Published in final edited form as:
Infancy. 2010 November 1; 15(6): . doi:10.1111/j.1532-7078.2010.00033.x.

Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning

Gwyneth C. Rost and
Department of Communication Sciences and Disorders, University of Iowa

Bob McMurray
Department of Psychology, University of Iowa

Abstract

It is well attested that 14-month olds have difficulty learning similar sounding words (e.g. bih/dih), despite their excellent phonetic discrimination abilities. In contrast, Rost and McMurray (2009) recently demonstrated that 14-month olds' minimal pair learning can be improved by the presentation of words by multiple talkers. This study investigates which components of the variability found in multi-talker input improved infants' processing, assessing both the phonologically contrastive aspects of the speech stream and phonologically irrelevant indexical and suprasegmental aspects. In the first two experiments, speaker was held constant while cues to word-initial voicing were systematically manipulated. Infants failed in both cases. The third experiment introduced variability in speaker, but voicing cues were invariant within each category. Infants in this condition learned the words. We conclude that aspects of the speech signal that have been typically thought of as noise are in fact valuable information – signal – for the young word learner.

Keywords

infant; speech perception; word learning, switch task, variability

Research in early language acquisition has been peppered with findings that very young infants have excellent abilities to discriminate speech categories (e.g., Eimas, Siqueland, Juczkyk & Vigorito, 1971; Werker & Tees, 1984; Werker & Curtin, 2005 for a review). However, Stager and Werker (1997; Werker & Fennell, 2006 for a review) reported that for somewhat older infants (14-month olds), some of these abilities appear to be ineffective when applied to word learning. Phonological skills such as the ability to discriminate between native-language phonemes (Werker & Tees, 1984), and to represent the phonology of words in detail (Ballem & Plunkett, 2007; Swingley & Aslin, 2002) seem to have little bearing on the ability to learn words that are phonologically similar (Stager & Werker, 1997; see also Swingley & Aslin, 2007; Werker, Cohen, Lloyd, Casasola, & Stager, 1998; but see Fennell & Werker, 2003 and Swingley & Aslin, 2002 for recognition of known words).

Explanations for the failure to learn phonologically similar words typically focus on top-down mechanisms such as task demands (Werker et al, 1998; Yoshida, Fennel, Swingley & Werker, 2009) or lexical access (Swingley & Aslin, 2007). Proponents of the former argue that the demands of laboratory word-learning tasks are heavy because the children are required to encode both visual and auditory forms in a short time period and then to connect them to one another. This requires children to allocate their limited resources to specific elements of the task (Werker & Fennell, 2006 for a review). PRIMIR (Werker & Curtin, 2005) describes this as a case where general perceptual processes overwhelm the child's system, leaving little room for phonetic ones. Additionally, the *switch* task typically used in these experiments (see Werker, et al., 1998) requires that information be represented and organized robustly, as success requires the infant to determine that something is *not* part of a category. Children this age succeed more easily at positive identification tasks in which they must map an auditory word form to an object (Ballem & Plunkett, 2007). Even infants trained in the style of Stager and Werker (1997) correctly identify word-object pairings when the test is presented using a two-alternative looking paradigm (Yoshida et al, 2009). Lack of capacity coupled to the difficulty of the switch task might negatively affect 14-month-olds' use of their discrimination skills in this task. However, as children get older, they become more adept, and by 20 months, they learn phonologically similar words in the switch task (Werker, Fennell, Corcoran, & Stager, 2002).

Alternatively, it has been suggested that processes involved in lexical access, particularly competition (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Luce & Pisoni, 1998), interfere with learning (Swingley & Aslin, 2007). In the small lexicon of 14-month olds, known words are accessed somewhat easily from phonetic input and compete with novel or newly learned words. New words that sound similar to existing words will activate both a novel representation and these existing known words, and do not fare well in the resulting competition. Thus, 14-month olds learning words like "tog" will have difficulty because they retrieve "dog" instead (Swingley & Aslin, 2007). Similarly, when infants learn two similar words at once, the word-forms compete with one another for representation. As a result, each inhibits the other and learning fails, or alternatively, both representations get linked to the referent (since they are both momentarily active in parallel). As children develop, the lexicon expands, resulting in more "balanced" competition – the strength of competitive interactions coming from *dog* for example, may be balanced by competition from words like *doll, tall, dot*, and *bog* (c.f., Thiessen, 2007).

Though both theories explain existing behavioral data, they imply that speech perception is well developed in children at this age, and that top-down factors impede it (Werker & Curtin, 2005). However, it is possible that bottom-up speech perception factors, that is, perceptual abilities that are relevant for speech but not completely developed, may contribute to this failure.

Though discrimination tasks indicate that some category boundaries are established by 1 year (e.g., Werker & Tees, 1984), there is also abundant evidence that children refine their phoneme categories well into the school years (Nittrouer, 2002; Ohde & Haley, 1997; Slawinsky & Fitzgerald, 1998). Thus, it is possible that 14-month-olds' phonetic categories are only partially developed, and the existing categories, while sufficient to succeed at discrimination tasks, may provide a weak platform for word learning.

Rost and McMurray (2009) assessed this by examining the role of acoustic variability in learning phonologically similar words. We hypothesized that if speech categories were still developing, the small set of acoustic exemplars provided in most studies (Stager & Werker, 1997; Werker et al, 1998; 2002) might leave ambiguity about the structure of the phonetic category. Variability could provide more structure to the phonetic category, supporting word learning. Similar effects of variability on category learning have been observed in both visual categorization (Quinn, Eimas & Rosenkrantz, 1992; Oakes, Coppage & Dingle, 1998) and in the acquisition of phonetic categories in a second language (Lively, Logan & Pisoni, 1993), suggesting that this simple manipulation may be an important way to support categories that are not yet fully developed.

Fourteen-month olds were tested in the switch task (Werker et al., 1998) by habituating them to two novel objects paired with two novel, phonologically similar, words (/buk/and/puk/, both rhyme with “luke”[i]). Infants were then tested on a *same* trial, where the word-object pairing was consistent with habituation, and a *switch* trial, where the word-object pairing was opposite of what it had been in habituation. If infants internalized the word-object mapping, they should dishabituate on the *switch* trials. Experiment 1 replicated prior work: infants hearing a small set of exemplars failed to notice the switch. However, Experiment 2 employed multiple exemplars of the words spoken by 18 speakers; infants hearing variable exemplars correctly acquired the two phonologically similar words.

At face value, successful learning in the multi-talker condition is surprising. Multi-talker variation imposes a significant cost on speech processing in adults (Mullenix, Pisoni, & Martin, 1989), toddlers (Ryalls & Pisoni, 1997), and infants (Jusczyk, Pisoni, & Mullinex, 1992) and could be expected to add to the task demands here. In fact, from a purely processing standpoint, this may add significant demands.

However, specific types of variability may also play a role in forming appropriate phonetic categories. Under both prototype (Miller, 1997, 2001; Kuhl, 1991) and exemplar (Goldinger, 1998; Pierrehumbert, 2003) theories of speech perception, variability is essential to defining the limits of a category (e.g., what tokens are *not* a /b/). Developmentally, it is important for the learner to hear variable exemplars in order to delineate the acoustic space encompassed by a phonological category and words. Moreover, as numerous authors have pointed out (Swingley & Aslin, 2002; Yoshida et al., 2009), the switch task relies on infants’ abilities to both identify a word and identify that a given auditory stimulus is *not* an exemplar of a lexical category. If variability is essential to defining the edge of a category, a lack of variability could be particularly problematic in the switch task.

The multi-talker input used in Rost and McMurray (2009) contained multiple sources of variability, both within and between speakers. This included variation in prosodic patterning, fundamental frequency, vowel quality, and voice timbre. These factors do not distinguish /buk/from/puk/, nor do they serve as cues for voicing more broadly. However, these tokens also contained variation in Voice Onset Time (VOT: the continuous cue that distinguishes voicing, hence the two words to be learned) that is *constrastive* for the voicing feature distinguishing /buk/ and /puk/. A number of studies have examined the role of such variation in the formation of speech categories. Phonetic investigations of cues like VOT reveal statistical distributions that maintain the separability of /b/ and /p/, but have significant within-category variation (Allen & Miller, 1999; Lisker & Abramson, 1964). Moreover, Maye, Werker & Gerken (2002; see also Maye, Weiss & Aslin, 2008; Teinonen, Aslin, Alku & Csibra, 2008) have demonstrated that infants are sensitive to these distributions and may use them to learn speech categories. In these studies, infants were exposed to a set of words in which the VOT statistically distributed into one or two clusters, after which, infants’ patterns of discrimination mirrored the number of clusters in the input. Thus, variation in contrastive cues may play a role in category learning (see McMurray, Aslin, & Toscano 2009) by providing an estimate of the width of the category or its edge.

In fact, Rost and McMurray’s (2009) stimuli contained variability in VOT that mirrored the statistical distributions of English. Figure 1A shows the distribution of tokens for VOT found by Allen and Miller (1999) along with the distributions in the stimulus set of Rost and

[i] These novel words contained the high, back, rounded vowel /u/. In the dialect of the participants of these experiments (as well as most dialects of American English), this vowel differs in both height and tenseness from the one in the known word “book” (/*/). Phonetic spelling has been used throughout to minimize confusion: spelling it “buke” suggests palatalization that is not present (e.g. puke), spelling it “book” suggests a lower, unrounded vowel.

