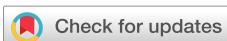


JULY 16 2010

## Effects of the acoustic properties of infant-directed speech on infant word recognition<sup>a)</sup>

Jae Yung Song; Katherine Demuth; James Morgan



*J. Acoust. Soc. Am.* 128, 389–400 (2010)  
<https://doi.org/10.1121/1.3419786>



### Articles You May Be Interested In

How acoustic cues in infant-directed speech facilitate 19-month-olds' word recognition: Evidence from time course analysis

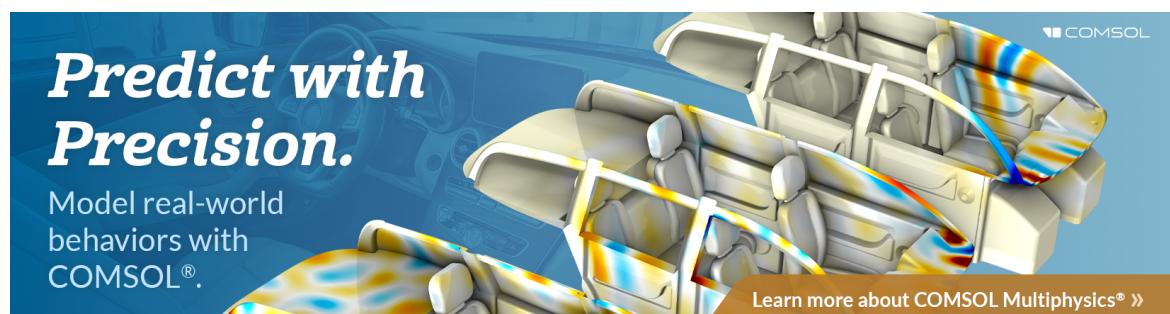
*J. Acoust. Soc. Am.* (November 2007)

Infant-directed speech reduces English-learning infants' preference for trochaic words

*J. Acoust. Soc. Am.* (December 2016)

Fine-grained variation in caregivers' /s/ predicts their infants' /s/ category

*J. Acoust. Soc. Am.* (May 2011)



**Predict with Precision.**  
Model real-world behaviors with COMSOL®.

Learn more about COMSOL Multiphysics® »

COMSOL

# Effects of the acoustic properties of infant-directed speech on infant word recognition<sup>a)</sup>

Jae Yung Song<sup>b)</sup>

Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island 02912

Katherine Demuth

Department of Linguistics, Macquarie University, New South Wales 2109, Australia

James Morgan

Department of Cognitive and Linguistic Sciences, Brown University, Providence, Rhode Island 02912

(Received 23 December 2008; revised 7 April 2010; accepted 8 April 2010)

A number of studies have examined the acoustic differences between infant-directed speech (IDS) and adult-directed speech, suggesting that the exaggerated acoustic properties of IDS might facilitate infants' language development. However, there has been little empirical investigation of the acoustic properties that infants use for word learning. The goal of this study was thus to examine how 19-month-olds' word recognition is affected by three acoustic properties of IDS: slow speaking rate, vowel hyper-articulation, and wide pitch range. Using the intermodal preferential looking procedure, infants were exposed to half of the test stimuli (e.g., *Where's the book?*) in typical IDS style. The other half of the stimuli were digitally altered to remove one of the three properties under investigation. After the target word (e.g., *book*) was spoken, infants' gaze toward target and distractor referents was measured frame by frame to examine the time course of word recognition. The results showed that slow speaking rate and vowel hyper-articulation significantly improved infants' ability to recognize words, whereas wide pitch range did not. These findings suggest that 19-month-olds' word recognition may be affected only by the linguistically relevant acoustic properties in IDS. © 2010 Acoustical Society of America. [DOI: 10.1121/1.3419786]

PACS number(s): 43.71.Bp, 43.71.Ft [RSN]

Pages: 389–400

## I. INTRODUCTION

When talking to babies, parents typically use a unique speech register characterized by a number of parameters including raised, exaggerated pitch, and slower speaking rate. Even young siblings and other adults who have had no experience with infants spontaneously change their speech register when interacting with an infant (Dunn and Kendrick, 1982; Jacobson *et al.*, 1983). This manner of talking is generally called motherese, baby talk, or infant-directed speech (IDS), as compared to adult-directed speech (ADS).

Acoustic modifications in IDS are widely found across various languages. For example, in a study with French, Italian, German, Japanese, British English, and American English-speaking parents, Fernald *et al.* (1989) showed that all parents commonly used higher mean fundamental frequency (F0), greater F0-variability, shorter utterances, and longer pauses during interactions with their 10-to 14-month-olds than with adults. Chinese-speaking mothers were also found to use higher F0 and wider pitch range when talking to their 2-month-old infants than to adults (Grieser and Kuhl, 1988). Thus, IDS in Mandarin Chinese, which is a tonal language, showed the same patterns of acoustic modification

as a nontonal language such as English (see Kitamura *et al.*, 2001, for similar results in Thai). These findings suggest that the exaggerated acoustic characteristics of IDS are universal, although a few studies have reported language-specific variation as well: Quiche-speaking mothers were found not to use a raised, high pitch when speaking to their children; high pitch may be reserved for addressing social superiors in this culture (Bernstein Ratner and Pye, 1984; Ingram, 1995).

