learning two phonologically similar words well enough to notice misnaming at test. They noticed the misnaming both in switch trials, where a familiar word was used with an incorrect (but familiar) object, and in novel trials, where a familiar word was used for an unfamiliar and perceptually dissimilar object. Thus, the multiple-exemplar task presented children with sufficient information to succeed at the switch task. It is therefore possible that switch-task failure can be attributed in part to acoustic/phonetic processes sensitive to the structure of the input.

Because it is likely that children take longer to habituate to a set of varied exemplars than to a set of similar ones, one immediate question was whether children in Experiment 2 simply got more experience with the word–object pairings than children in Experiment 1 due to longer habituation times. This was not the case, however, as there was no significant difference between infants in the two experiments in number of habituation trials (Experiment 1 = 18.4 habituation trials, Experiment 2 = 18.3; $T(15) = .02, p = .93$), nor in total looking time (Experiment 1 = 160.6 sec, Experiment 2 = 152.6; $T(15) = .46, p = .66$). Thus, it was not the amount of exposure, but the type of exposure, that helped children succeed at the switch task.

Figure 3 Looking times in Experiment 2. Error bars are Standard Error of the Mean.

© 2009 The Authors. Journal compilation © 2009 Blackwell Publishing Ltd.
A second question was whether the differences seen across the two experiments arose because of acoustic differences in the stimuli. Acoustically, pairs like /buk/ and /puk/ differ primarily in voice onset time (VOT), the time difference between the release of the lips and the onset of laryngeal vibration. If the difference between the VOTs of the two targets were greater in Experiment 2 than in Experiment 1, the learning effects could have arisen from the fact that the two words were, on average, more acoustically discriminable in this experiment. Again, this was not the case. The /buk/ used in Experiment 1 had a VOT of 9 msec, and the average VOT of /buk/ tokens in Experiment 2 was 11 msec (SD = 3). The /puk/ from Experiment 1 had a VOT of 79 msec, while the tokens in Experiment 2 had average VOT of 80 msec (SD = 26).3 The lack of difference in VOT between exemplars of the two experiments suggests that it was precisely the variability in the multiple-exemplar experiment that helped children succeed at this task.

General discussion

The variability in the multiple-exemplar version of the switch task improved infants’ performance by providing them with a richly defined category for the words. Though prior phonetic experience undoubtedly contributed to their performance, the failure in Experiment 1 suggests that it was insufficient for the task. However, Experiment 2 provided the infants with a set of input that allowed them either to estimate the variability of the categories or to maintain or augment their existing estimates. Given this, the ability to make lexically relevant phonological contrast emerged in the moment, allowing them to succeed at the task.

However, this is not to argue that top-down factors are irrelevant. Given evidence for online integration of lexical information and phonetic percepts (McClelland et al., 2006; Magnuson et al., 2003), phonetic boundaries, while not represented explicitly by the system, are best seen as an emergent property of bottom-up, graded phonetic categories, and top-down influence from the lexicon. Facing those difficulties imposed by their sparse lexicons and/or other more general cognitive limitations, 14-month-olds in the switch task may have to rely heavily on bottom-up perceptual processes to determine that a given auditory token is not a member of a lexical category. The results of the first experiment (as in prior switch-task studies) would seem to imply that these categories were insufficient to overcome these difficulties. However, brief exposure to multiple exemplars allowed infants to buttress their representation of the input to make a contrast between /buk/ and /puk/. Whether this represents a long-term strengthening of these phonetic categories or a short-term learning phenomenon is unclear. Moreover, it is not clear whether the variability during habituation or at test was more relevant. However, it importantly reveals that the phonological categories that underpin early word learning may not be fully formed, and that variability may play a critical role in the learning mechanisms that augment them.

There are at least two kinds of relevant variability—and hence two kinds of learning mechanisms—that may be important. First, variability along specifically phonetic dimensions (in this case, VOT) may have allowed the infants to define the phonetic or lexical categories that contrasted the words. This would require learning mechanisms of the sort demonstrated by Maye et al. (2002; see also Maye et al., 2008). This approach posits that infants track the frequencies of specific phonetic cues (e.g. VOT) and extract categories from the natural clusters (perhaps interacting with other mechanisms like competition: McMurray, Aslin & Toscano, in press). Accordingly, the variation within the relevant acoustic category found across the multiple exemplars in Experiment 2 is what is crucial for defining the category in this task. In fact, measurements of VOT reveal considerable variation along this dimension for both the voiced (M = 11 msec; SD = 3; Range = 5–20) and voiceless (M = 80 msec; SD = 26; range = 39–141) categories. Thus, this exposure offered by Experiment 2 may fit the bimodal distribution of VOTs required by this sort of learning mechanism.