McMurray (2009). Given this correspondence, it is possible that infants were simply engaging statistical learning mechanism of the sort identified by Maye et al. (2002).

Experiments 1 and 2 tested the hypothesis that variability along the contrastive dimension of voicing helps infants define the phonological categories for the words, while simultaneously eliminating non-contrastive variation that might be expected to impede processing. If true, it might suggest that further development of the internal statistical structure of VOT distributions is necessary for phonological categories to be engaged in this case.

Experiment 1

We used the same words as Rost and McMurray (2009): /buk/ and /puk/. These differ in voicing, for which VOT is the dominant cue. In the present study, the effects of variability in VOT alone were investigated by training and testing infants using auditory stimuli from a single speaker, but with a VOT distribution (Figure 1C) that mirrored distributions in the child’s language as well as the distribution found in the original Rost and McMurray study. This is an important contrast with the work of Maye and colleagues (2002, 2008), in that our continua spanned a dimension that infants had significant familiarity with, and used asymetrical (though more natural) distributions. Given the purpose of augmenting their natural categories (to explain our prior results), this seemed a better test.

If variation in VOT is sufficient to drive learning, then we should observe good word learning using a set of exemplars with this distribution of VOT, and no variation in any of the additional cues present in multi-talker input (e.g. pitch, vowel quality, prosody or timbre).

Methods

Participants—Infants between 13 and 15 months old were recruited from county birth records. Infants were eligible if they were monolingual English learning, with no history of developmental disorder or recurrent ear infection. 26 infants participated; data from 10 were excluded due to their failure to habituate (5), experimenter error (2), fussiness (2), and ear infection (1). 16 infants (9 boys; Mean age: 14 mo, 4 days; Range: 13,5-14,22) were included in the final analysis.

Stimuli—A female native speaker of the local dialect produced a series of /buk/ and /puk/ tokens in an infant directed register. In order to create a continuum with sufficient variation we included prevoicing (so that /b/ could be more variable while still being distinct from / p/). PRAAT (Boersma, 2001) was used for all stimulus manipulation.

One /buk/ token was chosen by five adults as being the “best” exemplar, and it was modified to have a VOT of close to 0 ms by cutting the prevoicing. One /puk/ token was chosen as having the most natural aspiration which was longer than 100 ms. From these we constructed a 29 point VOT continuum ranging from −40 to +100 in steps of approximately 5 ms (limited by the availability of splice-points) using the following procedure.

To construct positive voice onset times, we first cut the desired amount of aspiration from the onset of the /puk/ token, including the release burst as well as the following aspiration. We then cut the release burst from the /buk/ and added in the aspiration. The prevoiced portion of the continuum (from −40 to −5 ms) was constructed by adding prevoicing from the original recording back to the /buk/ in 5 ms increments. Each segment started at onset of the prevoiced period so as to preserve the natural amplitude envelope. This yielded a −40 to 100 ms continuum in which the coda (/uk/) was acoustically identical across exemplars while voicing (either prevoicing or aspiration) changed from −40 ms to 100 ms VOT. (See Figure 3)

The original waveform was 218 ms from the onset of the /b/ to the vowel closure. This was increased as a function of VOTs so that /p/s were up to 100 ms longer than /b/s, consistent with the approach to VOT/syllable length advocated by Kessinger and Blumstein (1998). The waveforms were surrounded by silence to increase the total length of the file to 2 s (so that when 7 files were spliced together, the total trial length would be 14 s). For all files, the release burst was timed to occur at exactly 500 ms into the file. This was done so that a sequence of files (within a trial) would be perceived as having a consistent rhythm.

Ten adult listeners piloted this continuum using a forced-choice (b/p) task. Results of the pilot indicated that VOTs of less than 15 ms were reliably perceived as /buk/, and VOTs greater than 20 ms were perceived as /puk/. (Both those tokens were ambiguous.) We did not observe any differences in overall rate of responding in the unambiguous regions (the good /buk/s and good /puk/s were both identified at 100%). In constructing the distribution of exemplars used for training infants, tokens within 10 ms of this boundary received a frequency of 0 and were not heard.

The /buk/ category extended from −40 ms of prevoicing to 5 ms VOT. The /puk/ category ranged from 35 to 100 ms. Similar to Maye et al. (2002), we assigned a frequency to each token, so that the most frequent /buk/ was at 0 ms, and /puk/ had a normal distribution with a mean of 70 (see Figure 1C). The particular values were chosen to simultaneously resemble the distribution of tokens in natural language while preserving the structure of Rost and McMurray (2009). Importantly, the difference between the modes for /buk/ and /puk/ was 70 ms, the same as that of previous work. Prior to the experiment, a custom Matlab script selected tokens for each phoneme at random, weighted by these probabilities. It then combined stimuli into a series of files containing 7 exemplars to be used during the experiment. Token selection was done separately for each trial (both training and test), so each trial had a unique set of exemplars.

The habituation and test trials were then prepared as in Rost and McMurray (2009), with the same photographic visual stimuli. Each trial lasted for 14 seconds during which one picture was shown and 7 different exemplars of the word were spoken at 2 second intervals.

Procedure—Infants were seated on a parent's lap throughout the procedure, and parents listened to music over headphones so they were unaware of the auditory stimulus. Stimuli were presented on a 50" plasma monitor and stereo speakers using HABIT software (Cohen, Atkinson, & Chaput, 2004). Looking time was coded online by an experimenter blinded to both visual and audio presentation, and inter-experimenter reliability for looking-time was over 90%.

The switch task was used (see Werker et al, 1998 for a complete description of the task). Infants were habituated to two objects paired with /buk/ and /puk/ in trials of a fixed length of 14 s. When looking time reached 50% of the initial value over a 4-trial moving window, the procedure automatically transitioned from the habituation phase to the test. Infants were then tested on one of the objects in a *same* trial (the word-object pairing was the same as in habituation) and a *switch* trial (the pairing was switched). As is typical practice, both trials used the same visual stimulus, but the auditory stimulus varied to either match or mismatch the object. After both experimental trials, infants were tested on a *control* trial, where a word from habituation was paired with a novel object to ensure that the procedure was successful.

Habituation trials were presented in pseudorandom order, with word-object pairing and test words counter-balanced across subjects. The *same* and *switch* trials were counter-balanced in the first two test positions, and the *novel* trial was always presented third.

Results and Discussion

Data were analyzed using a mixed design ANOVA, with test condition (*same*, *switch*, and *control*) as the primary within-subject variable. We also included test order (same-first or switch-first) and the word used for test (whether the same trial featured /buk/ or /puk/) as between-subjects factors. While these two factors were counterbalanced between subjects, it was important to demonstrate that they did not interact with our primary effect. We were particularly interested in the word used at test, as it was possible that infants' responses could have been affected by a preference for one of the words. This was important as one of our stimulus items, /buk/, is phonologically similar to "book", a word known to 90% of children this age (Dale & Fenson, 1996). Lexical familiarity could have created difficulty mapping /buk/ because of lexical competition (Swingley & Aslin, 2007) or conversely could allow children to map the word more easily due to lexical support (Theissen, 2007).

The analysis found a main effect of condition (*same, switch*, or *control*: $F(2, 24)=30.4$, $p<.001$). Planned comparisons (Figure 2) showed that the condition effect was driven entirely by looks to the *control* trial. The *control* trial was significantly different from *same* and *switch* trials ($F(1,12)=57.1$, $p<.001$), but there was no difference in looking time between *same* ($M=5.18$ sec., $SD=2.45$) and *switch* ($M=5.16$, $SD=3.45$) trials ($F<1$).

In addition, there was no effect of word used at switch ($F<1$) or test order ($F(2, 24) = 1.08$, p=.36), and no two- or three-way interaction (trial × word: $F(2,11)=1.1$, $p=.36$, trial × test order, $F<1$, trial × word × test order, $F(2,11)=2.1$, $p=.17$), indicating that children responded without preference for either word, and order of test trials did not affect responses.

The null result was unexpected, as work in infant speech perception has shown robustly that infants use variability in contrastive acoustic dimensions to learn phonemic contrasts (Maye et al., 2002, 2008), phonetic analyses support such structure in the input (Kuhl et al, 2007), and a number of computational model have shown that such processes can account for a range of behavioral data (McMurray et al., 2009; Vallabha, McClelland, Pons, Werker, and Amano, 2007; Toscano & McMurray, in press).

One possible explanation for this failure could be the method used to construct the stimuli. This method of continuum construction has the disadvantage of producing voiceless tokens without the F0 pitch-onset rise in naturally-produced speech. Younger infants in previous experiments have responded to voice distinctions in continua constructed this way (McMurray and Aslin, 2005), and data indicate that children do not perceive F0 as a cue before 4 years of age (Bernstein, 1983), yet it remains possible that the infants in Experiment 1 might have responded poorly to the /puk/ stimuli because of the unnatural properties of the continuum.