Given that IDS is commonly available for infant listening, it is important to determine how experience with such speech influences aspects of infants' development. Although researchers have proposed various hypotheses about the possible functions of IDS (e.g., Colombo *et al.*, 1995), there is general agreement on three functions of IDS (Cooper *et al.*, 1997; Grieser and Kuhl, 1988; Singh *et al.*, 2002). First, IDS may attract and maintain infants' attention. Second, it may communicate positive emotion or affect between a caregiver and an infant. Third, it may facilitate language acquisition. Note that these three potential functions may coexist during infant development, with attentional and affective functions dominating during early infancy, and linguistic functions gaining importance as infants progress in the language learning process.

Research on the linguistic function of IDS has been primarily descriptive in nature, reporting on the acoustic characteristics of IDS that could potentially benefit language development. A number of studies have found that syntactic boundaries are acoustically more salient in IDS than in ADS

<sup>a)</sup>Portions of this work were presented at the 154th Meeting of the Acoustical Society of America in New Orleans, LA (2007) and the 32nd Boston University Conference on Language Development in Boston, MA (2007).

<sup>b)</sup>Author to whom correspondence should be addressed. Electronic mail: jae\_yung\_song@brown.edu

(Bernstein Ratner, 1986; Morgan, 1986). For instance, Bernstein Ratner (1986) found that the degree of clause-final vowel lengthening was almost doubled in speech to 9–13-month-old infants as compared to ADS (100.74 vs 52.16 ms). Furthermore, Kemler Nelson *et al.* (1989) showed that infants were able to distinguish pauses inserted at clausal boundaries from pauses inserted at clause-internal locations only when they heard IDS. These findings suggest that exaggerated acoustic markers for syntactic boundaries in IDS may help infants identify syntactic units in the speech stream. Furthermore, unlike AD casual speech, in which phonemes are frequently reduced or deleted (Johnson, 2004), the vowels and consonants in IDS tend to be elongated and carefully articulated (Bernstein Ratner, 1984). Studies have also shown that phonemes in IDS are modified in a way that makes them acoustically more distinct from one another (Kuhl *et al.*, 1997; Malsheen, 1980; Werker *et al.*, 2007). This might aid infants in acquiring the phonological and/or phonetic categories in their native language by providing more detailed linguistic information.

These studies suggest various properties of IDS that might help infants acquire language. However, little is known about whether *infants* are sensitive to these properties in the speech input. In fact, unlike the attentional and affective functions of IDS, which have been well validated in studies showing infants' perceptual preferences for IDS over ADS (Cooper and Aslin, 1990; Fernald, 1985; Werker and McLeod, 1989; but see also Singh *et al.*, 2002), there has been very little direct evidence as to whether infants are able to use IDS for word learning. Karzon (1985) showed that 1–4-month-olds discriminated contrasts in [malana] and [maraṇa] only when the embedded syllables were spoken in IDS style. Similarly, Thiessen *et al.* (2005) showed that 7-month-olds were able to use transitional probabilities between syllables to distinguish words from partial words when they heard strings of nonsense words in IDS, but not in ADS. These findings suggest that the exaggerated acoustic characteristics of IDS might play a crucial role in speech segmentation, allowing infants to make discriminations they cannot make in ADS.

However, these studies focused on only a limited domain, i.e., segmentation of nonsense words by preverbal infants. Little is known about the possible facilitating effects of IDS in older infants whose vocabulary is rapidly expanding. Around 18 months, many infants gain speed in word learning, which is often cited as the 'vocabulary spurt' (Goldfield and Reznick, 1990). Around this age, infants not only increase the speed of acquiring new words, but also increase the efficiency with which they recognize the word in the speech stream (Fernald, 2000). Thus, it would be informative to examine how the unique acoustic characteristics of IDS affect infants' ability to recognize words around the time of the vocabulary spurt.

If IDS plays some role in infants' learning of language, it is also important to determine the mechanism of the facilitation. There are at least two possible ways in which the acoustic characteristics of IDS may facilitate speech processing in infants. One possibility is that IDS does this by effectively attracting and holding infants' attention. If infants pay

greater attention to speech, this may help them to discover and remember information about their native language more efficiently. The second possibility is that IDS enhances infants' language learning by providing better perceptual cues and more detailed linguistic information than ADS. That is, phonetic exaggeration in IDS may make individual linguistic units more distinct from one another, which may in turn help infants to learn the acoustic dimensions that distinguish linguistic units in their native language (Kuhl *et al.*, 1997).

The first step to address this issue would be to identify the acoustic properties that enhance infants' speech processing abilities. However, despite its importance, only a very limited amount of empirical evidence is available to address the question of which aspects of IDS are responsible for such facilitation. Among several typical acoustic characteristics of IDS, three candidates are of particular interest in the present study: slow speaking rate, hyper-articulation of vowels, and wide pitch range.

Interestingly, these three acoustic properties are also the ones that often characterize 'clear speech', a distinct speaking style that talkers adopt when they are aware of the listener's speech perception difficulty due to, for example, a hearing deficit, background noise, or a different native language (Smiljanic and Bradlow, 2005). A number of studies have investigated effects of clear speech on improving speech intelligibility for various listener populations, including hearing-impaired adult listeners (Picheny *et al.*, 1985), normal-hearing adult listeners (Payton *et al.*, 1994), non-native adult listeners (Bradlow and Bent, 2002), and school-aged children with and without learning disabilities (Bradlow *et al.*, 2003). Examining the effects of slow speaking rate, vowel hyper-articulation, and wide pitch range on infants' word recognition will fill the gap in the literature and provide a more comprehensive view of the role of clear speech. In the following sections, these three acoustic properties are discussed in more depth, paying special attention to the use of these properties in IDS, and how they might enhance infants' speech processing.

