Cognition 129 (2013) 362–378

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence?

Bob McMurray [a,b,c,d,*], Kristine A. Kovack-Lesh [e], Dresden Goodwin [e], William McEchron [a]

[a] Dept. of Psychology, University of Iowa, United States
[b] Dept. of Communication Sciences and Disorders, University of Iowa, United States
[c] Dept. of Linguistics, University of Iowa, United States
[d] The Delta Center, University of Iowa, United States
[e] Dept. of Psychology, Ripon College, United States

ARTICLE INFO

Article history:
Received 20 September 2012
Revised 18 July 2013
Accepted 22 July 2013
Available online 24 August 2013

Keywords:
Infant directed speech
Speech categorization
Statistical learning
Phonetic analysis
Vowels
Voices onset time

ABSTRACT

Infant directed speech (IDS) is a speech register characterized by simpler sentences, a slower rate, and more variable prosody. Recent work has implicated it in more subtle aspects of language development. Kuhl et al. (1997) demonstrated that segmental cues for vowels are affected by IDS in a way that may enhance development: the average locations of the extreme "point" vowels (/a/, /i/ and /u/) are further apart in acoustic space. If infants learn speech categories, in part, from the statistical distributions of such cues, these changes may specifically enhance speech category learning. We revisited this by asking (1) if these findings extend to a new cue (Voice Onset Time, a cue for voicing); (2) whether they extend to the interior vowels which are much harder to learn and/or discriminate; and (3) whether these changes may be an unintended phonetic consequence of factors like speaking rate or prosodic changes associated with IDS. Eighteen caregivers were recorded reading a picture book including minimal pairs for voicing (e.g., *beach*/*peach*) and a variety of vowels to either an adult or their infant. Acoustic measurements suggested that VOT was different in IDS, but not in a way that necessarily supports better development, and that these changes are almost entirely due to slower rate of speech of IDS. Measurements of the vowel suggested that in addition to changes in the mean, there was also an increase in variance, and statistical modeling suggests that this may counteract the benefit of any expansion of the vowel space. As a whole this suggests that changes in segmental cues associated with IDS may be an unintended by-product of the slower rate of speech and different prosodic structure, and do not necessarily derive from a motivation to enhance development.

1. Introduction

During the first year of life, infants' speech perception systems begin to be tuned to the characteristics of their native language (Werker & Curtin, 2005; Werker & Tees, 1984). Over the first 12–18 months, infants show a reduction in their ability to discriminate phonetic contrasts that are not used in their language (Werker & Lalonde, 1988; Werker & Tees, 1984); they gain the ability to discriminate difficult contrasts (Eilers & Minifie, 1975; Eilers, Wilson, & Moore, 1977); and they are continually refining existing categories (Kuhl, Stevens, Deguchi, Kiritani, & Iverson, 2006). A growing number of scholars have posited that this process is guided, in part, by the statistics of acoustic cues in the speech that infants hear (de Boer & Kuhl, 2003;

* Corresponding author. Address: Dept. of Psychology, University of Iowa, E11 SSH, Iowa City, IA 52242, United States. Tel.: +1 319 335 2408 (voice); fax: +1 319 335 0191.
E-mail address: bob-mcmurray@uiowa.edu (B. McMurray).

0010-0277/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cognition.2013.07.015

Guenther & Gjaja, 1996; Maye, Werker, & Gerken, 2003; McCandliss, Fiez, Protopapas, Conway, & McClelland, 2002; McMurray, Aslin, & Toscano, 2009; Pierrehumbert, 2003; Toscano & McMurray, 2010; Vallabha, McClelland, Pons, Werker, & Amano, 2007), and recent work shows that computational models of this learning mechanism can account for all three patterns of development (McMurray, Aslin, et al., 2009).

Statistical learning is based on the idea that phonological speech contrasts can be described by one or more continuous acoustic cues, which themselves are the product of articulation. For example, voicing (which distinguishes /b, d, g/ from /p, t, k/) is marked primarily by voice onset time (or VOT, the continuous time between the release of the articulators and the onset of voicing) (Lisker & Abramson, 1964). For voiced sounds, like /b,d,g/, the release of the articulators (the lips or tongue) occurs nearly simultaneously with the onset of voicing (in languages like English), resulting in VOTs near 0 ms. For voiceless sounds, like /p, t, k/, the onset of voicing is delayed by about 50 ms after the consonantal release. However, variation across talkers, speaking rates, and the effects of other phonetic properties of the signal creates some variation around these means resulting in statistical clusters (Fig. 1A; Allen & Miller, 1999; Lisker & Abramson, 1964).

Analogously, most vowels can be characterized by the frequency of the first three formants and their duration (Hillenbrand, Getty, Clark, & Wheeler, 1995; Peterson & Barney, 1952). The vowel /i/ as in *beet*, for example has a low F1 and a high F2; while /ɑ/ as in *Bob* has a high F1 and a low F2. These individual formant frequencies derive in part from the position of the tongue during the articulation of the vowel; as this is variable as a function of talker, coarticulation, etc., those cues also form statistical clusters around the prototypical values for the vowels of the language. Here, however, clusters may only be distinct when examined in two dimensions (Fig. 1B; data from Cole, Linebaugh, Munson, & McMurray, 2010; see also Hillenbrand et al., 1995; Peterson & Barney, 1952).

Given this description of the input, distributional learning posits a fairly simple mechanism for acquiring speech categories. By estimating the mean (or prototypical) cue-value and variance (or extent of allowable variation around this mean) of each cluster, children could arrive at a reasonable set of descriptors for the categories along a dimension or dimensions. There has been an explosion of computational models that show this can be done by a variety of learning mechanisms (de Boer & Kuhl, 2003; Guenther & Gjaja, 1996; McMurray, Aslin, et al., 2009; McMurray & Spivey, 2000; Toscano & McMurray, 2010; Vallabha et al., 2007). These models demonstrate how a variety of [largely] unsupervised clustering approaches can harness the statistical structure of the input to find the relevant categories, and thus establish the computational tractability of this hypothesis.

Evidence for such mechanisms comes from two sources. First, adult perceptual categories show a graded structure (Andruski, Blumstein, & Burton, 1994; Kuhl, 1991; McMurray, Aslin, Tanenhaus, Spivey, & Subik, 2008; McMurray, Tanenhaus, & Aslin, 2002; Miller, 1997; Miller & Volaitis, 1989; Toscano, McMurray, Dennhardt, & Luck, 2010; Utman, Blumstein, & Burton, 2000; Volaitis & Miller, 1992) that matches the graded clusters of speech cues. Infants are also sensitive to such gradations, contra earlier claims of categorical perception (Galle & McMurray, submitted for publication; McMurray & Aslin, 2005; Miller & Eimas, 1996). This correspondence suggests that this gradiency may be a remnant of the statistical learning process that undergirds development (McMurray & Farris-Trimble, 2012; McMurray, Horst, Toscano, & Samuelson, 2009).

Second, laboratory learning studies by Maye and colleagues have documented that distributional learning can occur over a short time span. Maye et al. (2003) exposed infants to a stream of speech sounds in which VOT clustered either bimodally (two categories) or unimodally (one category) and then tested their subsequent discrimination. Eight-month-olds that received bimodally structured input discriminated tokens that straddled the center of the continuum, while those receiving unimodally structured input did not. This suggests that this short (2 min) exposure to statistically structured speech was sufficient to bias discrimination, at least immediately after exposure. Given infants' likely abilities to discriminate these tokens prior to exposure, a unimodal distribution was sufficient to collapse categories. Subsequent work demonstrated the converse, that exposure to a bimodal distribution of speech sounds helps infants separate categories they do not already have (Maye, Weiss, & Aslin, 2008). Moreover, later in development, by 10 months, infants have difficulty using distributional statistics for speech sounds not in their native language (Yoshida, Pons, Maye, & Werker, 2010), suggesting that the perceptual

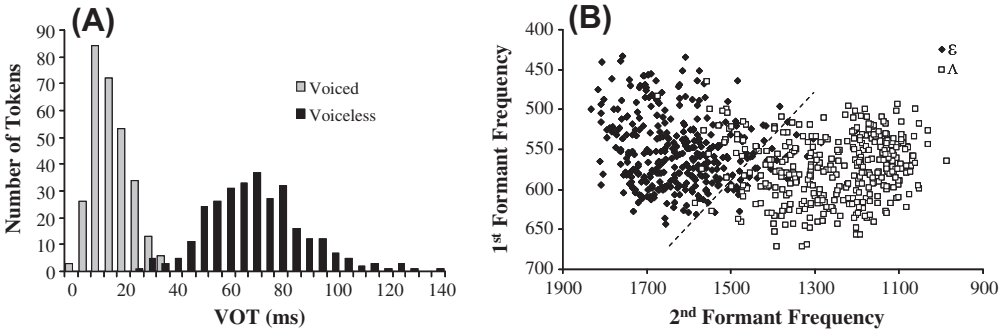

Fig. 1. The statistical distributions of various speech cues. (A) Voice Onset Time (from Allen & Miller, 1999); (B) formant frequencies for two vowels from the male speakers of Cole et al. (2010).

reorganization that is occurring at this time (the tuning speech perception to the categories of the native language) is not independent of distributional learning.

All of this suggests that the statistical properties of the input that infants receive are crucial for the development of speech perception. As a result, to understand development, we must understand the statistics of what infants hear. Consequently, there is increasing interest in the statistics of speech cues as they appear in infant-directed-speech or IDS, which is likely to be a large component of the input for most infants (cf., Bion, Miyazawa, Kikuchi, & Mazuka, 2013; Cristia & Seidl, 2013; Englund, 2005; Kuhl et al., 1997; Werker et al., 2007).

1.1. Infant Directed Speech (IDS)

IDS is marked by shorter utterances, a slowed speaking rate, longer pauses, higher absolute pitch, and much more variability in pitch (Fernald et al., 1989; Soderstrom, 2007). Infants' responses to IDS has been extensively studied and it is well documented that infants show a robust preference for IDS over adult-directed speech (ADS) from the neonatal period through 4 months (Cooper & Aslin, 1990; Fernald, 1985; Pegg, Werker, & McLeod, 1992). However, there is debate over whether this preference persists or changes after 8 months (Hayashi, Tamekawa, & Kiritani, 2001; Newman & Hussain, 2006; Zangl & Mills, 2007).