In fact, beyond F0, many cues to voicing are simultaneously present in natural speech (e.g. pitch, burst amplitude, vowel length, first formant frequency: Burton, Baum, & Blumstein, 1989; Burton & Blumstein 1995, Ohde & Haley, 1997). It is possible that variability in additional acoustic cues may be needed to establish a robust voicing contrast, cues that were likely to vary in Rost and McMurray (2009) within and across speakers. Experiment 2 therefore tested infants' use of variability in these additional contrastive cues by using a continuum that co-varied in VOT, pitch and burst amplitude.

Experiment 2

Methods

Participants—Recruitment and exclusion criteria were the same as in Experiment 1. 22 infants participated and data from six were excluded for failing to habituate (2), having ear infections (2), fussiness (1), and experimenter error (1). Analyses were run on data from the 16 remaining infants (10 boys; Mean age 14;13, Range: 13;10-15;0).

Stimuli—In Experiment 2 we modified the continuum from Experiment 1 to include additional covariation between VOT and two secondary voicing cues (burst amplitude and F0). Figure 3 details this process. The amplitude of the burst and aspiration was manipulated by excising the burst (including the entire VOT) from the voiced tokens and multiplying the wave form. The 5 ms token had a burst and aspiration whose amplitude was 5% of its maximum value, and the 100 ms token had 100% of its burst and aspiration (for intermediate VOTs the amplitude was varied in steps of 5%). The modified bursts were then spliced back onto the vocalic portion. Next, the initial F0 of this series was manipulated using PSOLA resynthesis. Pitch was shifted by an amount proportional to the VOT, started at the onset of the stimulus, remaining flat over the first 40 ms, and gradually reduced to the natural pitch by 100 ms. For VOTs of −40 ms, we subtracted 30 Hz from the onset pitch. For VOTs of 100 ms, we added 30 Hz (and interpolated for intermediate values[ii]). The 60 Hz difference in pitch change was chosen to mirror that reported in Bernstein (1983).

The resulting continuum simultaneously varied in VOT (from −40 to +100), in F0 at onset (from −30 Hz to +30 Hz over the unmodified pitch), and in amplitude of the burst (from 0 to 100% of the maximum value). Words lengths measured from consonant onset to vocalic closure varied systematically from 218 ms (0 ms VOT /buk/) to 258 ms (40 ms prevoicing / buk/) to 318 ms (100 ms VOT /puk/). Tokens were again validated by adult listeners in a two-alternative forced-choice task: the boundary was between 15 and 20 ms VOT, with tokens less than 5 ms VOT reliably perceived as /buk/ and greater than 30 ms VOT reliably perceived as /puk/. As these values were consistent with Experiment 1, the tokens were assigned to the same statistical distribution as in Experiment 1, and were chosen for habituation and test identically.

Procedure—Experimental set-up and procedures were identical to Experiment 1.

Results and Discussion

Data were analyzed similarly to Experiment 1, and results are shown in Figure 2. A repeated-measures ANOVA found a main effect of test condition (*same* vs. *switch* vs. *control*, *F*(2, 24)=30.6, *p*<.001). Planned comparisons again revealed that the effect was driven by responses to the *control* trial. Children looked at the *control* trial (*M*=10.1 s, *SD*=2.5) significantly longer then the *same* and *switch* trials (*F*(1,12)=58.7, *p*<.001), but did not look differently at the *same* (*M*=5.03 s, *SD*=2.37) and *switch* (M=5.55 s, *SD*=3.28) trials (*F*(1,12)=.56, *p*=.47).

[ii] We discovered two minor errors in stimulus construction after the experiment was finished. The −35 ms (prevoiced) /buk/ had an initial F0 that was about 25 Hz lower than it should have had. .This was not likely to influence the outcome of Experiment 2, as its frequency of exposure was 3/1000, and it was off in the direction that correlates with voicing (e.g., lower F0, more voiced). Additionally, the F0s for the 0 ms /buk/ and −5 ms prevoiced /buk/ were reversed, so 0 ms /buk/ was 2 Hz lower than −5 ms prevoiced /buk/, a relatively small difference Despite these errors, it is important to note that overall there was a robust correlation between VOT and F0 (R=.97), thus these stimuli nonetheless met the experimental requirements of systematically lowered F0 for /b/ tokens and higher for /p/ tokens.

There was no effect of test order (*F*(1, 12)=1.5, *p*=.24) or *switch* test word (/buk/ or /puk/, *F*<1) and no two- or three-way interaction (all *F*<1), indicating again that neither trial-order effects nor preference for either word affected responses.

As in Experiment 1, infants in Experiment 2 failed to map words well enough to react to the change in word-object pairing at test. It seems that distributional statistics of constrastive cues in the exemplars cannot account for the learning observed by Rost and McMurray (2009), even though those cues are fundamental to the voicing category. So, how did the infants in Rost and McMurray manage to learn the correct word-object mappings?

A set of multi-talker tokens naturally contains both contrastive and non-constrastive variability. Non-contrastive variability encompasses speaker-specific variables (i.e., pitch, vowel quality, timbre, sociolinguistic variation) and production-specific variables (i.e., prosody) that are not associated with lexical contrast (e.g., there are no English words that differ only by pitch). Since these do not cue phonemic or lexical contrasts, much work in speech perception has been devoted to explaining how listeners are able to overcome such variability to arrive at the underlying meaning (e.g., Perkell & Klatt, 1986).

Alternatively, it is possible that the auditory system would need to retain, rather than normalize, multiple forms of acoustic information to arrive at the correct categories (Goldinger, 1998; Klatt, 1979; Pisoni, 1997; Pierrehumbert, 2003). Prior work on this has focused on whether listeners use such detail during online perception (Creel, Aslin & Tanenhaus, 2008; Goldinger, 1998; Johnson, 1990; Ryalls & Pisoni, 1997). Importantly, it has been shown that infants might map both indexical and phonetic information of words in early word learning (Houston & Jusczyk, 2000). This suggests that irrelevant cues, such as indexical information, may help in the *acquisition* of speech contrasts.

Indeed there is evidence that variability along non-phonemic dimensions may help identify the underlying invariant structure of speech. Singh (2008) has shown that variation in the affective quality of speech improves word segmentation in infancy. Hollich, Jusczyk, and Brent (2002) report that word segmentation abilities are improved by multiple-talker familiarization in older infants. However, both studies looked at broad segmentation abilities, not at the perception of a single phonetic feature (e.g., voicing) in a highly ambiguous context. This was explicitly tested in Experiment 3.

The exemplar set used in Rost and McMurray (2009) was highly variable in non-contrastive aspects of the signal (such as vowel quality or pitch), but the range of variability within these dimensions did not differ between /buk/ and /puk/. If infants use highly variable information to isolate relatively invariant elements of the signal, they should succeed at the switch task when exemplars contain lots of variability, but minimal within-category variability in contrastive cues.

Experiment 3

Methods

Participants—Recruitment and exclusion criteria were the same as in Experiment 1. Twenty-three infants participated, and data from seven were excluded from analysis for experimenter error (4), fussiness (2), and failure to habituate (1). 16 infants (9 boys; Mean age 14;08, Range 13;05-15;01) were included in the experimental analysis.

Stimuli—Stimuli consisted of the original set of 54 exemplars recorded from 18 speakers from Rost and McMurray (2009). These were modified to maintain variation in all of the non-criterial (indexical and prosodic) cues but eliminate within-category variation in VOT.

To do this, all of the /buk/ tokens (56 total) were modified so that they had VOTs of approximately 2 ms (*M*=2 ms, *SD*=1) by clipping voice-onset time out of the sound files (since all had natural VOTs of great than 2 ms). Likewise, the /puk/ tokens were modified to have VOTs of approximately 70 ms (*M*=69, *SD*=2). These values are as identical to the means from Experiments 1 and 2 as was technically possible, and the difference between the means again mimics both exemplar sets in Rost and McMurray (2009). For the half of the tokens naturally produced with VOTs shorter than 70 ms, aspiration was copied from the center of the aspirated period and spliced again into the sound file to increase the total VOT. For tokens with VOTs longer than 70, aspiration was cut from the center of the aspirated period.

Stimuli in the /buk/ category varied in length from 217 ms to 705 ms, with a mean length of 425 ms (SD=11). Stimuli in the /puk/ category varied in length from 339 ms to 765 ms, with a mean of 487 (SD=.11). The length of the vocalic portion (measured from voicing onset to closure) between the two categories did not differ (/buk/ M= 237, SD=7; /puk/ M=220, SD=.8, *t*=1.09, *p*=.27), indicating that the mean difference of 62 ms between the /buk/ and /puk/ word sets was caused by the experimentally manipulated VOT difference between them.

The order of these items within and across trials was pseudo-randomized using a Matlab script so that infants heard 36 different exemplars of each word in random sets of 7 per trial during the habituation phase and 7 (previously unheard) exemplars of each word in random order during the test. These presentations were again at 2-second intervals for fixed habituation trials of 14 s.

Procedure—Experimental set-up and procedures were identical to Experiment 1, with the exception that all tokens were equally probable (for a given word).