## A. Slow speaking rate

Slow speaking rate is widely acknowledged as one of the distinctive characteristics of IDS. For example, when German mothers addressed their newborns (3–5-days-old), their speaking rate was on average significantly slower (4.2 syllables/s) than when they spoke with another adult (5.8 syllables/s) (Fernald and Simon, 1984). In Cooper and Aslin, 1990, an English-speaking female speaker produced a set of sentences as she would to her infant and to another adult. Mean duration of these same sentences differed significantly between IDS and ADS: on average, AD sentences were 0.54 times shorter than ID sentences.<sup>1</sup>

Similarly, pauses in IDS are known to be longer than those in ADS. For instance, Fernald *et al.* (1989) showed that the mean pause duration between utterances in speech addressed to 4-month-olds was significantly longer than in ADS by both American English-speaking mothers (1.31 vs. 1.13 s) and fathers (1.86 vs. 1 s). Stern *et al.* (1983) also reported that the duration of the median pause was different

between IDS and ADS, as well as across infant age (to neonate: 1.63 s, 4 months: 0.95 s, 12 months: 1.45 s, 24 months: 1.38 s, adults: 0.68 s).

There have been very few studies examining how speaking rate affects infants' speech processing abilities. [Zangl et al. \(2005\)](#) showed that 12- to 31-month-olds recognized the target words more accurately when listening to the 'unaltered' stimuli as compared to the 'time-compressed' stimuli, which were twice as fast as 'unaltered' stimuli. Further evidence of the effects of speaking rate on listeners' speech processing abilities comes from other age groups. For instance, [Nelson \(1976\)](#) showed that 5–9-year-old children's comprehension of spoken sentences was significantly impaired when they listened to the sentences at a fast speaking rate (4.9 syllables/s) compared to a slower speaking rate (2.5 syllables/s) (see also [Berry and Erickson, 1973](#)). Furthermore, slower speaking rate is generally associated with better speech intelligibility in hearing-impaired populations ([Uchanski et al., 1996](#)), as well as in normal adult listeners ([Bradlow and Pisoni, 1999](#)). These findings suggest that slower speaking rate in IDS may also improve infants' ability to process speech. Slower speaking rate presumably makes sounds more distinct from one another, providing infants with richer perceptual information as compared to the poorly specified information in ADS. In addition, slower speech likely provides infants with more time to process speech.

## B. Hyper-articulation of vowels

In addition to slower speaking rate, studies have also shown that IDS often contains vowels that are hyper-articulated or more clearly enunciated than those in ADS (see [Gay, 1978](#), for the relationship between speaking rate and vowel space). For example, [Kuhl et al. \(1997\)](#) examined language input directed to 2- to 5-month-old infants in the United States, Russia, and Sweden. In all three countries, mothers articulated vowels more clearly when talking to their infants than when talking to an adult, resulting in an expanded acoustic vowel space for these hyper-articulated vowels. On average, the area of the vowel triangle was expanded by 92% in IDS (English: 91%, Russian: 94%, Swedish: 90%). The authors therefore suggested that "language input to infants provided exceptionally well-specified information about the linguistic units that form the building blocks for words" ([Kuhl et al., 1997](#), p. 684). [Burnham et al. \(2002\)](#) reported similar results showing that mothers hyper-articulated vowels in speaking to their 6-month-old infants but not to their pets or another adult, although pitch was high and emotions were strong in both IDS and pet-directed speech.

To date, there has been no empirical investigation of whether expanded vowel space has a direct effect on infants' processing of speech. However, supporting evidence for a potential long-term advantage of listening to IDS with larger vowel spaces comes from [Liu et al., 2003](#). In their study, Chinese mothers who used a larger vowel space had 6–8-month-olds and 10–12-month-olds who performed better in a task involving discrimination of a Chinese affricate-fricative contrast. Their results suggested that mothers' vowel clarity

might have positive effects on infants' learning of phonemes. Furthermore, studies with both normal ([Bradlow et al., 1996](#)) and dysarthric ([Liu et al., 2005](#)) speakers indicate that speakers with larger vowel spaces were judged by normal adult listeners to be more intelligible than speakers with smaller vowel spaces. These results suggest that infants' perception of speech may be also benefit from vowel hyper-articulation.

## C. Wide pitch range

Expansion of pitch range in IDS has been extensively documented. For example, [Stern et al. \(1983\)](#) reported on 6 mothers' use of various prosodic characteristics of IDS at four different stages of development: at 2 to 6 days ('the neonatal stage'), at 4 months ('a pre-linguistic stage of face-to-face social interaction'), at 12 months ('the one-word stage of production'), and at 24 months ('when more complex language production is under way'). For infants at all ages, mothers used a wider pitch range in their speech to infants than adults, but particularly when infants were about 4 months of age (to neonates: 115 Hz, 4 months: 209 Hz, 12 months: 166 Hz, 24 months: 157 Hz, adults: 95 Hz).