A number of researchers have examined the statistical distributions of speech cues in IDS. Werker et al. (2007), for example, measured the vowel duration and the first and second formants for two vowel contrasts in IDS for English (/ɪ/ vs. /i/ and /ɛ/ vs. /eɪ/) and Japanese (/i/ vs. /iː/ and /ɛ/ vs. /ɛː/). Using logistic regression, they showed that the statistics across these cues were sufficient to discriminate the contrasts, and that the particular cues signaling these changes were appropriate to the language (e.g., duration was more informative in Japanese). This establishes that the statistics present in the input to children (IDS) are sufficient to support a learning mechanism such as distributional learning (though see, Bion et al., 2013). However, it leaves open the question of whether IDS changes the statistical properties of the input (relative to ADS) in a way that could affect infant speech development.

Kuhl et al. (1997) addressed this by examining free conversation between mothers and their infants (IDS) or with other adults (ADS), in mothers who spoke American English, Swedish, or Russian. They measured the first and second formant frequencies of three cardinal vowels (/i/, /ɑ/, /u/) and showed that the distance (in F1 × F2 space) between the prototype values of each of the vowels expanded in IDS, implying that the vowel prototypes are more discriminable from each other (on the basis of these two cues). This suggests that IDS is an ideal fodder for statistical learning; mothers appear to enhance the statistical structure of their vowels to support better phonetic category learning. Liu, Kuhl, and Tsao (2003) expanded on this showing that the infants of mothers who show greater expansion of their vowel space also showed better speech discrimination skills, implying that this modification to their speech helped enhance the development of speech perception abilities.

One critical question is why IDS affects the statistical distribution of segmental cues? One possibility is that at some level, caregivers are enhancing the statistical distinctiveness between vowels *in order to facilitate development or perception*. That is, these phonetic modifications may be undergone with the listener (the infant) in mind. While it is possible that such a motivation could be conscious and explicit, the fact that such changes are often pitched in an evolutionary framework (Kuhl et al., 1997), suggest it could be more instinctive or implicit on the part of caregivers (or even developed in dynamic interaction with the child; see, for example, Smith & Trainor, 2008, for an analogous situation with pitch). Alternatively, such enhancement may occur *as a natural consequence of other properties of IDS* like the slower rate of speech, the increased number of stressed syllables, or changes in prosodic or affective factors. However, effects of IDS on segmental cues have not been adequately assessed in a way that provides confidence in the generality of the effect, or with these questions in mind. This is the goal of the present study.

1.2. Does IDS improve perception and development?

Kuhl et al.'s (1997) analyses focused on vowels at the corners of the vowel space. These vowels should be easy for infants to discriminate and thus may not require enhancement. However, there are many vowels between these extremes. For example, between the back vowels, /ɑ/ and /u/ of English, most dialects also have /ɔ/, /ou/ and /ʊ/ at intermediate heights, and there are diphthongs that pass though this space like /ɔɪ/ (as in *boy*), /ɑʊ/ (as in *brown*) and /ɑɪ/ (as in *bite*). It is unclear if the enhancement observed in the point vowels is also seen in the interior vowels, where it would be more useful at discriminating acoustically close or overlapping vowels. This is particularly important, as Neel (2008) has demonstrated that the acoustic distinctiveness of neighboring vowels is more predictive of a talker's intelligibility than the area of their vowel space, the measure used by Kuhl et al. (1997).

Even if the point vowels do show expansion, this may not conclusively indicate a benefit for speech perception. Kuhl et al. (1997) statistical analyses focused on the *mean* formant values for a given vowel and the distance between them. However, they did not examine the *variability* in vowel productions. This may be very important for statistical learning: Toscano and McMurray (2010) for example, in a model of how people learn to combine cues in speech perception suggest that the discriminability of any two categories is a function of both the distance between them and their variability. Indeed the importance of variability in establishing the significance of a difference in means is central to statistics (Student, 1908).[1] It is unclear how IDS affects the variability of cues like vowel formants. However,

[1] The aforementioned studies of IDS did evaluate the significance of the vowel space expansion implying an analysis of their variance. However, they used the between-subject variance, averaging across all of the vowels of an individual talker. To an individual child, however, what matters is if the difference in mean articulations outweighs the variance across utterances *by their own mother*; that is the variance within a talker.

as IDS is typically described as more variable on other dimensions like pitch (Fernald et al., 1989), it seems likely that segmental cues will also be more variable. This raises the possibility that an increase in variability in IDS could outweigh any benefit of the expansion of the vowel prototypes. Thus, an important goal of this study is to reexamine how vowels change in IDS, by examining the interior vowels in a more carefully controlled task and by examining both mean formant frequencies and their variance.

In this regard, a recent study by Cristia and Seidl (2013) offers some insight. They measured several internal vowel contrasts in IDS and ADS (/ɪ/ vs. /i/, and /eɪ/ vs. /ɛ/ as well as the non-phonemic nasal/oral contrast between /æ/ vs. /æ~/ and / ɛ / and / ɛ /). The found little evidence for expansion among either type of interior contrast coupled with greater variance in the relevant cues. At the same time, point vowels did show expansion, suggesting that point vowels alone may not accurately describe the clarity of the vowel space in IDS. However, the contrasts they studied were restricted to two features (tenseness and nasality), one of which is non-phonemic. Thus, one of our goals was to extend this to a wider range of vowels which contrast in frontness, height and rounding.

1.3. Does the effect of IDS extend to new contrasts?

A second question is whether the effect of IDS on segmental properties appears in other cues or if it is an isolated property of vowels. Liu et al.'s (2003) study suggests a more widespread effect: while they measured mothers' vowel-spaces, their test of infant speech discrimination was on a fricative/affricate distinction. The most direct cause of such effects would be that the mothers who expand their vowel space are also likely to enhance the fricative/affricate distinction in their speech to infants (although Liu et al., did not measure this). However, an alternative is that the mothers who engage in more interaction with their children, or simply expose them to more language, also expand their vowel space. In this case, differences in parenting and language input more broadly lead to both improved speech perception and changes in the caregiver's vowel space. Thus, we cannot assume from Liu et al. (2003) that other cues are also affected.

There has been little work directly examining IDS effects on other phonetic cues, with the exception of a number of studies on VOT. However, these offer conflicting results. Sundberg and Lacerda (1999) tested six Swedish speaking mother/infant pairs in a period of free conversation with their infants and found that VOTs in IDS appear to shorten. In contrast, Englund (2005) examined six Norwegian mothers, also in free conversation, and found the opposite: both voiced and voiceless VOTs lengthen. These results conflict with each other, but, more importantly, neither are clearly consistent with enhancement – if both VOTs simply move in the same direction, this does not clearly enhance the distinction between them, and may make it worse by either reducing the distance, or by moving the mean VOTs away from the prototypical values for the language. However, both of these studies used relatively small samples of caregivers (though an admirably large sample of VOTs), and relied on free conversation. As a result, factors like word choice, prosodic position, or speaking rate that affect VOT may also vary between IDS and ADS. Neither study attempted to control for these factors, or to use measurements sensitive to speaking rate. Thus, a goal of this study was to examine for enhancement in VOT in a more systematic approach that may control for these factors.

1.4. Are IDS effects independent of more basic changes?

Such factors raises the possibility that even if speech categories are enhanced in IDS, these specific effects may not be intended and rather may be the byproduct of prosodic, word-choice, and speaking rate changes that also occur in IDS. That is, it is important to understand the effect of IDS on speech cues *over and above* these supra-segmental changes. The cardinal or point vowels tend to move in formant space due to a variety of low-level factors. Speaking rate (Van Son & Pols, 1990), prosodic accent (Cho, 2005), speaking style (Smiljanic & Bradlow, 2008), and even the number of phonological neighbors of a target word (Munson & Solomon, 2004) can all cause the vowel space to expand or contract. This is because a vowel's formants commonly undershoot their target frequency values in fast speech, unstressed syllables, and reduced syllables. By slowing down, and using more stressed single-syllable words in IDS, the articulators have more time to reach their targets which could account for the changes to phonetic cues in IDS. If this is the case, such effects may not reflect a motivation to improve statistical learning. Rather, they may be an unintended (though perhaps helpful) consequence of other changes in speaking style.

The three existing studies of mothers' vowel space expansion (Cristia & Seidl, 2013; Kuhl et al., 1997; Liu et al., 2003) used spontaneous speech, but preselected particular words for analysis. This ensures that differences in word-choice among the two registers cannot account for the effects of IDS. However, there is still the possibility that these words appear in different prosodic positions, or with different degrees of prosodic strength (cf., Cristia & Seidl, 2013, for a similar argument). Cho (2005) showed that /i/ and /ɑ/ generally expand as a function of prosodic position or accent, and if IDS used more prosodically strong positions, this could account for the expansion. It is also likely that words spoken in IDS are slower which could also account for the expansion. Thus, the effect of IDS on segmental cues could derive from prosody, and it is important to rule this out.

This last question may also best be addressed by looking at a different phonetic cue. VOT offers an ideal test case. As a temporal variable, the relationship of VOT to speaking rate is clear and well understood. Typically in slower speech (in English), the voiceless category (e.g., /p, t, k/) shows longer VOTs (Allen & Miller, 1999; Kessinger & Blumstein, 1998; Miller, Green, & Reeves, 1986), and the voiced category (e.g., /b, d, g/) either increases slightly (Magloire & Green, 1999) or shows little change (Kessinger & Blumstein, 1998). In contrast, if VOT is deliberately enhanced, voiceless VOTs should get longer, and voiced VOTs should get shorter (or negative/ pre-voiced). We also understand how to quantify the relationship between

speaking rate and VOT using the length of the subsequent vowel as a proxy for speaking rate. This relationship has been formally described as the Consonant/Vowel (CV) ratio, the ratio between the VOT and the length of the vowel (Boucher, 2002; Pind, 1995; Port & Dalby, 1982). If there is enhancement, understanding this effect in terms of the CV ratio can help determine how much of this effect is due to the infant's needs (enhancement) and how much is a consequence of slowing. Thus, even if the predictions of enhancement for the voiced category (shorter VOTs and/or pre-voicing) are not found, we can use the CV ratio to ask if there is an effect of IDS *over and above* the effect of speech rate. In contrast, if VOT changes derive solely from speaking rate, then phonetic changes due to IDS may be a by-product of other factors.