Results and Discussion

Data were collected and analyzed in the same manner as in Experiment 1. Figure 2 displays the results. A repeated-measures ANOVA revealed a main effect of test condition (*F*(2, 24)=22.7, *p*<.001). Planned comparisons revealed that this effect was driven by the fact that infants looked to the *switch* trial (*M*=7.16 s, *SD*=4.06) significantly longer than the *same* trial (*M*=4.19 s, *SD*=1.98; *F*(1, 12)=8.1, *p*=.015). Unlike Experiments 1 and 2, they dishabituated to the *switch*: that is, they represented both words well enough to notice the misnaming. Similarly to the prior experiments, infants also looked to the *control* trial (*M*=9.63 s, *SD*=3.17) longer than the *same* and *switch* trials (*F*(1, 14)=57.7, *p*<.001).

Importantly, we found no effect of test order (F<1) or switch test word (/buk/ or /puk/, *F*<1), and no two- or three-way interactions (all *F*<1). Dishabituation to the *switch* trials cannot be attributed to test order or word preference.

One concern was whether the highly salient speaker variability caused the infants in Experiment 3 to take longer to habituate than those in the prior experiments. A longer habituation would provide more experience with the words, resulting in better learning. This was not the case: infants took an average of 15.6 (*SD*=5.07) trials to reach habituation criterion in Experiment 3 while they averaged 16.6 (*SD*=6.37) trials in Experiment 1 and 17.6 (*SD*=6.02) in Experiment 2. Note that since trials were not terminated due to lack of attention, this means that Infants in Experiment 3 averaged 15.6×7=109.2 tokens of the words compared to 116.2 in Experiment 1 and 123.2 in Experiment 2. These differences were not significant (*F*<1), and if anything the infants in Experiments 1 and 2 got more exposure. Consequently, the learning observed here cannot be attributed to the number of words heard by the infants. Instead, it must be that the acoustic variability along non-criterial dimensions affected infants' learning.

A second concern was that we operationally defined the contrastive cues for voicing as the absolute VOT, rather than the relative duration of the aspiration and voiced period. As a timing cue, VOT varies as a function of the speaking rate, which can be approximated as the duration of the vowel. If infants perceive voicing using VOT *relative to the vowel length*, then there may be some contrastive variability embedded in this set. Any effect of speaking rate (vowel length) will be necessarily small: a 100 ms difference in vowel can only shift the VOT boundary by 5-10 ms in synthetic speech (Summerfield, 1981; McMurray, Clayards, Tanenhaus & Aslin, 2008), and barely at all in natural speech (Utman, 1998; Toscano & McMurray, submitted). Moreover, McMurray et al. (2008) demonstrate that listeners are capable of using VOT before they have heard the vowel length, suggesting the two function as independent cues to voicing, not as a single relative cue (see Toscano & McMurray, in press). Nonetheless it is important to determine whether, even when VOT is treated as a relative cue, we reduced the variability in contrastive cues from Rost and McMurray (2009).

One way to operationalize this relative measure is the ratio of VOT to Vowel Length. Analysis of the relationship between the original items reported in Rost & McMurray (2009) and the modified versions of those stimuli used in the experiment reported here indicated that our stimulus construction minimized, rather than contributed to, variability in this measure. For reference purposes, this measure lead to a mean ratio of .012 for /b/ in the modified set (.063 in the original), and .45 for /p/ (.51 original). Computing the standard deviations of this ratio measure of voicing showed a substantial decrement between the experiments for both /buk/ ($SD_{original}$ =.027, $SD_{modified}$=.0085) and /puk/ ($SD_{original}$=.227; $SD_{modified}$:=.18)[iii].

We can also operationalize this relative measure by using linear regression to partial out the effect of vowel length from VOT. An analysis of these residuals after linear regression also showed that the present stimuli have lower variance by an order of magnitude. Again for reference, the means of these residuals were −.044 (−.034 for the original stimuli) for /buk/ and .023 (.034 for the original stimulil) for /puk). Importantly the variance in both was much lower in the present experiment (/buk/: $SD_{original}$=.0046, $SD_{modified}$=.0023; /puk/: $SD_{original}$=.026, $SD_{modified}$=.0026).

Thus, by both relative measures, the variance in the information available for voicing was minimized dramatically. Given the relatively slight contribution of this cue to perception in adults, it is clear that we have significantly reduced (if not altogether eliminated) variation in contrastive information in Experiment 3.

A final concern was that the coda (/uk/) portion of the two words was not physically identical between /buk/ and /puk/ tokens within a speaker, as it was in Experiments 1 & 2. Coda information could have provided an additional source of constrastive information about voicing. It seems unlikely that such information would be sufficient to distinguish the words for two reasons: first, if coda information was necessary to distinguish the word-initial voicing, prior experiments using natural recordings that preserved coda information (Rost & McMurray, 2009; Pater, Stager, & Werker, 2004) would have provided sufficient information for categorization in this task. Second, the effect of voicing on the vowel is small: most of the established cues to word initial voicing are found at the release or the aspiration/voicing juncture (Allen & Miller, 1999).

Nonetheless, if there was information correlated with voicing, then variability in these cues could have helped the infants. Experiment 1 and 2 rule out contrastive variability alone

[iii] Inferential statistics to compare variances (e.g. the F-test) are not possible in this case, since the two values (the original and modified stimuli) are not independent of each other (both have the same vowel length in the denominator of the ratio).

(particularly since the contrastive cues varied there were much more robust cues to voicing than anything in the coda), but it is possible that these cues, *combined* with the non-contrastive variability we manipulated were driving the effect. To determine if the coda portions of the words contained any information that could contribute to a voicing decision, we measured a number of cues to voicing: the length of the syllable (measured from the release to the onset of closure), the pitch (F0), and the first and second formant frequencies. Measurements of F0, F1 and F2 were conducted twice, once during the first pitch pulse after the onset of voicing and once at the midpoint of the vowel (see Table 1).

All of the measurements showed substantial variability. For example, at voicing onset, F0 had an SD of 84 Hz for /buk/ at onset and 97 Hz for /puk/. Similarly, F2 varied by well over 150 Hz at both points. This is perhaps to be expected given the variability in speakers (especially the variability in gender) and register across the Experiment 3 stimulus set and it validates our assumption that these stimuli had substantial variation. However, none of these measures showed significant differences as a function of the word. In fact, F1 and F2 at onset actually showed effects that were the slight inverse of standard phonetic results (i.e., Allen & Miller, 1999): both were lower for /puk/ than /buk/.

F0 was the only cue near significance for distinguishing between /buk/ and /puk/. Phonetic data suggests that F0 should be lower for /b/ than /p/, and at voicing onset, /buk/'s F0 was indeed 27 ms lower than /puk/'s; this was not even marginally significant (*t*(106)=1.59, *p*=.11). However, it seems unlikely that F0 could serve even to augment the non-contrastive variability in Experiment 3: 28 /buk/s had F0 values less than the median, compared to 26 /puk/s. Though there was an almost marginal effect in the right direction, there weren't enough tokens showing this relationship to make F0 a worthwhile cue. Moreover, Experiment 2 ruled out that F0 in the absence of non-contrastive variability drives this effect. As a result, the cue that came the closest to distinguishing the words does not appear to have much utility as a constrastive cue in this particular set of stimuli.

General Discussion

These experiments investigated the role of contrastive and non-contrastive phonetic variability in infants' word learning in the switch-task procedure. Experiments 1 and 2 examined whether variability in a contrastive cue was necessary for minimal pair learning in the switch task. Our initial hypothesis was that the switch task requires children to determine that a given exemplar is *not* a member of the /buk/ (or /puk/) category, and as a result, some estimate of the extent of a category along the contrastive dimension may be needed to make this determination. However, this was not the case: across both experiments there was no evidence for learning, even when three cues to voicing varied simultaneously. Indirectly, this provides evidence that the kind of statistical learning first reported by Maye and her colleagues (Maye et al., 2002, 2008; see also Kuhl, et al., 2007; McMurray et al, 2009; Vallabha et al., 2007) cannot account for learning in Rost and McMurray (2009): variability along the contrastive dimension of voicing alone is not sufficient to support learning. We do not argue that infants ignore variability along dimensions such as VOT. Indeed, it is likely to be important in establishing the location of categories within a dimension. However, it seems this is not the information that they must glean to succeed here by this more advanced age. This suggests that whatever perceptual development must still occur to support performance on this task is not simply locating categories within a dimension, rather some other component must be developing.

In contrast, Experiment 3 suggests that variability along non-contrastive acoustic dimensions supports minimal-pair learning in the switch task, even when contrastive variability is minimized. Before reaching this conclusion, however, it is important to assess several alternatives. One possible explanation for this is that the stimuli presented in Experiment 3 are more natural than those in Experiments 1 and 2. It is not clear that this is the case: both sets of stimuli were created by manipulating natural speech using similar techniques (cross-splicing), and adult listeners did not report that either sounded unnatural. Nor is it clear that manipulated speech in this case poses a problem: previous switch-task studies (Stager & Werker, 1997; Werker and Fennell, 2006 for a review; see Pater, Stager, & Werker, 2004 for an example using voicing contrasts) all used un-manipulated natural speech, and 14-month-olds consistently failed to learn minimal-pair words.