Exaggerated pitch range is known to play an important role in attracting and maintaining infants' attention. For example, [Fernald and Kuhl \(1987\)](#) examined which of three acoustic characteristics of IDS 4-month-olds paid attention to: pitch, amplitude, or duration. Their results showed that infants listened longer to IDS than ADS when the two types of speech differed in pitch. In contrast, when IDS differed from ADS only in amplitude or duration, infants showed no preference for IDS. Thus, [Fernald and Kuhl \(1987\)](#) concluded that pitch was the determining factor that attracts infant's attention to IDS (cf. [Kitamura and Burnham, 1998](#); [Singh et al., 2002](#)).

If the pitch characteristics of IDS help infants to pay attention to speech, this may also help them notice linguistically relevant aspects of the speech input. Using a conditioned head-turn procedure, [Trainor and Desjardins \(2002\)](#) examined the effects of exaggerated pitch contours on vowel discrimination in 6-month-old infants. Their results showed that infants discriminated vowels with large pitch contours more easily than the same vowels with steady-state pitch, suggesting that the exaggerated pitch contours of IDS promoted vowel discrimination.

Alternatively, although wide pitch range is effective in drawing infants' attention to speech, it may not have a direct effect on speech processing in infants. Using a high-amplitude sucking procedure, [Karzon \(1985\)](#) examined whether wide pitch range would be sufficient to help infants to discriminate the syllables /la/ and /ra/ in the synthesized nonsense words [malana] and [marana]. It was found that 1- to 4-month-old infants failed to discriminate the embedded syllables regardless of whether the contrastive syllables had a wide pitch range (180–350 Hz) or not (180–220 Hz). One possible explanation for this failure to discriminate is that pitch range alone may have been insufficient to enhance the differences between syllables, even if it might be sufficient to command the infant's attention.

Furthermore, there is an emerging body of evidence

showing that infants' attentional preference for IDS decreases over time, with no apparent preference by 9 months (Newman and Hussain, 2006). This is also the time at which infants begin to tune into the fine phonetic details of their native language (Werker and Tees, 1984). Infants over 1 year of age might therefore focus more on the linguistically relevant acoustic properties in the speech signal rather than simply pitch variation, which has been shown to play an important role in attracting infants' attention (at least for learners of a language like English). Thus, it would be useful to compare the role of pitch range in older infants' word recognition to that of speaking rate and vowel space, which may help infants' word recognition by providing them more detailed linguistic information. The goal of this study was therefore to examine the extent to which slow speaking rate, hyper-articulation of vowels, and wide pitch range might facilitate 19-month-olds' ability to recognize the target word during an intermodal preferential looking task. If these acoustic properties facilitate infants' word recognition, it is predicted that their ability to identify the target referents would deteriorate when they listen to the test stimuli lacking them.

## II. METHOD

### A. Participants

Participants were 48 (21 females, 27 males) monolingual English-learning 19-month-old-infants (range: 18.6–20.7 months, mean: 19.4 months) recruited in Providence, RI. All participants were healthy infants with normal hearing and vision. An additional 7 infants were tested but were excluded from the analysis due to fussiness during testing ( $n=5$ ), or experimental error ( $n=2$ ).

### B. Stimuli

Visual stimuli included pictures of the target and a distractor which were presented after listening to the audio stimuli *Where's the target?* (e.g., *Where's the cup?*). There were 12 target-distractor pairs in total (see Appendix A). The names of the target and distractor objects were chosen to be easily picturable and highly familiar to 19-month-olds. The familiarity of words was determined by the percentage of 16-month-old infants reported to understand the words in the MacArthur Communicative Development Inventory (MacArthur CDI) (Dale and Fenson, 1996). Familiarity was matched between the target and distractor word pairs in order to prevent potential word familiarity effects.

Half of the test stimuli (*Where's the target?*) were presented in typical IDS style. The other half lacked one of the three acoustic properties under investigation. For simplicity, we will call the former *Typical-IDS* and the latter *Modified-IDS*. There were three kinds of Modified-IDS stimuli depending on which acoustic characteristic was removed: *Fast-IDS* stimuli lacked the characteristic of slow speaking rate. *Hypo-articulated-IDS* stimuli lacked the characteristic of hyper-articulated vowels. *Monotonous-IDS* stimuli lacked the characteristic of wide pitch range. Except for the single acoustic characteristic that was lacking in each condition, other acoustic characteristics remained the same. The degree

of modification was determined based on the values reported in the literature to reflect the differences in speaking rate, pitch range, and vowel space between IDS and ADS (e.g., Cooper and Aslin, 1994; Kuhl *et al.*, 1997). Due to the different methods and measures taken in different studies, we focused on the difference in ratio of acoustic values between IDS and ADS found in previous studies, rather than the difference in absolute values. In addition, amplitude was controlled, with the mean value for all stimuli at 75 dB.