1.5. Logic and goals

The present study was designed to answer each of these three questions. Our primary goal was to conduct a more thorough and controlled assessment of VOTs in IDS and ADS. This was done to both extend the investigation of IDS effects on segmental cues, and to investigate a cue for which enhancement makes different predictions from processes like speaking rate. Our secondary goal was to reassess vowel space expansion by looking at the interior vowels and by examining the relationship between the mean and variance. In doing so, it was crucial to control as much of the linguistic and situational content as possible to ensure that any differences found were due to IDS. Thus, we asked mothers to read simple picture books to either their infant or an experimenter. This allowed us to control the specific words being used and the prosodic frame in which they were situated. These picture books included minimal pair words spanning the voicing contrast at all three places of articulation, to control the vowel and final-consonants across voicing pairs, as both of these factors can also affect VOT and word length. We also measured syllable length as an estimate of speaking rate to examine whether IDS-induced modifications to VOT could be seen over and above the rate changes. With regard to the vowels, our word list was primarily geared toward providing a precise investigation of VOT (where the clearest predictions can be made), emphasizing minimal pairs for voicing (that were reasonable in a picture book). As a result, we were not able to control the vowels to the same degree. Thus, our goal was to ensure a fairly a wide range of vowels across the minimal pairs, and critically to include substantial interior vowels for analysis.

2. Experiment

2.1. Method

2.1.1. Participants

Participants were 18 parent-infant dyads from the Ripon, WI area. Two were male, and 16 were female. Infants ranged from 9 to 13 months (M = 325.83 days, SD = 32.85 days, range 273–400 days; 13 males and 5 females). All were Caucasian and lived in homes where English was the primary language. Infants' names were obtained from local birth announcements and contacted first through a letter, then followed up with a phone call. Infants received a small gift for participating. An additional 4 mother-infant dyads were tested but excluded from the final analyses due to equipment or experimenter error (n = 3) or sibling interference (n = 1).

2.1.2. Design and stimulus

There were 24 target words in this experiment. The target words consisted of 12 minimal pairs (e.g., *bear*/*pear*), spanning the three consonantal places of articulation (coronal, labial, or velar), and a range of vowels (see Table 1).

Picture books depicted each word on one page. On each page, there was a picture corresponding to the word and three sentences. Sentences were constructed such that the word would appear in three positions: the first word in the sentence (*Pears are yummy*); the second word after a determiner (*The pears are for eating*) and as the final word in the sentence (*There are three pears*). Ultimately, we wanted recordings of the parent speaking the same word in ADS and IDS, but we did not want the different versions of the same word to be spoken in much proximity. Thus, we split the word list into two books. Each book contained two (of four) labial pairs, two coronal pairs, and two velar pairs. Parents read one book to their infant, and the other to the adult experimenter, thus reading the entire list over a single session. They then returned several days later (M = 3.83 days; SD = 2.96 days) and the assignment was swapped, such that the words read to the infant on day 1 were read to the experimenter on day 2. In order to eliminate any order effects, the counterbalanced books used on the second day had their pages in a different random order. This led to four counterbalanced books (one pair of books for each day) that were used for each participant, with each book containing 12 of the 24 words. As a result of this, each word was spoken in both registers, and both registers were spoken on both days, but a word was never spoken in the same register on the same day. Two sets of these books were constructed with different random assignments of words per day (and in different orders) and the set of books was randomly selected for each dyad on the first day. As a consequence of this design, for each participant, we had recordings of 12 minimal pairs for

Table 1
Words used in the experiment.

Voiced	Voiceless	Place of articulation	Vowel
Bugs	Pugs	Labial	ʌ
Baths	Paths	Labial	æ
Bowls	Poles	Labial	ou
Bears	Pears	Labial	eɹ
Dime	Time	Coronal	ɑɪ
Dunes	Tunes	Coronal	u
Darts	Tarts	Coronal	ɑɹ
Deer	Tear	Coronal	ɪɹ
Guards	Cards	Velar	ɑɹ
Goats	Coats	Velar	ou
Gaps	Caps	Velar	æ
Girls	Curls	Velar	ɝ

voicing (4 labial, 4 coronal, 4 velar) × 2 words/pair (voiced and voiceless) × 3 sentence positions × 2 registers (ADS/IDS), yielding 144 utterances per dyad.

2.1.3. Procedure and apparatus

Testing was conducted in a small, quiet room. When parents read to their infants, they were seated in a chair large enough for the infant to sit on their parent's lap or next to them. Parents read the book aloud to their infants as naturally as they would at home. The experimenter stood outside of the room to minimize distractions. When parents read to the adult experimenter, they were seated in the same chair and room with the experimenter seated across from the parent. They were asked to read the book as they would to an adult. Infants were cared for by a second experimenter in an adjacent room during this time. During each appointment, half of the words (one book) were read to the infant and the other half were read to the experimenter, and the order of these conditions was counterbalanced across participants and appointments.

Speech was recorded using a Marantz Solid State PMD670 Recorder and a Shure SM48 Microphone. Microphone placement was tricky as infants liked to play with head-mounted or lavalier style microphones, and even a traditional microphone needed to be out of reach to be ignored. Thus, we mounted a directional microphone on a microphone stand located to the left of the chair, about 40 cm from their mouth. Recordings were made directly to WAV files at 44,100 kHz, 16 bit A/D conversion.

2.1.4. Phonetic measurements

For each recording, a text-grid was built using Praat (Boersma & Weenink, 2009) by one of three trained naïve coders. This grid marked the onset of the release burst, the onset of laryngeal voicing, and the closure of the vowel/syllable. The release burst and voicing onsets were marked directly from the waveform at the closest zero crossing and the closure was marked from the spectrogram at the point when the upper formants (F2 and F3) were no longer visible[2] or the onset of frication for plurals. From these grids the VOT, vowel length, and pitch were automatically generated. Grids were checked by one of the authors with phonetic training.

VOT was extracted from these grids as the difference between the onset of voicing and the release burst. If voicing preceded the release burst, this was coded as a negative VOT (pre-voicing). VOTs were further subject to a series of audits which flagged any utterance whose VOT may be out of range for its intended articulation. This was based on the range of variation from studies of VOT in various rates of speech, (Allen & Miller, 1999; Kessinger & Blumstein, 1998) along with the first author's extensive experience measuring VOT. For voiced sounds, VOTs less than −50 ms or greater than 40 ms were flagged; for voiceless sounds, VOTs less than 25 ms or greater than 100 ms were flagged. These VOTs were measured manually by one of the authors and included in the analysis, or marked as uncodeable. Next, *vowel length* was coded as the time difference between the release burst and the closure duration.

Once the TextGrids were established and verified, we computed the formant frequencies and pitches of the vowels. While the vowel-space is often informally described by F1 × F2 alone, we also measured F3, which is a useful cue for rounding, and a secondary cue for height and/or backness. Formant frequencies for the first, second, and third formant were automatically estimated using the *Hack-SL* method in Praat (Boersma & Weenink, 2009) over a 50 ms window straddling the center of the vowel. However, in our experience formant frequency computations are rarely 100% (or even 80%) accurate, and accuracy can be greatly affected by the free parameters of the algorithms. Thus, we extracted formant frequencies twice using two different parameter sets (changing the maximum frequency parameter to 6000 and 7000). After this, one of the coders used a specially designed Matlab script to view each set of formant tracks on top of the spectrogram, and to choose which was correct or to indicate that neither was. If neither of the automatic values was accurate, formants were coded by the fourth author from the spectrogram. Over, the course of manually verifying the formants we noticed that the *dune*/*tune* pair was consistently mismeasured. Its high back rounded vowel naturally has a very high F1 and a very low F2; when combined with the following nasal, the formants "blended" making it very difficult to disambiguate them. Thus, *dune* and*tune* were excluded from the formant analyses.

Finally, we extracted the pitch. Pitch was estimated at 10 ms windows, and these were averaged within particular time-windows to obtain the mean pitch for the word. Most pitch tracks were averaged across a 50 ms window straddling the center of the vowel; however, if Praat failed to extract a track in this window, it was extended to 100 ms. Those that failed at 100 ms were hand measured or marked as unmeasurable by one of the authors.

2.2. Results

The dataset from this experiment permitted a wide range of analyses. In the interest of focus, we present only the most important ones here. A number of additional analyses can be found in the online supplement, and we refer to them as they relate to our findings.

Data were analyzed using linear mixed effects models with the LME4 package Bates, Maechler, & Bolker, (2011) of R. This approach was chosen for two reasons. First, our study had two random effects, the dyad and the word-pair (item) and mixed models can account for both sources of variance in a single model (Baayen, Davidson, & Bates, 2008), yielding more power. Second, mixed effects models, as a variant of regression, cope with missing data more gracefully. There are a number of choices for how to code random and fixed effects and which to include in the model. Thus, we started by comparing a range of models—without examining the fixed effects—to determine the optimal model for the data (Supplement S1). The result of this was a model with participant and word-pair as random intercepts. To compute*p*-values for parameters in the model, we used Markov-Chain Monte Carlo sampling (MCMC)

[2] For words ending in approximants like *deer*/*tear*, there was often no closure before the word (e.g., *The deer are*...). In this case, we used the point at which the formants reached their minima to determine the word offset.

with 20,000 iterations. To establish the significance of main effects that spanned two parameters (e.g., the main effect of place of articulation had three levels so required two variables), we added the main effect in question to a model with only random effects and used the χ^2 test of model comparison to compute significance of the factor as a whole. A similar procedure was followed for interactions, comparing a model with only main effects to one with just the interaction of interest. *P*-values reported in the text are based on χ^2 unless noted.

Before analyzing the VOT or formant frequencies, we first analyzed word length, pitch, and pitch variability to confirm that our talkers were speaking in the appropriate registers. This analysis (Online Supplement S2) shows that mothers used longer words, higher pitch and more pitch variability in IDS than ADS.

2.2.1. Voice onset time

Our analysis of VOT addressed three questions: (1) Does IDS affect VOT; (2) Does it do so in a way consistent with enhancement; and (3) Can these changes be attributed to other factors like speaking rate? If IDS enhances phonetic distinctions (to benefit the infant), VOTs for aspirated segments (/p, t, k/) should get longer, and those for voiced segments (/b, d, g/) should get shorter. In contrast, if the effect of IDS/ADS is due to speaking rate, the voiced segments should remain unchanged (or get slightly longer) and the voiceless segments should increase. Fig. 2 shows the effect of register and voicing on VOT and appears more consistent with speaking rate, with longer VOTs in both categories for IDS.