A second possibility is that the highly salient variation between speakers was more engaging and thus resulted in better learning. But our analysis of infant habituation times renders unlikely the possibility that infants were more engaged: they had slightly fewer trials to habituation in Experiment 3 than Experiment 1 and 2.

A third possibility is that the more naturalistic variation in Experiment 3 also contained secondary cues to voicing. Yet measurements of our stimuli rule out the possibility that the items retained perceptible variability of cues related to voicing. Moreover, if VOT was treated as a relative cue (which is unlikely given the adult work), Experiment 3 substantially minimized variation in this contrastive dimension, and infants still learned the words.

Finally, as we will discuss in more detail, the task-demands (Yoshida et al, 2009) and lexical competition (Swingley & Aslin, 2007) frameworks offered as prior explanations for children’s failures in this task also do not predict the findings reported in Experiment 3.

Because neither naturalness, saliency, contrastive acoustic cues, nor task demands explanations adequately explain the results of Experiment 3, we are left with irrelevant speaker information as the driving force of this effect. It must therefore be that variability along dimensions that do not typically distinguish words, in fact helps 14-month-olds to acquire lexically contrastive phonetic representations.

One simple account for this is that infants might not be fully committed to which cues are relevant for voicing by this age. If this were the case, then, variability along indexical dimensions helps infants learn that they are not relevant; conversely the relative invariance of VOT points to its utility in contrasting words. Multi-talker variability helps the infants with *dimensional weighting* (Toscano & McMurray, in press), the assignment of weight or importance to perceptual dimensions.

Ongoing computational work (Apfelbaum & McMurray, submitted) shows how simple associative learning mechanisms can give rise to this. This model suggests that without speaker variability infants erroneously associate indexical and pitch cues with both words—when the same speaker is heard at test, then, both words receive partial support making it difficult to rule one out. The constant indexical cues, thus, interfere with establishing contrast. Variability in speaker prevents this by spreading association across many possible speakers.

By this account, multi-talker variability might be only one of many types of variability that could yield this same effect. Variability in non-contrastive cues (as is prevalent in infant directed speech) has been thought to be helpful for word and language learning in young infants, although relatively few reports indicate that this is indeed supportive of learning, as opposed to merely preferred by infants. Singh (2008) is a notable exception: she familiarized 7.5-month-olds to words using both high- and low-affect productions, and found that infants only segmented the words in the presence of high affective variability, that is, high prosodic variability. Similarly, infants segment words from IDS-inflected speech but not ADS-inflected speech in novel speech strings containing statistical cues to word boundaries

(Thiessen, Hill, and Saffran, 2005). This raises the possibility that highly variable prosody alone may be sufficient to support word learning in this task, as well.

This suggests that the established view that infants use the statistical structure of contrastive cues to learn phonological categories (Kuhl et al, 2007; Maye et al., 2002; 2008; McMurray et al., 2009; Vallabha, et al, 2007) may be incomplete. We suggest that by 14 months, even though infants appear to discriminate tokens *within a dimension*, they might not be fully committed to VOT as a relevant dimension for distinguishing words that vary in voicing, and must determine which dimensions are relevant by examining relative variability.

Of course, the behavioral experiments reported here and in Rost and McMurray (2009) do not offer definitive proof of our dimensional weighting account. Further empirical and computational work will be necessary to fully establish this account. However, as we argue in the subsequent sections, the dimensional weighting account is consistent with both the task-demands framework for explaining the switch task and with broader exemplar models of speech (e.g., Pierrehumbert, 2003). Moreover, the use of relative variability as a mechanism of weighting crops up in numerous domains of learning and may represent a general principle of learning. Thus, when the present behavioral data is coupled with the seeming universality of such mechanisms and strong computational models (Toscano & McMurray, in press; Apfelbaum & McMurray, submitted), this seems to be quite a reasonable explanation.

Implications for task demands and phonological development

In the task-demands framework (Werker & Fennell, 2006; Werker & Curtin, 2005), attentional demands on the infant create an apparent U-shaped developmental trend where infants' speech perception abilities are intact and preserved, but infants are unable to access them in a difficult task, as they struggle to balance perceptual, phonological, and lexical representations.

There is no doubt that the switch task is particularly hard. Infants fail at the switch-task test but succeed at the easier looking-preference test (Yoshida et al, 2009). Nazzi's (2005) sorting-by-name task may yet be more difficult. However, this may not be simply an issue of general capacity limits, but the unique way in which word-object mappings must be used in the switch task may also create task-specific difficulty (e.g. Swingley & Aslin, 2002). However, there are two interpretations of infants' difficulties with this task: it could indicate that phoneme perception is robust at this age, but that a difficult task masks children's ability to deploy these skills (e.g., Werker & Fennell, 2006). Alternatively, our work suggests that this difficult task reveals specific difficulties in speech perception.

In an easy task, such as a checkerboard dishabituation or a looking-preference task, the nature of the task only requires infants to discriminate pairs of speech sounds – it is not necessary to ignore any dimensions as a detectable difference on any of them should be sufficient to drive discrimination. In Maye et al's (2002, 2008) work, the relevant statistics within a cue were sufficient to alter discrimination. However, the switch task is closer to a categorization task, in which many sources of information (irrelevant or relevant) may be associated with the response. Thus, it may reveal a second component of perceptual development, dimensional weighting. Dimensional weighting is a key feature of PRIMIR (Werker & Curtin, 2005), but it was not explicitly tied to switch-task failure due to lack of empirical evidence. The results of our experiments suggest this explicit relationship: for 14-month olds in the switch task, the statistics of contrastive cues are less helpful (because they are relevant to a problem that is already solved) than the statistics of non-contrastive cues (which are relevant to the problem of weighting).

Thus, as numerous researchers have pointed out, the nature of the task is of fundamental importance to understanding results like these (Yoshida et al, 2009; Swingley & Aslin, 2002; Werker & Fennel, 2006). However, the overall difficulty of task perhaps does not fully describe why. Rather, what is important is the way that the task shapes how particular (and perhaps non-obvious) sources of information contribute to learning, the particular mappings that must be employed at test, and the kind of information used in those mappings (see Yoshida et al, 2009 for a similar discussion). Our interpretation of these results is that it is not that a difficult switch task masks intact phoneme perception, but rather that this difficult task highlights an aspect of speech perception is not yet well developed at this age. We may be left with the original conclusion of Stager & Werker (1997), that speech perception may not be developed sufficiently in 14-month-olds to fully support word learning.

Importantly, the ability of variation to shape dimensional learning is likely to break down differently depending on the acoustic/phonetic properties in question. This could explain the differences we see between perception of consonant, vowels, fricatives, and liquids (Havy & Nazi, 2009, Nazzi, 2005; but see Mani & Plunkett, 2007, 2008; Nazzi & New, 2007) Ultimately, differences in performance across phonetic contrasts may derive less from their phonological status (e.g., consonant vs. vowel) and more from the statistical structure of the cues to these contrasts, particularly when we look across multiple relevant and irrelevant dimensions. For example, cues like VOT do vary between speakers (Allen, Miller & DeSteno, 2003), but they are largely discriminable without taking this into account, therefore high variance in speaker cues will quickly reveal the more invariant contrastive VOT cues. However, for vowels, and to a lesser extent, fricatives and liquids, contrastive and non-contrastive acoustic dimensions overlap substantially (vowels: Hillenbrand, Getty, Clark & Wheeler, 1995; fricatives: Jongman, Wayland & Wong, 2000). These contrastive dimensions, such as formant structure, F0, and length, are also cues that vary considerably by speaker. In order to use speaker variation to detect such differences, infants may need more sophisticated ways of dealing with this variability or may simply need to learn more about those things that contribute to variance (Cole, Linebaugh, Munson & McMurray, in press) before vowel can be a cue to word identity.

As a result, failures in the switch task at 14 months do not represent a reversal of development, a U-shaped curve, or a discontinuity. We suggest rather that speech perception never was fully developed at 12 months, as is evidenced by studies of older children (Nittrouer, 2002; Ohde & Haley, 1997; Slawinsky & Fitzgerald, 1998). Reliance largely on discrimination measures resulted in a failure to consider other factors (like dimensional weighting) that are revealed by this task.

Mechanisms of Learning

This study hints that dimensional weighting is sensitive to the relative variation along different dimensions. This may in fact represent a general principle of learning. For example, the role of irrelevant variation suggested by this work parallels mechanisms proposed by Gómez (2002) for statistically-determined grammatical dependency structures. Adults and infants learned a novel grammar with non-adjacent dependency structure. When intervening elements in the dependency were long and variable, both adults and infants detected the non-adjacent dependencies. When intervening elements did not vary, participants were unable to learn the grammatical dependencies. Consequently, learning of grammatical dependencies in Gómez's experiment requires high variability in those elements that are not criterial for determining the grammar.

Yu & Smith's (2007) work on learning word-object mappings via cross-situational statistics illustrates the same point. In this study, subjects learned a small set of word-object mappings solely by noticing the statistical relationship between the sound and the object: whenever a given word was heard, the referent was consistently present. Importantly, *competing* objects were variable (with respect to the auditory word form). When the competing words were less variable, (i.e., there were fewer words each competing more systematically with the referent) subjects struggled much more to learn the word-object pairings.