Typical-IDS stimuli were naturally spoken and recorded by one female speaker. Both Fast-IDS and Monotonous-IDS stimuli were created by digitally manipulating the Typical-IDS stimuli using PRAAT software (Boersma and Weenink, 2005). Either the speaking rate or the pitch range of the *entire* sentence was manipulated uniformly throughout the utterance. The average speaking rate of Typical-IDS stimuli was 1.94 syllables/s (average duration: 1.68 s). Fast-IDS stimuli were twice as fast as the Typical-IDS stimuli (average speaking rate: 3.88 syllables/s, average duration: 0.84 s). The average duration of Typical-IDS target words and Fast-IDS target words was 0.73 and 0.365 s, respectively. Similarly, Monotonous-IDS stimuli had pitch ranges half of the original value in Typical-IDS stimuli (average pitch range: 194 vs 388 Hz, in semitones: 12 vs 21). The only difference between Typical-IDS and Monotonous-IDS stimuli was the range of pitch; the average pitch level between the two stimuli was kept the same. Pitch values were obtained using PRAAT, using 75 and 800 Hz for the minimum and maximum pitch, respectively, in tracking pitch values.

We now turn to Hypo-articulated-IDS stimuli. Vowel articulation is generally defined in terms of vowel formants. Due to the technical challenges of manipulating vowel formants digitally, good exemplars of hypo-articulated vowels from natural speech were used; the same female speaker was asked to utter the target words more casually or less clearly, and then the best exemplars were chosen among the words produced. Speaking rate and pitch range of these stimuli were digitally controlled using PRAAT software. Formants were measured at the mid-point or at the most stable region in the vowel using WAVESURFER (Sjölander and Beskow, 2000). The following parameters were employed: number of formants: 4, analysis window length: 0.049, pre-emphasis factor: 0.97, LPC order: 12. Various adjustments were made in the cases where the automatic formant tracking showed numbers too different from the values reported in the literature (e.g., Kuhl *et al.*, 1997) or if they did not agree with the visual estimate of formant locations on the spectrograms. Adjusting the number of formants to 3 and LPC order to 10 also improved the formant tracking depending on the tokens. FFT and LPC spectra, as well as the automatic formant tracking in PRAAT, were also referred to as needed.

The stimulus exemplars were chosen using the following steps: There were target words which contained vowels /a/ (*block, car*), /i/ (*bunny, kitty*), and /u/ (*spoon*). First, the area of the vowel space in Typical-IDS was calculated for the vowels /a/, /i/, and /u/. The triangulation formula used to calculate the vowel space area was the same as that employed in Liu *et al.*, 2003:

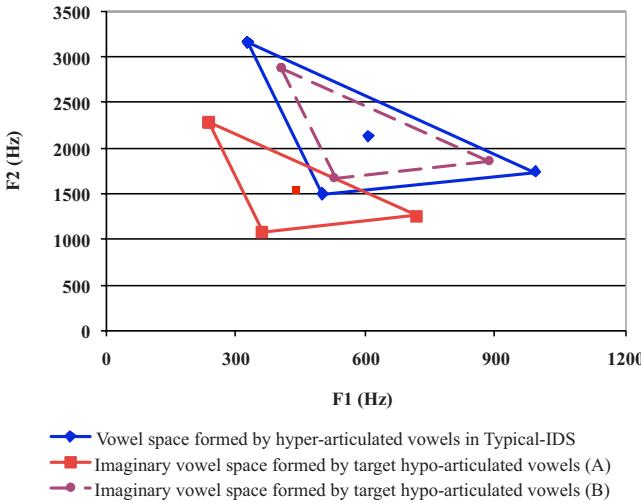


FIG. 1. (Color online) Vowel spaces formed by /a/, /i/, /u/ in Typical-IDS and the imaginary ADS vowel spaces that were used as reference points for Hypo-articulated-IDS. The imaginary ADS vowel space was created by first shrinking Typical-IDS vowel space (A), and then equalizing the circumstances of the triangles (B).

$$\{F1i^*(F2a - F2u) + F1a^*(F2u - F2i) + F1u^*(F2i - F2a)\}/2. \quad (1)$$

Here,  $F1i$  represents the  $F1$  of vowel /i/,  $F2a$  represents the  $F2$  of vowel /a/, and so forth. The vowel space formed by /a/, /i/, and /u/ in the present study was  $429111 \text{ Hz}^2$ . This value is slightly larger than the average vowel space that English-speaking mothers used in speaking to their 2- to 5-month-olds in [Kuhl et al., 1997](#) (area=412900 Hz $^2$ ). Second, the circumcenter of the vowel triangle was calculated by averaging  $F1$  and  $F2$  of /a/, /i/, and /u/. Third, we calculated an ‘imaginary’ vowel space triangle for ADS so as to establish a baseline for selecting  $F1$  and  $F2$  values corresponding to hypo-articulated vowels. [Kuhl et al. \(1997\)](#) showed that, on average, mothers’ vowel space area was expanded by 91% in IDS as compared to ADS. We therefore determined the end points of the imaginary triangle by decreasing those of the Typical-IDS triangle by this same percentage. This was achieved by multiplying the  $F1$  and  $F2$  values of vowels /a/, /i/, and /u/ in Typical-IDS by the vector of 0.72 [see Fig. 1(A)]. The circumcenters of the IDS and ADS triangles were then equalized by adding 167 and 589 Hz to the reduced  $F1$  and  $F2$ , respectively [see Fig. 1(B)]. This was partly motivated by [Kuhl et al.’s \(1997\)](#) finding that the center of vowel space in ADS did not shift extremely from that in IDS. It was also expected that the formants of a vowel produced by the same speaker would generally occur in a similar frequency region.