To validate this statistically, we used a linear mixed effects model to examine the effect of register, place, and position on VOT. Register was sum coded (ADS = −.5; IDS = +.5). Position was coded as two sum-coded variables (Position 1 vs. other: +.667/−.333; and Position 2 vs. other: +.667/−.333) as was place of articulation (Labial vs. not-labial: +.667/−.333; Velar vs. not-velar: +.667/−.333). Each main effect was allowed to interact with register, but not with each other (and there were no higher level interactions) as model selection (Online Supplement S1) suggested such interactions did not improve model fit. As we anticipated the possibility of different effects of register for voiced and voiceless sounds, separate models were run for each category. For all of our analyses of VOT we excluded any token for which either VOT or word length (WL) was marked as uncodeable during the audit and subsequent manual measurement. (We excluded based on both cues since they were to be combined in the next analysis). Of the 2592 total tokens, this excluded 71 tokens (M = 3.94/sub or 2.7%/sub). More of these were for IDS (55) than ADS (16; $T(17)$ = 3.3, p = .0039). We also excluded 36 pre-voiced tokens (M = 2.0/sub), and discuss them separately.

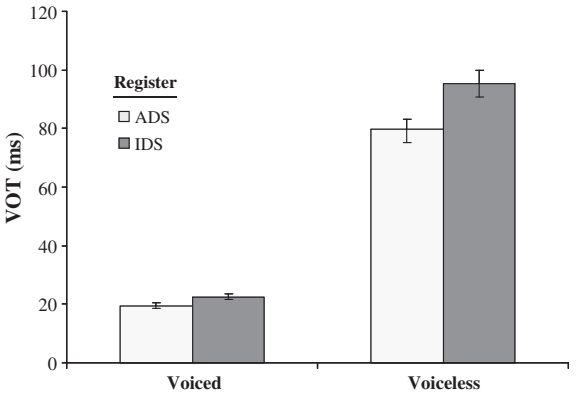

Fig. 2. The effect of voicing and speech register on VOT.

For voiced tokens (Table 2a), we found a significant main effect of register (p < .0001). Consistent with the predictions of a speech rate account, voiced VOTs increased slightly from ADS (M = 19.5 ms, SD = 3.4) to IDS (M = 22.5 ms, SD = 4.5). As seen in prior work (Lisker & Abramson, 1964), place of articulation was also significant (p < .0001). Labials had the shortest VOTs (M = 13.1, SD = 3.0), followed by coronals (M = 20.6, SD = 5.2) and then by velars (M = 29.2, SD = 4.8). Position was significant (p = .045), though the differences were numerically very small (1st position: M = 21.9 ms, SD = 4.7; 2nd position: M = 20.8 ms, SD = 4.9; 3rd position: M = 20.1 ms, SD = 3.3). Neither interacted with register.

Voiceless tokens (Table 2b) showed a similar pattern. Again, register was significant (p < .0001): VOTs were longer in IDS (M = 95.4, SD = 19.3) than ADS (M = 80 ms, SD = 13.9). We also found a significant effect of place of articulation (p = .03). Labials had the shortest VOTs (M = 79.5 ms, SD = 16.7), followed by velars (M = 89.0 ms, SD = 14.3) then coronals (M = 94.0 ms, SD = 18.7). Position was significant (p = .013), though this was largely a function of the words in 3rd position (M = 83.9 ms, SD = 17.4) which were shorter than the other two (1st position: M = 89.3, SD = 16.3; 2nd position: M = 89.3, SD = 18.6). Register did not interact with place of articulation (p = .53), however it did interact with position (p = .0098). Follow-up analyses at each position showed that the main effect of register was significant and in the same direction at all three positions (1st Position: B = 21.6, SE = 2.5, p_{mcmc} < .0001; 2nd position: B = 10.9, SE = 2.6, p_{mcmc} < .0001; 3rd position: B = 13.8, SE = 2.6, p_{mcmc} < .0001).

Thus, for both voiced and voiceless sounds, VOTs increase in IDS. This suggests a locus in speaking rate, not enhancement. To test this possibility, we next computed the ratio of VOT to vowel length (CV ratio) to create a measure of voicing that accounts for differences in rate. Fig. 3 shows a remarkably reduced difference between IDS and ADS using this measure.

CV-ratio was entered into a pair of mixed effects models for voiced and voiceless tokens ($|R_{max}|$ = −.013 and −.009, respectively) and the results are shown in Table 3. For both analysis, there was no effect of register (Voiced: B = .0002, SE = .0021, p = .82; Voiceless: B = .0057, SE = .0038, p = .15). Place of articulation was significant for Voiced (p = .00013), but not voiceless (p = .12), though trends were in the same direction as the analysis of VOT. Position was significant (Voiced: p < .0001; Voiceless: p < .0001) and followed the same pattern as VOT. This suggests that prosodic position can exert an effect on voicing over and above that of slowing. None of the interactions with register were significant.

Table 2a
Results of a linear mixed effects model examining VOT for voiced sounds.

Factor	*B*	SE	p_{mcmc}	χ^2	df	*p*
Register	2.9	0.6	<.0001	26.5	1	<.0001
Place of artic						
Labial vs. other	−7.5	2.2	0.0057	22.4	2	<.0001
Velar vs. other	8.5	2.2	0.0026			
Position						
1st vs. Other	1.7	0.7	0.014	6.2	2	0.045
2nd vs. Other	0.6	0.7	0.42			
IDS Place						
× Labial	−1.7	1.4	0.20	1.7	2	0.42
× Velar	−0.6	1.4	0.64			
IDS × Position						
× P1	1.7	1.4	0.22	2.1	2	0.35
× P2	0.0	1.4	0.99			

Table 2b
Results of a linear mixed effects model examining VOT for voiceless sounds.

Factor	*B*	SE	p_{mcmc}	χ^2	df	*p*
Register	15.6	1.5	<.0001	101.5	1	<.0001
Place of artic						
Labial vs. other	−14.4	5.3	0.021	7.0	2	.03
Velar vs. other	−4.9	5.3	0.38			
Position						
1st vs. Other	4.8	1.9	0.011	8.7	2	0.013
2nd vs. Other	4.8	1.9	0.0086			
IDS × Place						
× Labial	−1.6	3.7	0.65	1.3	2	0.53
× Velar	2.5	3.7	0.51			
IDS × Position						
× P1	7.5	3.7	0.045	9.2	2	0.0099
× P2	−3.4	3.7	0.36			

Thus, when we account for the slowing observed in IDS, there was not any unique effect of IDS on VOT.

One alternative possibility is that VOTs obey a lawful relationship with speaking rate, but that caregivers may add an additional enhancing gesture like pre-voicing (for voiced sounds) in IDS. Of the 18 caregivers, 10 used pre-voicing on at least one of the 36 voiced tokens. However, within these participants, pre-voicing was rare. Parents averaged 1.6 pre-voiced tokens in ADS and 2.0 tokens in IDS, though this difference was not significant ($t(9) = .77$, $p = .46$). Thus, pre-voicing appears to be rare in general, occurring in little more than half the participants and even among them it occurs infrequently, and not differentially more in IDS.

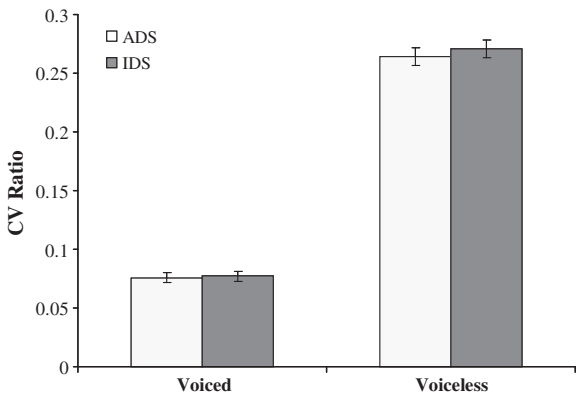

Fig. 3. CV ratio as a function of stop-type and register.

To summarize, the effect of IDS on raw VOTs was not in the direction predicted by intentional enhancement. Rather, VOTs were simply lengthened in IDS. The analysis of CV ratio confirms this story: when speaking rate is accounted for, the effects of IDS disappear, suggesting that they may be largely due to the slower rate of speech in IDS.

2.2.2. Vowels

We next examined the vowels. First, we asked if IDS altered vowel production. For this, we evaluated both the point vowels to replicate prior work and the interior vowels which have not been previously studied. Second, we asked if IDS changed the variance within vowel categories. Third, directly assessed the discriminability of the vowels in IDS and ADS in a way that was sensitive to changes in both mean cue-values and the variances. These analyses assessed the three formants (F1, F2, F3) along with pitch and WL. Our stimulus set contained nine vowels (Table 1). However, due the aforementioned difficulties in coding the

Table 3a
Results of a linear mixed effects model examining CV ratio for voiced sounds.

Factor	*B*	SE	p_{mcmc}	χ^2	df	*p*
Register	0.0002	0.0056	0.91	0.1	1	0.82
Place of artic						
Labial vs. other	−0.0205	0.0100	0.063	17.9	2	0.00013
Velar vs. other	0.0368	0.0100	0.0029			
Position						
1st vs. Other	0.0168	0.0026	<.0001	73.2	2	<.0001
2nd vs. Other	0.0211	0.0026	<.0001			
IDS × *Place*						
× Labial	−0.0042	0.0051	0.41	3.0	2	.22
× Velar	−0.0088	0.0051	0.088			
IDS × *Position*						
× P1	0.0007	0.0051	0.89	1.2	2	.54
× P2	0.0052	0.0051	0.31			

Table 3b
Results of a linear mixed effects model examining VOT for voiceless sounds.

Factor	*B*	SE	p_{mcmc}	χ^2	df	*p*
Register	0.0057	0.0038	0.125	2.1	1	0.15
Place of artic						
Labial vs. other	−0.0268	0.0145	0.094	4.2	2	0.12
Velar vs. other	−0.0026	0.0145	0.86			
Position						
1st vs. Other	0.0644	0.0046	<.0001	328.9	2	<.0001
2nd vs. Other	0.0867	0.0046	<.0001			
IDS × *Place*						
× Labial	0.0052	0.0092	0.58	0.3	2	0.85
× Velar	0.0015	0.0093	0.88			
IDS × *Position*						
× P1	0.0137	0.0092	0.14	2.4	2	.30
× P2	0.0033	0.0093	0.72			

formants of the high-back vowel, /u/, the *dune*/*tune* pair was dropped from this analysis, leaving 2376 tokens in the dataset. Of these, 327 were not codeable (M = 18.2/sub or 13.8%/sub; ADS: 124; IDS: 203, $T(17) = 4.3$, $p < .0001$), leaving 2049.