The variability of irrelevant rules, associations, or dimensions may be fundamental to learning. This in turn hearkens back to much older work on *cue adaptation* or *cue neutrality* (Bourne & Restle, 1959; Bush & Mosteller, 1951; Restle, 1955), from the learning theoretic tradition. In these studies, animals or adult humans learned two-alternative categorization amongst stimuli that varied in multiple dimensions (some informative, some not). Crucially, subjects did not know in advance what dimensions to attend to and had to determine this from the relative amount of variability. Thus, an analysis of the relative variability in the input (or its utility in predicting the word/category) may be a core mechanism of learning.

More broadly, one of the critiques commonly leveled at (and by) the statistical learning community is its necessity to know *a priori* what units to compute statistics over (Marcus & Berent, 2003; Newport & Aslin, 2004; Remez, 2005; Saffran, 2003; but see Spencer, Blumberg, McMurray, Robinson, Samuelson & Tomblin, 2009). This work suggests a response to that critique: the system might compute statistics over multiple dimensions simultaneously to "discover" the right ones (using simple estimates of variability or something more complex). The system thereby forms knowledge of the statistical structure of the dimension.

Implications for Models of Speech Perception

This description of dimensional weighting also dovetails with work showing that speech perception in both adults and children is improved in known voices (Creel, Aslin, & Tanenhaus, 2008; Nygaard, Sommers & Pisoni, 1994; see also Goldinger, 1998 for a review). Because each speaker uses production cues differently and even has his/her own habitual VOT (Allen, Miller, & DeSteno 2003), listeners must learn to be sensitive to talker-specific intra-category differences (Allen & Miller, 2004). In light of our data, such effects could be interpreted as the remnants of dimensions that are not fully down-weighted. Speaker-specific effects have been taken to support exemplar models of speech (e.g., Goldinger, 1998; Pierrehumbert, 2003) in which contrastive- and non-contrastive information are stored together as part of the word-form. Our results suggest that such models might need to consider the ways that multiple dimensions are encoded and weighted, and how this changes over development.

Perhaps more importantly, a classic issue in speech perception has been the problem of invariance – how can listeners perceive the same word from highly variable acoustic streams? Classic theories have parsed "signal" (that is, the acoustic information we have labeled as being criterial) from "noise" and have attempted to explain category selection on only a few dimensions. In contrast, this work suggests that at least developmentally, the "noise" may be essential to acquiring the signal.

References

Allen SJ, Miller JL. Effects of syllable-initial voicing and speaking rate on the temporal charactaristics of monosyllabic words. Journal of the Acoustical Society of America. 1999; 106:2031–2039. [PubMed: 10530026]

Allen SJ, Miller JL. Listener sensitivity to individual talker differences in voice-onset time. Journal of the Acoustical Society of America. 2004; 115(6):3171–3183. [PubMed: 15237841]

Allen SJ, Miller JL, de Steno D. Individual talker differences in voice-onset-time. The Journal of the Acoustical Society of America. 2003; 113(1):544–552. [PubMed: 12558290]

Apfelbaum KS, McMurray B. Successes and failures in early word learning: An emergent property of basic learning principles. submitted.

Ballem KD, Plunkett K. Phonological specificity in children at 1;2. Journal of Child Language. 2005; 2005; 32:159–173. [PubMed: 15779881]

Bernstein LE. Perceptual development for labeling words varying in voice onset time and fundamental frequency. Journal of Phonetics. 1983; 11(4):383–393.

Boersma, Paul. Praat, a system for doing phonetics by computer. Glot International. 2001; 5(9/10): 341–345.

Bourne LE, Restle F. Mathematical theory of concept identification. Psychological review. 1959; 66:278–296. [PubMed: 13803353]

Burton MW, Baum SR, Blumstein SE. Lexical effects on the phonetic categorization of speech: The role of acoustic structure. Journal of Experimental Psychology: Human Perception and Performance. 1989; 15(3):567–575. [PubMed: 2527963]

Burton MW, Blumstein SE. Lexical Effects on phonetic categorization: The role of naturalness and stimulus quality. Journal of Experimental Psychology: Human Perception and Performance. 1995; 21(5):1230–1235. [PubMed: 7595247]

Bush RR, Mosteller F. A model for stimulus generalization and discrimination. Psychological review. 1951; 58(6):413–423. [PubMed: 14900302]

Cohen, LB.; Atkinson, DJ.; Chaput, HH. Habit X: A new program for obtaining and organizing data in infant perception and cognition studies (Version 1.0). University of Texas; Austin: 2004.

Cole JS, Linebaugh G, Munson C, McMurray B. Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach Journal of Phonetics. in press.

Creel SC, 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]

Dahan D, Magnuson JS, Tanenhaus MK, Hogan EM. Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Language and Cognitive Processes. 2001; 16(5/6):507–534.

Dale P, Fenson L. Lexical development norms for young children. Behavior Research Methods, Instruments, & Computers. 1996; 28:125–127.

Eimas PD, Siqueland ER, Jusczyk P, Vigorito J. Speech perception in infants. Science. 1971; 171:303–306. [PubMed: 5538846]

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]

Goldinger SD. Echoes of Echos? An episodic theory of lexical access. Psychological Review. 1998; 105(2):251–279. [PubMed: 9577239]

Gómez RL. Variability and detection of invariant structure. Psychological Science. 2002; 13(5):431–436. [PubMed: 12219809]

Havy M, Nazzi T. Better Processing of Consonantal Over Vocalic Information in Word Learning at 16 Months of Age. Infancy. 2009; 14(4):439–456.

Hillenbrand JM, Getty L, Clark MJ, Wheeler K. Acoustic Characteristics of American English vowels. Journal of the Acoustical Society of America. 1995; 97(5):3099–3111. [PubMed: 7759650]

Houston DM, Jusczyk PW. The role of talker-specific information in word segmentation by infants. Journal of experimental psychology. Human perception and performance. 2000; 26(5):1570–1582. [PubMed: 11039485]

Hollich, GH.; Jusczyk, PW.; Brent, M. Talker Variability and Infant Word Learning; Poster presented at the International Conference on Infant Studies; Toronto, Canada. 2002; Apr.

Johnson K. The role of perceived speaker identity in F0 normalization of vowels. Journal of the Acoustical Society of America. 1990; 88(2):642–654. [PubMed: 2212287]

Jongman A, Wayland R, Wong S. Acoustic characteristics of English fricatives. Journal of the Acoustical Society of America. 2000; 106:1252–1263. [PubMed: 11008825]

Jusczyk PW, Pisoni DB, Mullinex J. Some consequences of stimulus variability on speech processing by 2-month-old infants. Cognition. 1992; 43(3):253–291. [PubMed: 1643815]

Kessinger RH, Blumstein SE. Effects of speaking rate on voice onset time and vowel production: some implications for perception studies. Journal of Phonetics. 1998; 26:117–128.

Klatt DH. Speech perception: a model of acoustic-phonetic analysis and lexical access. Journal of Phonetics. 1979; 7:279–312.

Kuhl PK. Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. Perception & Psychophysics. 1991; 50(2):93–107. [PubMed: 1945741]

Kuhl PK, Andruski JE, Chistovich I, Ludmilla A, Chistovich A, Kozhevnikova EV, Ryskina VL, Stolvarova EI, Sindberg U, Lacerda F. Cross-language analysis of phonetic units in language addressed to infants. Science. 2007; 277(5326):684–686. [PubMed: 9235890]

Lisker L, Abramson AS. A cross-language study of voicing in initial stops. Word. 1964; 20:384–422.

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 Acoustical Society of America. 1993; 94(3):1242–1255. [PubMed: 8408964]

Luce PA, Pisoni DB. Recognizing Spoken Words: The Neighborhood Activation Model. Ear & Hearing. 1998; 19(1):1–36. [PubMed: 9504270]

Mani N, Plunkett K. Phonological specificity of vowels and consonants in early lexical representations. Journal of Memory and Language. 2007; 57(2):252–272.

Mani N, Plunkett K. Fourteen-month-olds pay attention to vowels in novel words. Developmental Science. 2008; 11(1):53–59. [PubMed: 18171367]

Marcus GF, Berent I. Are there limits to statistical learning? Science. 2003; 300:53–54. [PubMed: 12677042]

Maye J, Weiss DJ, Aslin RN. Statistical phonetic learning in Infants: facilitation and feature generalization. Developmental Science. 2008; 11(1):122–134. [PubMed: 18171374]

Maye J, Werker JF, Gerken L. Infant sensitivity to distributional information can affect phonetic discrimination. Cognition. 2002; 82(3):B101–B111. [PubMed: 11747867]

McMurray B, Aslin RN. Infants are sensitive to within-category variation in speech perception. Cognition. 2005; 95(2):B15–26. [PubMed: 15694642]

McMurray B, Aslin RN, Toscano J. Statistical learning of phonetic categories: Computational insights and limitations. Developmental Science. 2009; 12(3):369–378. [PubMed: 19371359]

McMurray B, Clayards M, Tanenhaus M, Aslin R. Tracking the timecourse of phonetic cue integration during spoken word recognition. Psychonomic Bulletin and Review. 2008; 15(6):1064–1071. [PubMed: 19001568]

Miller JL. Internal structure of phonetic categories. Language and Cognitive Processes. 1997; 12:865–869.