Fourth, we determined the  $F1$  and  $F2$  values of target hypo-articulated vowels for each of the 12 target words by multiplying the same vector (0.72) and then adding the same numbers (167, 589) to the original  $F1$  and  $F2$  values of hyper-articulated vowels in Typical-IDS. Finally, among the attempted hypo-articulated vowels, we chose those having  $F1$  and  $F2$  values closest to those of the target hypo-articulated vowels or those with formant values falling inside the imaginary vowel space formed by the target hypo-

TABLE I. The video and audio stimuli for one block of trials.

	Duration (s)	Center light	Monitor 1	Audio stimuli	Monitor 2
Familiarization trials	3	Off		<i>What's this!</i>	Blank screen
	3	Off	Blank screen	<i>Look at this!</i>	
Salience trial	3	Off		<i>What are these!</i>	
Test trial		On	Blank screen	<i>Where's the cup?</i>	Blank screen
	2.5	Off			

articulated vowels. Appendix B presents the  $F1$  and  $F2$  values of hyper-articulated vowels in Typical-IDS, target hypo-articulated vowels, and the actual formant values of vowels that were used as Hypo-articulated-IDS stimuli.

### C. Procedure

The intermodal preferential looking procedure (IPLP) ([Hirsh-Pasek and Golinkoff, 1996](#)) was used to examine infants’ preferential looking to one of the two video events that corresponded to audio stimuli. The logic of IPLP is that infants should look more quickly and longer at the visual stimulus that matches the audio input if they understand what they are hearing. During the procedure, the infant sat on the caretaker’s lap and was presented with the visual and audio stimuli. Caretakers listened to music over headphones to mask the audio stimuli.

### D. Test design

Each of 12 blocks of trials consisted of three different kinds of sub-trials: familiarization, salience, and test trials (see Table I). During the familiarization and salience trials, infants were presented with pictures of the target and a distractor while listening to ‘neutral’ sentences such as *What’s this!/Look at this!* (familiarization trials) and *What are these?* (salience trials) spoken in typical IDS style. These trials served to create the expectation that something would appear on each screen and introduce the video events before the infant had to find the match for the audio stimulus ([Hirsh-Pasek and Golinkoff, 1996](#), p. 64). Infants’ performance on the familiarization trials was not included in the analyses, but their looking to the two pictures during the salience trials were measured and used as a baseline to calculate the increase in the proportion of looking time to the target during the test trials.

During the following test trial, infants listened to a sentence *Where’s the target?* either in Typical-IDS or Modified-IDS. Stimulus type was a within-subject factor, and each infant listened to 6 test stimuli in Typical-IDS and the other 6 test stimuli in Modified-IDS. There were three conditions depending on which acoustic property was modified in Modified-IDS stimuli: speaking rate, vowel space, and pitch range. As shown in Table II, the three different acoustic properties were between-subject factors, with 16 infants in each condition. There were two subgroups of infants in each condition to counterbalance the target words across two different speech styles. While infants listened to the test stimu-

TABLE II. Design of the word recognition experiment.

Conditions (between-subject factor)		
Speaking rate ( <i>n</i> =16)	Vowel space ( <i>n</i> =16)	Pitch range ( <i>n</i> =16)
Typical-IDS	Typical-IDS	Typical-IDS
Fast-IDS	Hypo-articulated-IDS	Monotonous-IDS

lus *Where's the target?*, a light mounted between two video monitors blinked to provide them something to fixate. Target words were always placed at the end of the sentence, and as soon as the target word ended, the light was extinguished and pictures of target and distractor were displayed on the monitors.

### E. Measures and predictions

Infants' gaze was video-recorded and then coded frame by frame (1 frame=33 ms) using the SUPERCODER software (Hollich, 2003) by a coder who was blind to which side the target picture was on. There were three measures to examine how accurately and fast infants recognize the target words during the 2.5 s of the test trials: proportion of looking time, latency of the first look, and time course of infants' looking at the target picture. We examined infants' looking behavior for 2.5 s, because they were expected to be familiar with the pictures through the familiarization and salience trials, and also because many studies used a time window smaller than 3 s from the presentation of the visual stimuli to provide reliable measures of infants' speech processing abilities (e.g., Swingley and Fernald, 2002).

If the slower speaking rate, hyper-articulated vowels, and wide pitch range of IDS enhanced infants' recognition of words, we would expect more accurate and faster responses to the target words than when the speech stimuli lacked one of these characteristics. However, if these acoustic properties of IDS did not influence infants' word-recognition abilities, performance should be equally good when listening to speech with or without each of these acoustic properties.

## III. RESULTS

### A. Analysis 1: Proportion of looking time to the target

The proportion of looking time to the target was defined as the looking time to the target over the time that the infant spent looking at either target or distractor, and it is typically used to measure infants' accuracy in recognizing the target word (e.g., Zangl *et al.*, 2005). For each of 12 blocks of trials (6 Typical-IDS, 6 Modified-IDS) that each infant received, we computed the proportion of looking time to the target during the salience trial, which served as a baseline for comparison, and the following test trial. Then the difference in the proportion of looking time to the target between the salience trial and the test trial was computed. Finally, the average of differences in the proportion looking time to target was calculated for each set of 6 Typical-IDS and 6 Modified-IDS stimuli for individual infants and the two averages per infant were used in computing test statistics.