The first and second formants are typically considered the most important cues for vowels. Fig. 4 shows the mean location of each of the nine vowels in the F1 × F2 space; filled markers represent ADS vowels and open markers represent IDS. There is movement in the vowel space as a function of register, and the corner vowels show roughly the expansion reported by Kuhl et al. (1997). Our highest and furthest back vowel, /ou/, gets higher and backer (lower F1, lower F2), while the low-front vowel /æ/ both lowers and moves slightly to the front. While /i/ shows only a little movement, it is in the correct direction. However, at the same time, this movement is not in a uniformly enhancing direction. For example, /eI/ moves closer to /i/; /ɔ/ and /ʌ/ barely move at all, and /ɑr/ moves to the center of the space. Thus, considering more than the point vowels, movement due to IDS does not uniformly appear to be beneficial. Fig. 5 shows an analogous plot for the effect of position. Here, the 1st position is the strongest prosodically, and we see clear evidence for expansion as the 1st position for the three corner vowels is more extreme in IDS than ADS. Simultaneously, the pattern of movement for the interior vowels looks similar to the effects of IDS, with interior vowels moving fairly haphazardly in the stronger positions.

We did not expect a uniform shift in either formant across all vowels as a function of register (e.g., IDS was not predicted to shift F1 upward or downward for all vowels). Thus, we evaluated each vowel in a separate mixed effects model with register, voicing, position, and their interactions with register as fixed effects.[3] Participant was the only random effect.

The most relevant results are shown in Table 4 (left side). Bearing in mind the fairly weak power (each participant only had 6 tokens for some vowels), there was a difference between IDS and ADS in at last one dimension for four of the eight vowels (/eIr, æ, ɑr, ou/), and marginal evidence for another (/ɑI/). For the others, there was a register × VOT interaction with significant movement in at least one of the two voicing conditions. For example, /ir/ showed a significantly higher F2 in IDS (B = 85.5, SE = 36, $\chi^2(1) = 5.0$, $p = .025$) for voiceless sounds, but not for

[3] We could not evaluate place of articulation as a fixed effect or word-pair as a random-effect, since many of the vowels had only a single word-pair.

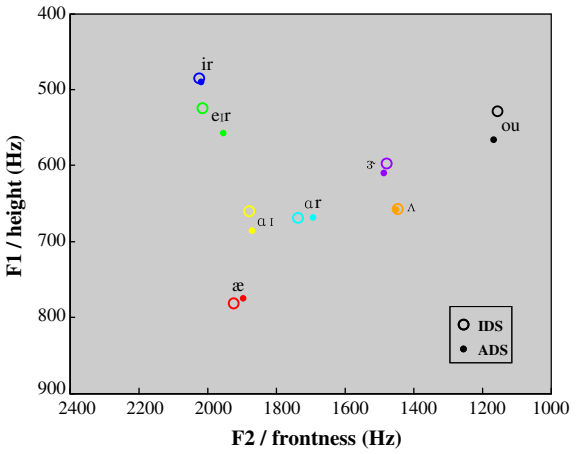

Fig. 4. Location in F1 × F2 space of each of the nine vowels in IDS (open circles) and ADS (filled points). Note the direction of the axes is reversed so that the orientation of the space corresponds roughly to height and frontness.

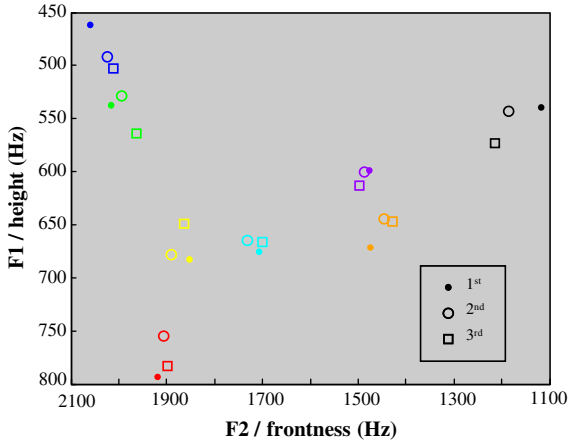

Fig. 5. Location in F1 × F2 space of each of the nine vowels as a function of prosodic position.

voiced sounds ($B = -28.0$, SE = 34, $\chi^2(1) = .7$, $p = .41$). Similarly, /ɜ˞/ had a marginally higher F3 in IDS for voiced sounds ($B = 120.2$, SE = 65, $\chi^2(1) = 3.3$, $p = .068$) but not voiceless ($B = -49.6$, SE = 89.4, $\chi^2(1) = .3$, $p = .57$); it was also marginally higher in IDS in position 2 ($B = 141.6$, SE = 75, $\chi^2(1) = .062$), but not in position 1 or 3 (1: $B = 49.8$, SE = 87, $\chi^2(1) = .34$, $p = .55$; $B = -101.1$, SE = 86.3, $\chi^2(1) = 1.4$, $p = .24$). This may have been due to differences in lip rounding (which changes F3) in the articulation of the /r/ between IDS and ADS. Thus, for seven of the eight vowels, IDS affected the formant frequencies of some or all of the words containing them. The one exception was the central vowel /ʌ/ where no effect of register was observed (perhaps predictably as there is nowhere for the vowel to move).

Thus, the effect of register in Fig. 4 is for the most part significant, even if it is not uniformly in a direction predicted by enhancement. However, it is also complex. Cues like F1 and F2 are obviously affected by vowel identity, but also by voicing (three vowels showed main effects of voicing on F1, and five showed effects of voicing on F2); and lip-rounding (which affects F3 and can derive from both rounded vowels like /ou/ and /u/ and consonants [/ɹ/]). IDS may affect the articulatory/acoustic forms of these phonemes, in the way it interacts with more basic phonetic properties like speech rate, jaw opening or coarticulation. This further undermines the case for any direct enhancement of vowels by IDS.

In contrast, the effect of position (Fig. 5, Table 4, right) was straightforward: every vowel except /ɜ˞/ showed an effect of position on at least one of the three formants. As Fig. 5 suggests, this took the form of expansion in the point vowels, and some more haphazard movement in the interior vowels. This suggests that the effect of prosodic strengthening may be somewhat similar to that of IDS, and like more consistent or robust.

2.2.3. Variance

While there was clearly movement in the average locations of each vowel, we next examined the variance. Fig. 6 shows the vowel space of all of the tokens in the dataset. Comparing to ADS (Fig. 6A), the vowel space in IDS (Fig. 6B) is more dispersed, and this results in substantial overlap between the categories. Of course, this conflates between- and within-talker variation. It is possible that individual talkers are less variable, but that they move in different directions creating more variation overall. Our focus is on the behavior of individuals – do caregivers create more or less variance in their utterances? However, for descriptive analyses of between participant variability see Online Supplement S3.

To assess the specifically within-participant variability, we computed the variability within each participant along F1, F2, and F3 for each vowel separately (and see Supplement S3 for vowel space plots that have been adjusted for talker variation). Fig. 7 displays the average variability along both F1 and F2 for each vowel as ellipses centered at the mean location of that vowel. It suggests that the variability of all of the vowels (except /ou/) increases in IDS, and in some cases this leads to substantial overlap. For example, /ir/ and /eɪr/ and /ɑɪ/ and /ɑr/ overlap minimally in ADS but substantially in IDS. This cannot be attributed to between-subject variability in IDS as the SDs represents the average of within-subject/within-vowel SD. These SDs were submitted to a linear mixed effects model with vowel and register as fixed effects, and participant as a random intercept (a similar model with random slopes of IDS by participant did not improve the fit). Results are shown in Table 5. For all three formants, there was a significant main effect of Register with more variability in IDS than ADS gain for all of these(Fig. 8). There was also a significant interaction for F2. Thus, we conducted additional analyses for F2 separated by vowel, but including only the fixed effect of register (Table 6). We found significant effects of register on F2 for /æ, eɪr, ir/ and a marginal effect for /ɑɪ/; again for all of these, IDS showed greater variability. Thus, it appears that across the board, IDS strongly increases the variability in F1 and F3, and increases or makes no change in F2. These challenge both the utility of the movement of

Table 4
Significance of the main effect of IDS or position for each vowel for each formant. $p > .2$ is indicated by a "–". Also shown are the interactions of IDS with Voicing or position (if significant). The number of tokens per participant is given in parenthesis in the first column.

	Effect of register			Effect of position		
	F1	F2	F3	F1	F2	F3
ir (6)	–	–	–	.0007	–	–
IDS × Voicing		.011				
eɪr (6)	.0087	.00061	.0015	.0048	.016	–
ɑɪ (6)	.07	–	.17	.037	–	–
æ (12)	–	.068	.016	.00013	.19	–
ɑr (12)	–	.027	.018	.15	.022	–
IDS × Voicing	.023	.048				
ɝ (6)	.11	–	–	–	–	–
IDS × Voicing	.065	.086	.078			
IDS × Position			.033			
ʌ (6)	–	–	.11	.015	.042	.006
IDS × Voicing		.081				
ou (12)	.0011	–	<.0001	.00021	.00067	.00015

the vowels in IDS for perceptual learning and the reliability of it across infants.

2.3. Identification: weighing means and variances

Our final analyses asked directly about the quantity of information in the speech signal to support vowel categorization and learning. The increased variance in IDS suggests that it may not enhance vowel category learning. However, at the same time, we also observed complex changes in the mean vowels that could enhance some contrasts. To resolve this discrepancy, we used logistic regression as a model of ideal categorization performance given the structure of the information in this dataset (Cole et al., 2010; McMurray & Jongman, 2011; Werker et al., 2007). Logistic regression maps a set of cues onto the corresponding categories by computing the optimal linear weighting of cues to separate the categories. This is a form of supervised learning, which is more powerful than what infants are likely to employ. However, in this way, it serves as an ideal observer, asking how well an infant could conceivably learn speech categories from the input provided by IDS.

While traditional logistic regression outputs a binary choice, we used multinomial logistic regression which can predict any number of categories. These were trained to categorize each token as one of the eight vowels based on the raw F1, F2, and F3 measurements, along with WL and pitch. Models were trained separately on IDS and ADS to evaluate each independently. To avoid over fitting the data we randomly selected 85% of the data to use as training data, and then tested the model on the remaining 15%. This was repeated 1000 times to estimate the proportion correct (on the held-out data) as well as its variability across sampling (we computed confidence intervals, as the range at which 95% of the proportions fell). To convert the probabilities from the logistic regression to a response, we used the "discrete rule" of McMurray and Jongman (2011) in which the most probable choice is the model's discrete response.

Overall both models performed well averaging 62.2% correct. Models trained on ADS (Raw cues: $M = 64.7\%$, CI = 57.8–71.6) performed slightly better than those trained on IDS (Raw cues: 59.7%, CI = 52.3–67.3). Fig. 9A shows the performance broken down by specific vowel for both classes of models. The interior vowels like /ɑɪ, ɝ, ʌ/ clearly showed much lower performance than the corner vowels. However, across all the vowels, there was either a benefit for ADS over IDS or the two registers were equal.