Second, it is possible that variability in non-phonetic information helped infants extract the relatively invariant phonetic dimensions. That is, infants at 14 months may still be unsure about what dimensions are relevant for the task, and variability in irrelevant aspects of the stimuli improve performance by focusing attention on those aspects of the input that are comparatively stable. Such a mechanism would be analogous to learning processes posited by Gómez (2002). She demonstrated that when learning sequences of syllables in which nonadjacent syllables were predictable, the variability in the irrelevant, intervening stimuli was a crucial determinant of learning: when the set size of possible intervening elements was large, infants learned the non-adjacent dependencies, while a small set-size led to failure. Thus, the relatively stable elements of a stimulus set become increasingly easier to extract as variability increases (see also Yu & Smith, 2007, for an example in word learning). In fact, measurements of pitch and the first four formants (measurements of vowel quality) of Experiment 2 stimuli were all highly variable (see Table 1). Most importantly, none of these cues differed significantly between /buk/ and /puk/, suggesting that they would not be available to directly contrast the words. Nonetheless, the immense amount of irrelevant variation present would provide the

---

3 Eighty msec seems like a relatively long VOT for a /pf/ given Lisker and Abramson’s (1964) classic study (which found a mean VOT for voiceless stops in English of 53 msec). However, Allen and Miller (1999) found that in slow rates of speech, voiceless labials averaged 64 ms, and stops as a whole averaged 78 ms. Moreover, the small literature in infant-directed speech (e.g. Englund, 2005), along with data in preparation from our lab, indicates that VOTs are significantly increased in infant-directed speech. Thus, this is not an unexpected VOT.
necessary fodder for the sort of learning mechanism that uses non-criterial variation to extract the invariant elements from a noisy signal.

Both sources of variability – the criterial voice onset time, and the non-criterial indexical and prosodic information – were available in the stimulus set of Experiment 2. The current data offer no insight into which of these two coexisting sources of variability provided infants with the necessary information to learn two lexical neighbors, or if infants harnessed both. It will be important for future work to control each type of variability more precisely (e.g. hold VOT constant, and vary only speaker, or use a single speaker and vary VOT) to determine which type of variability predicts word learning.

These results do not argue that variability is a good thing in general. It is more important that the input contain the appropriate statistical structure for a given learning mechanism, and that variability along some dimensions (criterial or non-criterial) will typically be part of this. This may point toward an explanation for Nazzi's (2005) demonstration of difficulty learning minimal pairs that differ by vowel. Vowels in general have more overlapping distributions along criterial dimensions, but also (unlike consonants) these cues are not necessarily independent of non-criterial cues like pitch, duration, and timbre. Children's behavior in tasks like these will thus be an emergent property of the statistics of prior history, the statistics of immediate learning, the structure of the lexicon (which may support some contrasts over others) and the demands of the task.

Nonetheless, this work makes it clear that children do not fail at the switch task because they have a limited capacity for acoustic detail. Capacity limits or task demands do not provide a compelling explanation for these results because (a) we did not manipulate task-specific aspects, and (b) our purely acoustic/phonetic stimulus manipulations actually add necessary processes to the task (speaker normalization, memory). Rather, these results suggest that the failure may arise from an interaction between incompletely developed phonetic categories, statistically impoverished input, and the unique demands of the switch task. When the bottom-up input is manipulated in a way that is sensitive to the mechanisms used to extract phonetic categories, infants can learn lexical neighbors.

Acknowledgements

We would like to thank Brandon Abbs, Tracy Ball, Allison Bean, Katie Bresson, Angelo LaRocca, John Lipinski, Dan McEchron, Cheyenne Munson, Amanda Murphy, Amanda Nemtabashk, Brooke Overgard, Sammy Perone, Molly Robinson, Scott Spilger, Joe Toscano, Beth Walker, and Jed White for recording /buk/ and /puk/s for us. We are also indebted to Kristine Kovack-Lesh and Sammy Perone for assistance with HABIT, and particularly thank Sammy for suggesting that we include the novel object trial. We thank Janet Werker and Chris Fennell for helpful comments during the development of this project; Karla McGregor for her comments on an early draft; and two anonymous reviewers for their thoughtful and careful comments. Finally, we thank the many families who have so generously donated their time to our efforts.

References


Received: 12 July 2007
Accepted: 30 March 2008