Recall that during the salience trial, infants were presented with two pictures while listening to a neutral sentence

*What are these!* During the following test trial, the infants were presented with the same pictures, but this time, they were expected to find the target upon listening to a sentence *Where's the target?* Thus, it was predicted that the proportion looking to the target would increase from the neutral salience trial to the test trial. Furthermore, if slower speaking rate, hyper-articulation of vowels, and wide pitch range helped infants recognize the target word more accurately, it was predicted that the increase in the proportion of their looking to the target would be greater when listening to the test stimuli *Where's the target?* in Typical-IDS as compared to Modified-IDS.

On average, across the conditions and the stimuli types, infants looked at the two pictures for 2.45 ( $SD=0.16$ ) s during the 3 s of salience trial. The proportion of looking time to the target object was 54% ( $SD=9.95$ ). During the test trial, the time that infants spent looking at either target or distractor was 1.97 ( $SD=0.24$ ) s out of 2.5 s, and the proportion of looking time to the target object was 68% ( $SD=11.63$ ). As expected, a series of one-sample t-tests indicated that the increase in the proportion looking time to the target from the salience trial to the test trial was significantly different from the chance value of 0% for both Typical-IDS and Modified-IDS stimuli in all three conditions, speaking rate condition: Typical-IDS,  $t(15)=7.45$ ,  $p<0.001$ , Modified-IDS,  $t(15)=3.92$ ,  $p<0.01$ , vowel articulation condition: Typical-IDS,  $t(15)=4.13$ ,  $p<0.01$ , Modified-IDS,  $t(15)=3.25$ ,  $p<0.01$ , pitch range condition: Typical-IDS,  $t(15)=2.15$ ,  $p<0.05$ , Modified-IDS,  $t(15)=2.74$ ,  $p<0.05$ .

We next examined whether the increase in proportion looking to target from the salience trial to the test trial differed as a function of stimulus types and the three acoustic properties under investigation. To this end, a mixed analysis of variance (ANOVA) was conducted with stimulus type (Typical-IDS, Modified-IDS) as the within-subjects factor, and the three conditions (speaking rate, vowel space, pitch range) as the between-subjects factor. The dependent variable was the difference in infants' proportion looking to the target between the salience trial and the test trial. The results showed a significant main effect of stimulus type,  $F(1,45)=6.34$ ,  $p<0.05$ . Overall, the change in the proportion looking to the target was greater when they listened to Typical-IDS (17%) as compared to Modified-IDS (11%). The main effect of condition was not significant,  $F(2,45)=2.03$ ,  $p=0.14$ , suggesting that infants' accuracy in recognizing the target did not differ as a function of condition (speaking rate: 18%, vowel space: 15%, pitch range: 9%). The interaction between stimulus type and condition was also not significant,  $F(2,45)=0.95$ ,  $p=0.40$ .

Nevertheless, we were interested in whether the change in the proportion of looking time to the target from the salience trial to the test trial was significantly different between Typical-IDS and Modified-IDS for each of the three modification conditions. A paired t-test demonstrated that the change in the proportion of looking time to the target from the salience trial to the test trial was significantly greater for Typical-IDS ( $Mean=23.43$ ,  $SD=12.59$ ) compared to Fast-IDS ( $Mean=12.65$ ,  $SD=12.91$ ),  $t(15)=3.18$ ,  $p<0.01$ . Thus, although the proportion looking time to the target generally

increased during the test trial for both Typical-IDS and Fast-IDS, the amount of increase was significantly larger for Typical-IDS than for Fast-IDS. This suggests that slower speaking rate helps infants recognize words more accurately. On the other hand, the amount of increase in proportion looking to target did not differ significantly for (hyper-articulated) Typical-IDS ( $Mean=18.33$ ,  $SD=17.77$ ) versus Hypo-articulated-IDS ( $Mean=11.87$ ,  $SD=14.59$ ),  $t(15)=1.26$ ,  $p=0.23$ . Finally, as discussed earlier, infants' proportion looking to the target significantly increased from the salience trial to the test trial for Typical-IDS and Monotonous-IDS stimuli in pitch range condition as well, but the amount of increase did not differ as a function of pitch range (Typical-IDS:  $Mean=10.48$ ,  $SD=19.53$ , Monotonous-IDS:  $Mean=8.34$ ,  $SD=12.20$ ),  $t(15)=0.46$ ,  $p=0.65$ . In sum, Analysis 1 demonstrated that slower speaking rate significantly improved infants' accuracy in recognizing the target words in the test trial as compared to the salience trial. In contrast, their proportion looking time to the target was not affected by either vowel articulation or pitch range.

## B. Analysis 2: Latency of the first look to the target

In Analysis 2, the average latency of the first look to the target was calculated for each infant to investigate how typical IDS versus modified speech without these characteristics might affect infants' speed at identifying the target word. Response latency was defined as the time taken for infants to first look at the target within 2.5 s after the target word offset, regardless of whether they looked directly at the target picture after they listened to the test stimuli, or whether they incorrectly looked at the distractor first and then shifted to the target. If infants did not look at the target until 2.5 s after the target word offset, the latency time of the trial was calculated as 2.5 s (see Gout *et al.*, 2002, for the same method for calculating latencies). This involved 44 out of 576 total trials in three conditions. In addition, 5 trials out of 576 trials where infants were looking at the target at the word offset were excluded in the analysis, as it was not clear how to interpret them.