One concern here is that the raw measurements conflate both within- and between-talker variation, while our focus is on within-talker variation as the most important aspect of the infants' environment. To eliminate the between-category variation, we subtracted mean cue-values for each participant (across all their vowels) from their raw formant frequencies, pitches and durations. While we use this here as a statistical technique to approximately align talkers, it is useful to point out that this has been

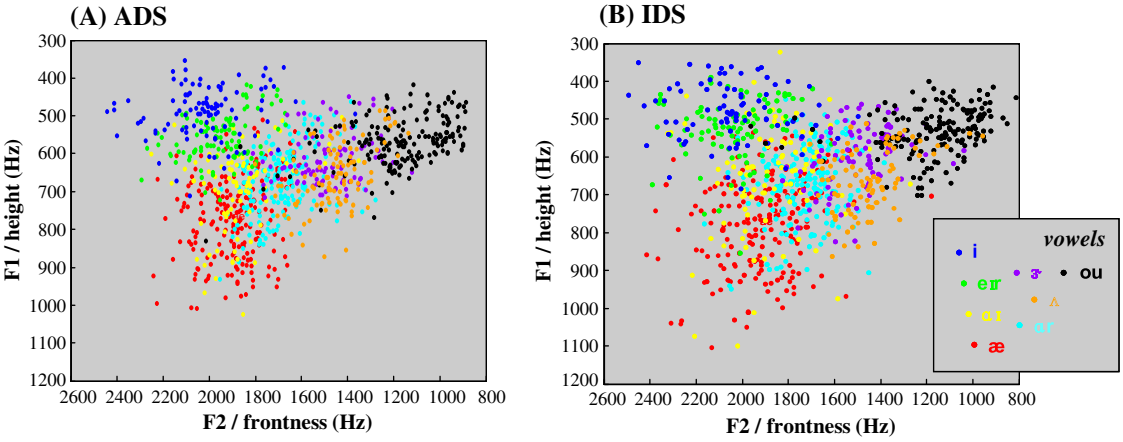

Fig. 6. Scatter plots showing each individual token in F1 × F2 space. (A) Raw measurements for ADS and (B) IDS. For analogous plots after talker related variance has been removed see Online Supplement S3.

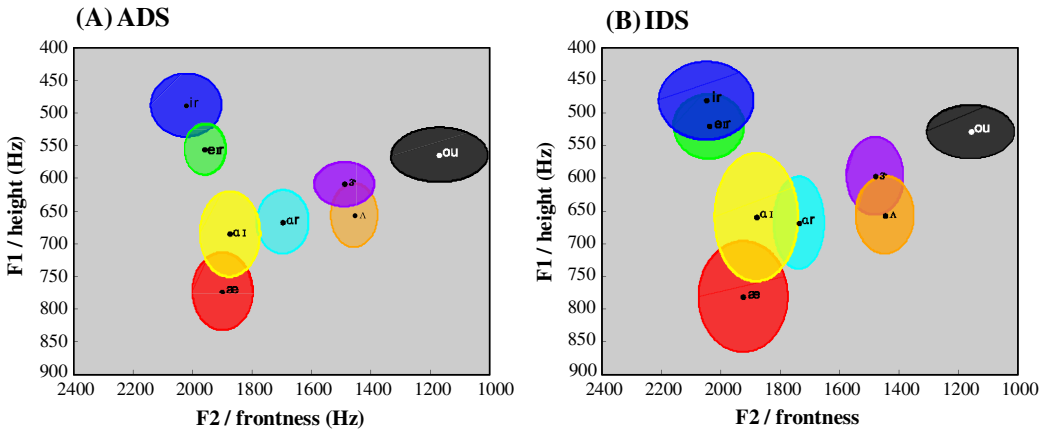

Fig. 7. Variability among vowels. Each ellipse is centered at the grand mean of F1 and F2 for each vowel; the height corresponds to one SD in F1 and the width corresponds to one SD of F2. SDs are the average of within-subject SDs. (A) For ADS; (B) for IDS.

Table 5
Results of a series of mixed effects models analyzing the SD of F1, F2 and F3 as a function of register and vowel.

	χ^2	df	*p*
F1			
Register	24.1	1	<.0001
Vowel	60.1	7	<.0001
Register × Vowel	12.7	7	0.079
F2			
Register	9.7	1	0.0018
Vowel	52.2	7	<.0001
Register × Vowel	18.3	7	0.011
F3			
Register	4.1	1	0.042
Vowel	66.5	7	<.0001
Register × Vowel	11.5	7	0.12

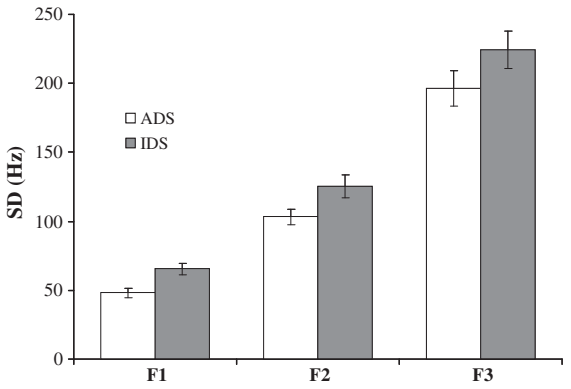

Fig. 8. Average within-vowel/within-talker standard deviation of each of the three formants. Error bars represent SEM.

shown to be a good approximation of how adult listeners compensate for talker variation (Cole et al., 2010; McMurray & Jongman, 2011). When we replicated the logistic analysis using residualized data, this model showed even better performance than the raw cue model (*M* = 71.4%).

Table 6
Individual effect of register on variability in F2 for each vowel.

Vowel	$\chi^2(1)$	*p*	IDS–ADS
æ	10.7	0.0011	51.3
ɑr	0.0	–	−0.7
ɑɪ	2.9	0.088	41.1
eɪr	14.9	0.0001	52.6
ɜ˞	0.2	–	−6.7
ir	3.9	0.048	41.1
ou	0.8	–	−17.2
ʌ	1.2	–	17.8

However, it also continued to show an advantage for ADS (*M* = 72.9%, CI = 66.0–79.2) over IDS (*M* = 69.9%, CI = 63.0–76.7). Fig. 9B shows performance for each vowel and again, models trained on ADS outperformed IDS for every vowel except /ɑɪ/ and /ʌ/ (and for these the vowels the difference was small). Thus, there is not any evidence to support the notion that IDS is superior to ADS for learning speech categories under a statistical learning model. In fact under some assumptions about this learning mechanism, ADS may be superior, which fits with Cristia and Seidl's (2013) finding that for some interior vowels IDS may partially collapse the distinction.

3. Discussion

This experiment used measurements of VOT and vowel formants to more precisely characterize the changes in phonetic cues that derive from IDS. Specifically, we asked whether these changes are intended to enhance statistical learning or are unintended consequences of supra-segmental properties of IDS. To do this, we measured vowel length to control for rate changes on VOT. We also controlled for word choice and prosodic frame differences associated with IDS by measuring the same caregivers reading the same sentences—the only difference was the audience.

Across our analyses a striking picture emerged. Our analysis of VOT suggested clear effects of IDS. However,

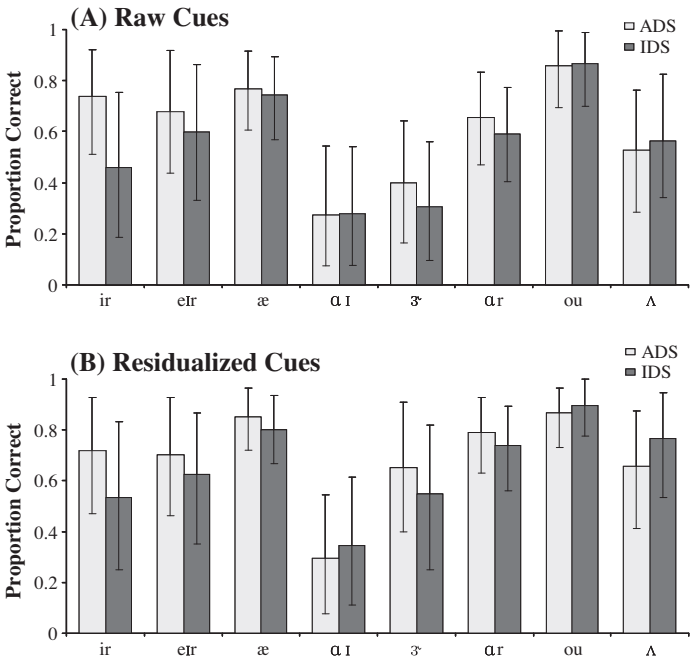

Fig. 9. Performance of the logistic regression classifiers as a function of vowel and register. (A) Trained on raw cues. (B) Trained on residualized cues. Error bars represent 95% confidence intervals.

IDS lengthened both voiced and voiceless sounds, rather than increasing the contrast between them, and when we accounted for speaking rate, the effect of IDS disappeared. This suggests that IDS-induced changes to voicing may have little to do with enhancement of the contrast, and instead are a by-product of slower speech. Our analysis of the vowel cues (F1, F2, F3, pitch, and WL) showed a similar pattern. Again, we found significant movement of the vowels in IDS, but again it was not uniformly in a beneficial direction. While the corner vowels did expand, as seen in prior studies (Kuhl et al., 1997), the interior vowels moved more idiosyncratically (see also, Cristia & Seidl, 2013). If anything, the effect of prosodic position was larger and more systematic than that of register. There was also substantially more variation in IDS for all three formants and our logistic regression analyses suggest that this increase in variance may outweigh any benefit of changes in the means. This suggests that the modification inherent in IDS do not serve to enhance phonetic discrimination or development via statistical learning mechanisms.

The relationship of our results on vowels to prior studies of IDS (Cristia & Seidl, 2013; Kuhl et al., 1997; Liu et al., 2003) is clear. While Kuhl et al., and Liu et al., both report enhancement, they also looked only at the point vowels (not the interior vowels) and did not examine variability. When we looked at both we reached a different conclusion. Similarly, our results with the interior vowels strongly parallel Christia and Seidl. In contrast, the relationship of our results on VOT to prior studies is less clear. Our results clearly accord with those of Englund (2005), and they match three prior studies of VOT that manipulated speaking rate (Allen & Miller, 1999; Beckman, Helgason, McMurray, & Ringen, 2011; Kessinger & Blumstein, 1998), reinforcing our conclusions regarding the role of speaking rate. However, they differ from Sundberg and Lacerda (1999) who found that both voiced and voiceless categories *shorten* in IDS. While at the highest level, this too conflicts with enhancement, it is unclear why they found shortening. It is possible that the use of free conversation in their study created inadvertent confounds such as differences in prosodic position, or word choice that could have given rise to such effects. One likely factor is the number of syllables – the authors report coding both word initial and word-medial VOTs (which are typically shorter than word initial VOTs). If one register had a different proportion of mono-syllabic or multi-syllabic words this could lead to systematic effects. However, the authors do not report much detail about the specific lexical items, their prosodic positions, number of syllables, or their speaking rate.