Miller JL. Mapping from acoustic signal to phonetic category: Internal structure, context effects and speeded categorization. Language and Cognitive Processes. 2001; 16:683–690.

Mullennix JW, Pisoni DB, Martin CS. Some effects of talker variability on spoken word recognition. Journal of the Acoustical Society of America. 1989; 85(1):365–378. [PubMed: 2921419]

Nazzi T. Use of phonetic specificity during the acquisition of new words: Differences between consonants and vowels. Cognition. 2005; 98(1):13–30. [PubMed: 16297674]

Nazzi T, New B. Beyond stop consonants: Consonantal specificity in early lexical acquisition. Cognitive Development. 2007; 22(2):271–279.

Newport EL, Aslin RN. Learning at a distance: I. Statistical learning of non-adjacent dependencies. Cognitive Psychology. 2004; 48:127–162. [PubMed: 14732409]

Nittrouer S. Learning to perceive speech: How fricative perception changes, and how it stays the same. The Journal of the Acoustical Society of America. 2002; 112(2):711–719. [PubMed: 12186050]

Nygaard L, Sommers M, Pisoni D. Speech perception as a talker-contingent process. Psychological Science. 1993; 5(1):42–46. [PubMed: 21526138]

Oakes LM, Coppage DJ, Dingel A. By land or by sea: The role of perceptual similarity in infants’ categorization of animals. Developmental Psychology. 1997; 33(3):396–407. [PubMed: 9149919]

Ohde RN, Haley KL. Stop-consonant and vowel perception in 3- and 4-year-old children. Journal of the Acoustical Society of America. 1997; 102(6):3711–3722. [PubMed: 9407663]

Pater J, Stager C, Werker JF. The perceptual acquisition of phonological contrasts. Language. 2004; 80(3):384–402.

Perkell, JS.; Klatt, DH. Invariance & Variability in speech processes. Lawrence Erlbaum Associates; Hillsdale, N.J.: 1986.

Pierrehumbert JB. Phonetic diversity, statistical learning, and the acquisition of phonology. Language and Speech. 2003; 43(2-3):115–154. [PubMed: 14748442]

Pisoni, D. Some thoughts on normalization in speech perception. In: Johnson, K.; Mullinex, JW., editors. Talker Variability and Speech Processing. Academic Press; San Diego: 1997. p. 9-32.

Quinn PC, Eimas PD, Rosenkrantz SL. Evidence for representations of perceptually similar natural categories by 3-month-old and 4-month-old infants. Perception. 1993; 22(4):463–475. [PubMed: 8378134]

Remez, R. Perceptual Organization of Speech. In: Pisoni, D.; Remez, R., editors. Handbook of Speech Perception. Blackwell publishing; Oxford: 2005. p. 28-50.

Restle F. A theory of discrimination learning. Psychological Review. 1955; 62(1):11–19. [PubMed: 14357523]

Rost GC, McMurray B. Speaker variability augments phonological processing in early word learning. Developmental Science. 2009; 12(2):339–349. [PubMed: 19143806]

Ryalls BO, Pisoni DB. The effect of talker variability on word recognition in preschool children. Developmental Psychology. 1997; 33(3):441–452. [PubMed: 9149923]

Saffran J. Statistical language learning: Mechanisms and Constraints. Current Directions in Psychological Science. 2003; 12:110–114.

Singh L. Influences of High and Low Variability on Infant Word Recognition. Cognition. 2008; 106(2):833–870. [PubMed: 17586482]

Slawinski EB, Fitzgerald LK. Perceptual development of the categorization of the /r-w/ contrast in normal children. Journal of Phonetics. 1998; 26:27–43.

Spencer J, Blumberg M, McMurray B, Robinson S, Samuelson L, Tomblin JB. Short arms and talking eggs: Why we should no longer abide the nativist-empiricist debate. Child Development Perspectives. 2009; 3(2):79–87. [PubMed: 19784383]

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]

Summerfield Q. Articulatory rate and perceptual constancy in phonetic perception. Journal of the Acoustical Society of America. 1981; 7(5):1074–1095.

Swingley D, Aslin RN. Lexical Neighborhoods and the Word-Form Representations of 14-Month-olds. Psychological Science. 2002; 13(5):480–484. [PubMed: 12219818]

Swingley D, Aslin RN. Lexical competition in young children's word learning. Cognitive Psychology. 2007; 54(2):99–132. [PubMed: 17054932]

Theissen ED. The effect of distributional information on children's use of phonemic contrasts. Journal of Memory and Language. 2007; 56(1):16–34.

Theissen ED, Hill EA, Saffran JR. Infant-directed speech facilitates word segmentation. Infancy. 2005; 7(1):53–71.

Teinonen T, Aslin RN, Alku P, Csibra G. Visual speech contributes to phonetic learning in 6-month-old infants. Cognition. 2008; 108(3):850–855. [PubMed: 18590910]

Toscano J, McMurray B. Cue integration with categories: A statistical approach to cue weighing and combination in speech perception. Cognitive Science. in press.

Toscano J, McMurray B. Online integration of acoustic cues to voicing: Natural vs. synthetic speech. submitted.

Utman J. Effects of local speaking rate context on the perception of voice-onset time in initial stop consonants. Journal of the Acoustical Society of America. 1998; 103(3):1640–1653. [PubMed: 9514028]

Vallabha GK, McClelland JL, Pons F, Werker JF, Amano S. Unsupervised learning of vowel categories from infant-directed speech. Proceedings of the National Academy of Sciences. 2007; 104:13273–13278.

Werker JF, Curtin S. PRIMIR: A Developmental Framework of Infant Speech Processing. Language Learning and Development. 2005; 1(2):197–234.

Werker, JF.; Fennell, CT. Listening to sounds versus listening to words: Early steps in word learning. In: Hall, DG.; Waxman, SR., editors. Weaving a Lexicon. MIT Press; Cambridge, MA: 2006.

Werker JF, Tees RC. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behavior and Development. 1984; 1984; 7:49–63.

Werker JF, Cohen LB, Lloyd VL, Casasola M, Stager CL. Acquisition of word-object associations by 14-month-old infants. Developmental Psychology. 1998; 34(6):1289–1309. [PubMed: 9823513]

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.

Yoshida KA, Fennell CT, Swingley D, Werker JF. Fourteen month-old infants learn similar-sounding words. Developmental Science. 2009; 12(3):412–418. [PubMed: 19371365]

Yu C, Smith LB. Rapid word learning under uncertainty via cross-situational statistics. Psychological Science. 2007; 18(5):414–420. [PubMed: 17576281]

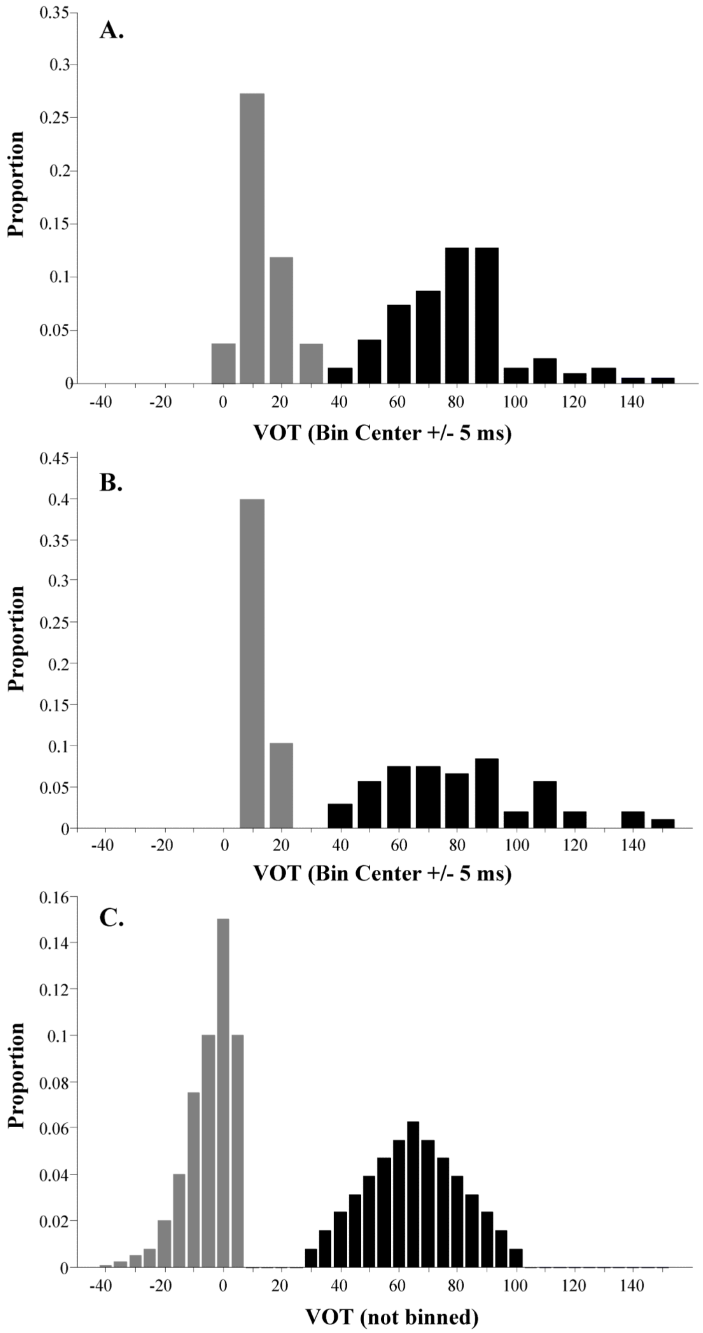

Figure 1.