In our more controlled approach, however, the results are systematic and suggest that the changes to phonetic cues in IDS do not uniformly benefit statistical learning. So where do these changes derive from? The first thing that we must account for is the fact that caregivers are clearly modifying supra-segmental properties of the signal like speaking rate, prosody, and affect. These modifications have consequences for segmental properties. Our data on VOT directly implicate speaking rate in this regard. Additionally, the movement in the vowels looks similar to what is observed during deliberately slow speech (Van Son & Pols, 1990), which also shows an almost haphazard pattern of movement. In this case, rather than reflecting a deliberate enhancement of phonetic cues, a slower speaking rate causes the jaw to be more open, which in turn changes tongue position independently of the intended vowel articulation. Similarly, changes in speech rate can also affect the

timing of coarticulation or dipthongs in a nonlinear way which is not directly relevant for phonetic categories.

However, speaking rate is not the only relevant supra-segmental factor. We found effects of prosodic position on all of our cues, and Fig. 5 suggests that sentence-initial words (the prosodically strongest) show an enhancement effect similar to, if not bigger than, what has been reported for IDS. As IDS commonly affects both the prosodic strength of key words (words may be accented to point out their novelty), and the structure of the sentences that are being used, such factors may mimic the vowel expansion effect. However, as our work shows, such effects do not uniformly enhance segmental cues and can also cause more idiosyncratic changes in the middle of the vowel space.

Beyond the present study, other research suggests an important role of affect in accounting for vowel-space changes in IDS. Benders (2013) suggests that increased smiling in IDS affects lip position in a way that can alter phonetic cues for fricatives (and likely other phonemes as well). This may explain the dissociation observed by Burnham, Kitamura, and Vollmer-Conna (2002) between IDS and pet-directed speech which has the higher pitch and variability of IDS, but no vowel space expansion (although again, prosodic position was not controlled in this study). They also found significant differences in affect between pet-directed and infant directed speech, and ongoing work by Panneton (unpublished) suggests that *puppy*-directed speech (which has much higher affect) also shows the hyper-articulation of IDS.

Given the constellation of supra-segmental factors involved in IDS, it seems likely that any coarse grained changes in speech register may create broad-based changes in phonetic cues. However, these changes may or may not result in any direct enhancement of phonetic distinctions, and cannot be clearly interpreted as evidence of a motivation (implicit or explicit) to enhance infants' abilities to learn speech categories. Given this, what are caregivers intending (implicitly of course) to modify in the signal?

Perhaps the simplest explanation is that caregivers' desire for clarity is not at the level of fine-grained articulator position. Rather, their goal is to emphasize broader prosodic and supra-segmental features of speech. These are perhaps more relevant for transmitting the communicative intent, even if emphasizing them comes at the expense of individual phonemes (see, Sundberg & Lacerda, 1999, for a similar argument). Given the role of IDS in initiating and maintaining attention, modifying arousal, and communicating affect, it would not be surprising if this was the primary or even sole motivation. This is consistent with our data as most of our effects appeared similar to prosodic or rate effects. This is obviously relevant for learning other aspects of language, but it is also possible that such changes improve speech category learning. That is, even if the specific segmental changes associated with IDS do not improve infants' ability to learn speech categories, the broader set of prosodic and pragmatic factors associated with IDS could facilitate learning by modulating arousal, maintaining attention, or presenting shorter chunks of speech. Indeed, there is evidence that the closely related "clear speech" register improves intelligibility somewhat independently of its effect on specific phonetic cues (Ferguson & Kewley-Port, 2007), perhaps by making phonetic cues more detectable, by changing which cues are available or important, or by giving more time for processing. IDS may also offer such benefits, which could conceivably affect learning, though at this point there is not any evidence for such claims.

An alternative is that parents are attempting to speak more clearly at the level of segmental cues, but they cannot exert specific control over phonetic cues like VOT or formant frequencies—they can only really control gross properties like pitch or speaking rate. Indeed, this fits with a number of recent developmental studies showing less than optimal control over articulators in situations in which clear speech and feedback may be important. Julien and Munson (2012), for example, found that when people corrected a child's mispronounced fricatives, their own fricative productions were longer, but the spectral content did not reflect hyper-articulation. Thus, even when trying to produce a clearer sound, they only made it longer. Perhaps more strikingly, Lam and Kitimura (2010) compared the vowel space of a mother when talking to twins, one of which was hearing impaired. They found the vowel space *contracted* for the hearing impaired twin. This was subsequently replicated with mothers of normal-hearing infants interacting over closed-circuit television. When infants could not hear the mother well (and provided differential feedback), the mother's vowel space contracted (Lam & Kitimura, 2011). In both cases (hearing impairment and corrective feedback), mothers most likely need to phonetically enhance cues, and yet we see that, as in our data here, they only have control over much coarser grained properties.

Work on so-called "clear speech" offers a useful analogue to these issues, but also offers mixed evidence as to whether people have fine-grained control over segmental cues. While there are a large number of studies supporting vowel space expansion when talkers are trying to speak clearly (Ferguson & Kewley-Port, 2002; Moon & Lindblom, 1994; Smiljanic & Bradlow, 2005; and see, Smiljanic & Bradlow, 2009, for a review); at the same time, the magnitude of changes in formant frequencies are not always associated with differences in intelligibility (Ferguson & Kewley-Port, 2007; Neel, 2008). Similarly, while VOT changes in clear or slow speech (e.g., Smiljanic & Bradlow, 2008), such changes may not be intended for the listener. Beckman et al. (2011) showed that in Central Standard Swedish, where voicing is over-specified[4] and there is literally no ambiguity, when talking slowly talkers lengthen aspirated and pre-voiced VOTs anyways—enhancing the voicing contrast when it is not needed. Much like we observed here, this implies a locus closer to speaking rate dynamics than audience design. However, at the same time, there is evidence for fairly precise articulator enhancement in some situations. Maniwa, Jongman, and Wade (2009) showed that talkers could adjust specific properties of fric-

[4] CS Swedish uses both pre-voiced and aspirated stops with no short-lag stops in the middle.

atives (e.g., place of articulation or voicing) in response to specific types of recognition errors along these dimensions (although the effect was highly variable). Thus, the clear speech literature offers some support for the idea that listeners may have fine-grained articulator control; and if this is the case, this would imply that caregivers speaking in IDS are simply focusing on other (supra-segmental) aspects of the signal. However, given the inconsistencies in this body of work, we cannot rule out that caregivers are attempting phonetic enhancement and are simply not effective at achieving it.

In practice, however, no matter what the motivation, caregivers are clearly modulating the statistical environment in which infants learn, even if the precise segmental properties are not "designed" for the child's developmental needs. However, the debate around this issue frames the child's "needs" largely in terms of long-term developmental outcomes. Yet this is not the only relevant factor. One important consideration is the timescale over which the motivation to use IDS operates. We must consider both the real-time processes of language use (e.g., the immediate communicative and social needs of the parent and child), and the longer timescale processes of language development (cf., McMurray, Horst, & Samuelson, 2012).

The tempting conclusion from prior work is that parents may be using IDS to enhance development—their primary concern is the long-term developmental outcomes. However, this study, along the prior work we discussed (Benders, 2013; Englund, 2005; Lam & Kitimura, 2010, 2011) suggests the changes in phonetic cues afforded by IDS are not well suited to enhancing the development of phonetic categorization. In contrast, at the more immediate timescale, IDS may play more valuable roles – regulating infants' affect and attention (Smith & Trainor, 2008), highlighting key words or phrases (Fernald & Mazzie, 1991), or even as a sociolinguistic marker to other parents. Some of these things may have long-term developmental benefits, but the more proximal cause may be the real-time demands of communication. These real-time demands on parents require more coarse-grained changes like affect, pitch and timing, rather than fine-grained changes in articulation. Even when caregivers' real-time goals are relevant to phonetic issues, caregivers may be more motivated to ensure that they are understood, than to ensure that the child learns anything about language. Indeed this would fit with work by Song, Demuth, and Morgan (2010) showing that vowel hyper-articulation can improve older infants' (19 m.o.) recognition of familiar words.

The idea that IDS may be uniquely tuned to the real-time interactions and communicative needs of parents and children also fits with work by Smith and Trainor (2008) showing that the dynamics of interactions can shape parents' use of IDS. Again this suggests caregivers are responding more to real-time communicative demands than to a motivation for enhancing long-term outcomes. Given this, it is perhaps optimistic to assume that caregivers are simultaneously responding to both the immediate and developmental goals. Moreover, even caregivers wanted to simultaneously enhance segmental cues, this may be too difficult, as the real time changes affect, pitch, and timing indirectly affect segmental cues like VOT.

Thus, IDS does not appear to differentially support either the learning of phonetic categories, or their discrimination. Nonetheless, our work suggests that IDS modifies the distribution of phonetic cues in ways that have consequences for statistical learning. Given complex interactions between things like speaking rate, prosody, and segmental cues the early language environment may present considerable complexity to children learning phonetic categories.

Acknowledgements

The authors would like to thank Robin Panneton and Joy Wu for helpful feedback throughout this project, Susan Levasseur for helpful ideas as we were getting this project off the ground, and Kelsey Wiggs, Kallie Tebbe, Tiffany Born and Jessica Powell for help with phonetic coding. This work supported by NIH Grant R01 DC008089 to B.M., and a McNair Scholars Summer Fellowship to D.G. (mentored by K.K.).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2013.07.015.

References

Allen, J. S., & Miller, J. L. (1999). Effects of syllable-initial voicing and speaking rate on the temporal characteristics of monosyllabic words. *Journal of the Acoustical Society of America, 106*, 2031–2039.

Andruski, J. E., Blumstein, S. E., & Burton, M. W. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52*, 163–187.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes.

Beckman, J., Helgason, P., McMurray, B., & Ringen, C. (2011). Rate effects on Swedish VOT: Evidence for phonological overspecification. *Journal of Phonetics, 39*, 39–49.

Benders, T. (2013). *Nature's distributional-learning experiment*. Ph.D., The University of Amsterdam, Amsterdam, The Netherlands.

Bion, R. A. H., Miyazawa, K., Kikuchi, H., & Mazuka, R. (2013). Learning phonemic vowel length from naturalistic recordings of Japanese infant-directed speech. *PLoS One, 8*(2), e51594. http://dx.doi.org/10.1371/journal.pone.0051594.

Boersma, P., & Weenink, D. (2009). *Praat: Doing phonetics by computer (version version 5.1.05)*. <http://www.praat.org/>.

Boucher, V. J. (2002). Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception & Psychophysics, 64*, 121–130.

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's new, pussycat? On talking to babies and animals. *Science, 296*(5572), 1435. http://dx.doi.org/10.1126/science.1069587.

Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a, i/ in English. *Journal of the Acoustical Society of America, 117*(6), 3867–3878.

Cole, J. S., Linebaugh, G., Munson, C., & McMurray, B. (2010). Unmasking the acoustic effects of vowel-to-vowel coarticulation: A statistical modeling approach. *Journal of Phonetics, 38*(2), 167–184.

Cooper, R. P., & Aslin, R. N. (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*(5), 1584–1595. http://dx.doi.org/10.2307/1130766.

Cristia, A., & Seidl, A. (2013). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language, FirstView, 1–22*. http://dx.doi.org/10.1017/S0305000912000669.

de Boer, B., & Kuhl, P. K. (2003). Investigating the role of infant-directed speech with a computer model. *Auditory Research Letters On-Line (ARLO), 4*, 129–134.
Eilers, R., & Minifie, F. (1975). Fricative discrimination in early infancy. *Journal of Speech and Hearing Research, 18*(1), 158–167.
Eilers, R., Wilson, W., & Moore, J. (1977). Developmental changes in speech discrimination in infants. *Perception & Psychophysics, 16*, 513–521.
Englund, K. T. (2005). Voice onset time in infant directed speech over the first six months. *First Language, 25*(2), 219–234. http://dx.doi.org/10.1177/0142723705050286.
Ferguson, S. H., & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 112*, 259–271.
Ferguson, S. H., & Kewley-Port, D. (2007). Talker differences in clear and conversational speech: Acoustic characteristics of vowels. *Journal of Speech, Language and Hearing Research, 50*(5), 1241–1255. http://dx.doi.org/10.1044/1092-4388(2007/087.
Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior & Development, 8*(2), 181–195.
Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*(2), 209–221.
Fernald, A., Taeschner, T., Dunn, J., Papousek, M., de Boysson-Bardies, B., & Fukui, I. (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language, 16*(3), 477–501.
Galle, M., & McMurray, B. (submitted for publication). The development of voicing categories: A meta-analysis of 40 years of infant research. *Psychonomic Bulletin and Review*.
Guenther, F., & Gjaja, M. (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America, 100*, 1111–1112.
Hayashi, A., Tamekawa, Y., & Kiritani, S. (2001). Developmental change in auditory preferences for speech stimuli in Japanese infants. *Journal of Speech, Language and Hearing Research, 44*(6), 1189–1200. http://dx.doi.org/10.1044/1092-4388(2001/092.
Hillenbrand, J. M., Getty, L., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*(5), 3099–3111.
Julien, H., & Munson, B. (2012). Modifying speech to children based on their perceived phonetic accuracy. *Journal of Speech Language and Hearing Research, 55*(6), 1836–1849.
Kessinger, R. H., & Blumstein, S. E. (1998). Effects of speaking rate on voice-onset time in Thai, French, & English. *Journal of Phonetics, 25*, 143–168.
Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, 50.
Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., et al. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*(5326), 684–686. http://dx.doi.org/10.1126/science.277.5326.684.
Kuhl, P. K., Stevens, E. H. A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*, F13–F21.
Lam, C., & Kitimura, C. (2010). Maternal interactions with a hearing and hearing impaired twin: Similarities and differences in speech input, interaction, quality and word production. *Journal of Speech Language and Hearing Research, 53*, 543–555.
Lam, C., & Kitimura, C. (2011). Mommy, speak clearly: Induced hearing loss shapes vowel hyperarticulation. *Developmental Science, 15*(2), 212–221.
Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word, 20*, 384–422.
Liu, H.-M., Kuhl, P. K., & Tsao, F.-M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), F1–F10. http://dx.doi.org/10.1111/1467-7687.00275.
Magloire, J., & Green, K. P. (1999). A cross-language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica, 56*(3–4), 158–185.
Maniwa, K., Jongman, A., & Wade, T. (2009). Acoustic characteristics of clearly spoken English fricatives. *Journal of the Acoustical Society of America, 125*(6), 3962–3973.
Maye, J., Weiss, D. J., & Aslin, R. N. (2008). Statistical phonetic learning in infants: Facilitation and feature generalization. *Developmental Science, 11*(1), 122–134. http://dx.doi.org/10.1111/j.1467-7687.2007.00653.x.
Maye, J., Werker, J. F., & Gerken, L. (2003). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82*, 101–111.
McCandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & McClelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2*, 89–108.
McMurray, B., & Aslin, R. N. (2005). Infants are sensitive to within-category variation in speech perception. *Cognition, 95*(2), B15–B26.
McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology, Human Perception and Performance, 34*(6), 1609–1631.
McMurray, B., Aslin, R. N., & Toscano, J. C. (2009). Statistical learning of phonetic categories: Insights from a computational approach. *Developmental Science, 12*(3), 369–379.
McMurray, B., & Farris-Trimble, A. (2012). Emergent information-level coupling between perception and production. In A. Cohn, C. Fougeron, & M. Huffman (Eds.), *The Oxford handbook of laboratory phonology*. Oxford, UK: The Oxford University Press.
McMurray, B., Horst, J. S., & Samuelson, L. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*.
McMurray, B., Horst, J. S., Toscano, J. C., & Samuelson, L. (2009). Towards an integration of connectionist learning and dynamical systems processing: Case studies in speech and lexical development. In J. Spencer, M. Thomas, & J. L. McClelland (Eds.), *Towards an integration of connectionist learning and dynamical systems processing: Case studies in speech and lexical development*. London: Oxford University Press.
McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? Harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological Review, 118*(2), 219–246.
McMurray, B., & Spivey, M. J. (2000). The categorical perception of consonants: The interaction of learning and processing. *Proceedings of the Chicago Linguistics Society, 34*(2), 205–220.
McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*(2), B33–B42.
Miller, J. L. (1997). Internal structure of phonetic categories. *Language and Cognitive Processes, 12*, 865–869.
Miller, J. L., & Eimas, P. D. (1996). Internal structure of voicing categories in early infancy. *Perception & Psychophysics, 58*(8), 1157–1167.
Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica, 43*, 106–115.
Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46*(6), 505–512.
Moon, S. J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustical Society of America, 96*, 40–55.
Munson, B., & Solomon, N. P. (2004). The effect of phonological neighborhood density on vowel articulation. *Journal of Speech Language and Hearing Research, 47*, 1048–1058.
Neel, A. T. (2008). Vowel space characteristics and vowel identification accuracy. *Journal of Speech, Language and Hearing Research, 51*(3), 574–585. http://dx.doi.org/10.1044/1092-4388(2008/041.
Newman, R., & Hussain, I. (2006). Changes in preference for infant-directed speech in low and moderate noise by 4.5- to 13-month-olds. *Infancy, 10*(1), 61–76.
Pegg, J. E., Werker, J. F., & McLeod, P. J. (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior & Development, 15*(3), 325–345.
Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175–184.
Pierrehumbert, J. B. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech, 46*, 115–154.
Pind, J. (1995). Speaking rate, voice-onset time, and quantity: The search for higherorder invariants for two Icelandic speech cues. *Perception & Psychophysics, 57*, 291–304.
Port, R. F., & Dalby, J. (1982). Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics, 32*(2), 141–152.
Smiljanic, R., & Bradlow, A. R. (2005). Production and perception of clear speech in Croatian and English. *Journal of the Acoustical Society of America, 118*(3), 1677–1688.
Smiljanic, R., & Bradlow, A. R. (2008). Stability of temporal contrasts across speaking styles in English and Croatian. *Journal of Phonetics, 36*(1), 91–113.

Smiljanic, R., & Bradlow, A. R. (2009). Speaking and hearing clearly: Talker and listener factors in speaking style changes. *Language and Linguistics Compass, 3*(1), 236–264.

Smith, N. A., & Trainor, L. J. (2008). Infant-directed speech is modulated by infant feedback. *Infancy, 13*(4), 410–420. http://dx.doi.org/10.1080/15250000802188719.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*(4), 501–532.

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *Journal of the Acoustical Society of America, 128*(1), 389–400.

Student (1908). The probable error of a mean. *Biometrika, 6*, 1–25.

Sundberg, U., & Lacerda, F. (1999). Voice onset time in speech to infants and adults. *Phonetica, 56*(3–4), 186–199.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: A statistical approach to cue weighting and combination in speech perception. *Cognitive Science, 34*(3), 436–464.

Toscano, J. C., McMurray, B., Dennhardt, J., & Luck, S. (2010). Continuous perception and graded categorization electrophysiological evidence for a linear relationship between the acoustic signal and perceptual encoding of speech. *Psychological Science, 21*(10), 1532–1540.

Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics, 62*(6), 1297–1311.

Vallabha, G. K., McClelland, J. L., Pons, F., Werker, J. F., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *Proceedings of the National Academy of Sciences of the United States of America, 104*, 13273–13278.

Van Son, R., & Pols, L. (1990). Formant frequencies of Dutch vowels in a text, read at normal and fast rate. *Journal of the Acoustical Society of America, 88*(4), 1683–1693.

Volaitis, L., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America, 92*(2), 723–735.

Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197–234. http://dx.doi.org/10.1080/15475441.2005.9684216.

Werker, J. F., & Lalonde, C. E. (1988). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology, 24*(5), 672–683. http://dx.doi.org/10.1037/0012-1649.24.5.672.

Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., & Amano, S. (2007). Infant-directed speech supports phonetic category learning in English and Japanese. *Cognition, 103*(1), 147–162. http://dx.doi.org/10.1016/j.cognition.2006.03.006.

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 7*, 49–63.

Yoshida, K. A., Pons, F., Maye, J., & Werker, J. F. (2010). Distributional phonetic learning at 10 months of age. *Infancy, 15*(4), 420–433. http://dx.doi.org/10.1111/j.1532-7078.2009.00024.x.

Zangl, R., & Mills, D. L. (2007). Increased brain activity to infant-directed speech in 6-and 13-month-old infants. *Infancy, 11*(1), 31–62.