Frequency distribution of word-initial VOT from (A) slow speech produced in Allen and Miller (1999), (B) items used in Rost and McMurray (2009), and (C) The distribution used in Experiments 1 & 2. Grey bars indicate tokens perceived as /b/, and black bars indicate tokens perceived as /p/.

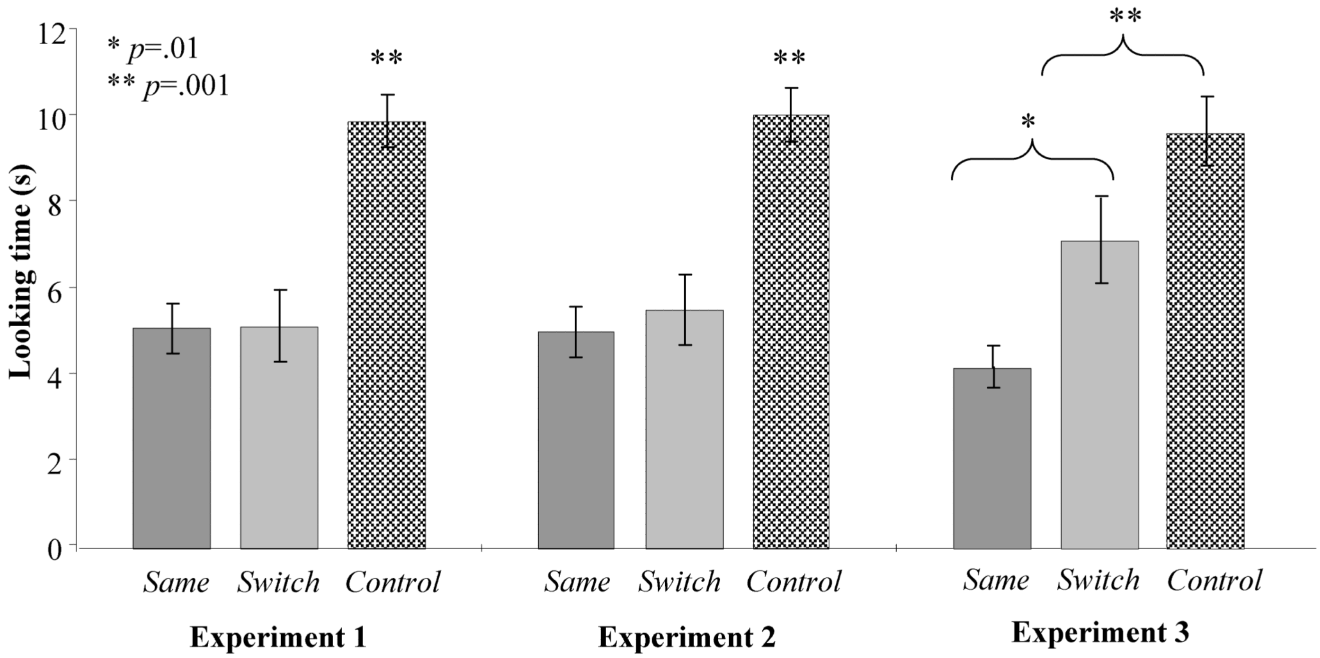

Figure 2.

Looking time in the same, switch and novel trials for Experiments 1, 2, and 3. Error bars represent SEM.

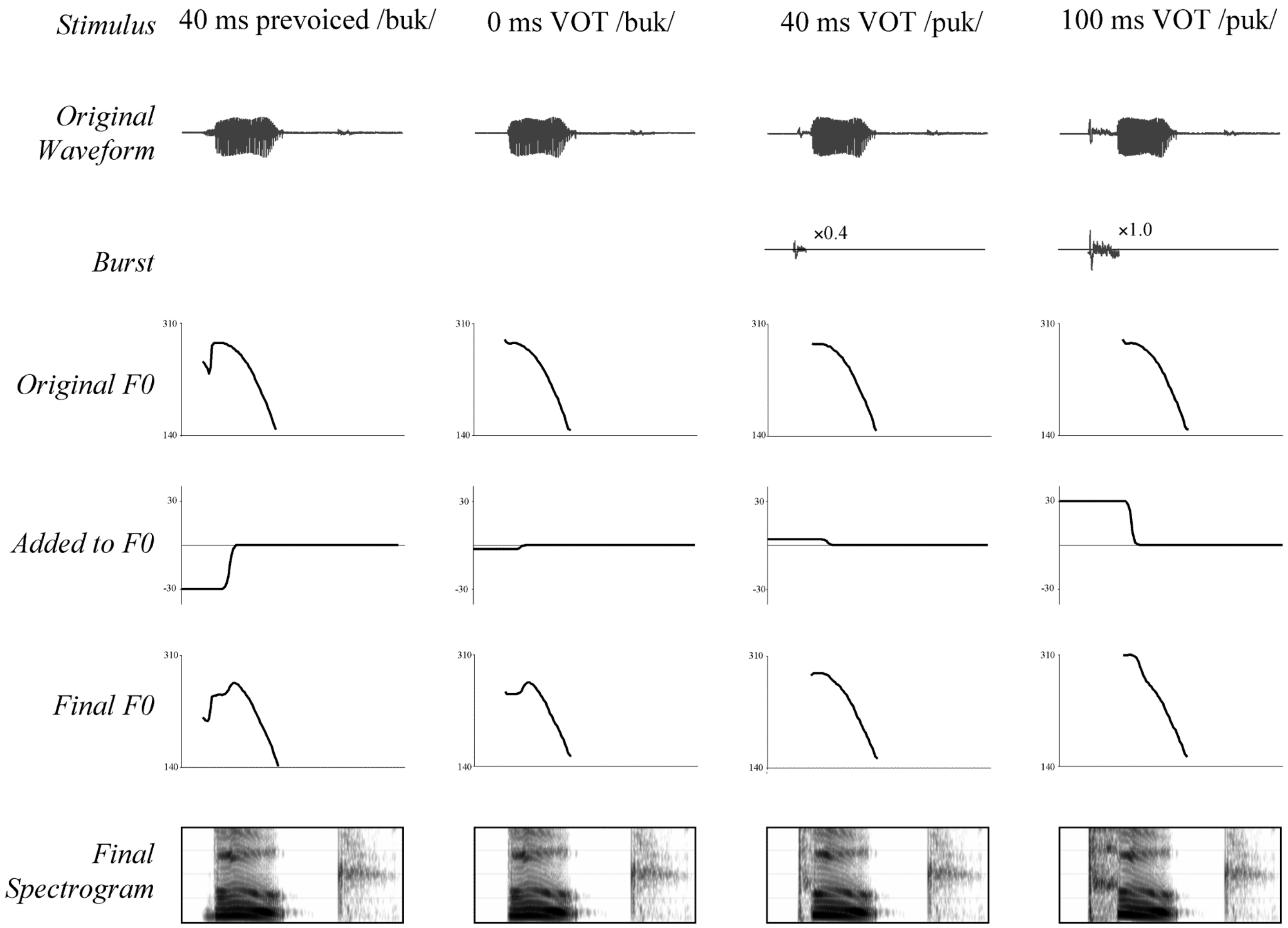

Figure 3.
Stimulus construction for Experiment 2. We start with the continuum from Experiment 1 (row 1). Bursts were excised and de-amplified proportional to their VOT (row 2). Next, the pitch track was extracted (row 3). We then constructed a modifier function (Row 4) which reduced the pitch by 30 Hz (at VOT=-40), and increased it by 30 Hz (at VOT=100). This modifier function only affected the first 50 ms of voicing. This function was added to the original pitch to get the pitch tracks in Row 5, and spliced into the final stimulus.

Table 1

Means of pitch and vowel quality measurements for items in Experiment 3. Standard deviations in parentheses

Cue	Location	/buk/	/puk/	*t(106)*	*p*
Vowel length (ms)	release-closure	237(8)	220(7)	1.09	.27
F0 (Hz)	onset	226 (84)	253 (97)	1.59	.11
	midpoint	237 (82)	232 (100)	.52	.60
F1 (Hz)	onset	386 (52)	372 (51)	1.36	.18
	midpoint	372 (51)	376 (48)	.35	.72
F2 (Hz)	onset	1356 (175)	1315 (175)	1.21	.23
	midpoint	1247 (195)	1208 (161)	1.11	.27