A mixed ANOVA was conducted with stimulus type (Typical-IDS, Modified-IDS) as the within-subjects variable, and the three conditions (speaking rate, vowel space, pitch range) as the between-subjects variable. The dependent variable was average latency of the first look to the target. The results showed a main effect of stimulus type,  $F(1,45)=5.10$ ,  $p<0.05$ . Overall, latency was shorter when infants listened to Typical-IDS (0.70 s) compared to Modified-IDS (0.80 s). The main effect of condition was not significant,  $F(2,45)=0.30$ ,  $p=0.74$ , suggesting that latency did not differ as a function of condition (speaking rate: 0.72 s, vowel space: 0.75 s, pitch range: 0.78 s). There was no significant interaction between stimulus type and condition,  $F(2,45)=1.67$ ,  $p=0.20$ .

Again, we were interested in whether differences in latency between Typical-IDS and Modified-IDS were significant within each of the three modification conditions. A paired t-test demonstrated that infants were faster to find the

target when they listened to Typical-IDS ( $Mean=0.63$ ,  $SD=0.29$ ) as compared to Fast-IDS ( $Mean=0.81$ ,  $SD=0.24$ ),  $t(15)=-2.45$ ,  $p<0.05$ . On average, infants tended to be slightly faster to look at the target when they listened to Typical-IDS ( $Mean=0.68$ ,  $SD=0.25$ ) as compared to Hypo-articulated-IDS ( $Mean=0.82$ ,  $SD=0.32$ ). However, this difference did not reach significance,  $t(15)=-1.56$ ,  $p=0.14$ . Nor was there a significant difference in the third condition: infants were equally fast orienting to the target whether they listened to Typical-IDS ( $Mean=0.78$ ,  $SD=0.32$ ) or Monotonous-IDS ( $Mean=0.78$ ,  $SD=0.25$ ),  $t(15)=0.17$ ,  $p=0.87$ . Thus, as in Analysis 1, Analysis 2 showed that latency of the first look to the target differed only as a function of speaking rate. On the other hand, latency did not differ as a function of vowel space or pitch range.

## C. Analysis 3: Time course of infants' looking to targets and distractors

Analyses 1 and 2 demonstrated that 19-month-olds showed more accurate and faster responses to the target words when they listened to Typical-IDS as compared to Fast-IDS. For vowel space, the proportion looking to the target was on average greater and the latency of the first look to the target was on average shorter for Typical-IDS than for Hypo-articulated-IDS, but the differences were not significant. Thus, we examined the time course of infants' word recognition process within 2.5 s after target word offset, providing a more fine-grained view of infants' processing abilities than a single, summary score based on the proportion looking time or latency time to the target (Fernald *et al.*, 2008).

Figures 2–4 show the time course of infants' looking to targets and distractors in listening to Typical-IDS and Modified-IDS. The x-axis is the time from the target word offset, plotted at 33 ms intervals; 0 on the x-axis indicates the time point when the target word ended. The y-axis is the proportion of trials where infants looked at the target (two thick lines on the top) and the distractor (two thin lines in the bottom). There was a third alternative in this case, which was 'look away', to provide more comprehensive information on infants' looking behavior during the test trials. Thus, the proportion of trials where infants looked at the target and a distractor does not necessarily add to 100% in Fig. 2–4, reflecting the fact that there were also some trials where infants looked away. The proportion of trials is averaged across 16 infants in each condition for each 33 ms interval. To better understand the time-course pattern of infants' behavior, the 33 ms frames were then grouped into 100 ms bins, and analyses were done in each of the 100 ms windows. To anticipate, looking to the target diverged from looking to the distractor earlier for Typical-IDS stimuli than for Fast-IDS (Fig. 2) or Hypo-articulated-IDS stimuli (Fig. 3). In contrast, looking to target and distractor closely tracked across time for Typical-IDS and Monotonous-IDS stimuli (Fig. 4).

### 1. Speaking rate condition

Figure 2 shows the proportion of trials in which infants looked at the target and distractor over time in listening to

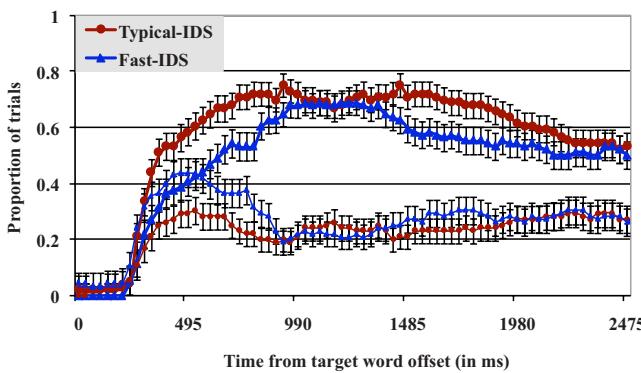


FIG. 2. (Color online) Speaking rate condition: Proportion of trials where infants ( $n=16$ ) looked to the target and the distractor over time. Error bars represent standard errors.

Typical-IDS and Fast-IDS. For example, 495 ms after the target word was spoken in Typical-IDS, infants were correctly looking at the target in about 60% of the total trials. On the other hand, they were looking at the target in only about 40% of the trials when they listened to Fast-IDS.

We first discuss infants' performance on Typical-IDS stimuli. Between 0 ms (word offset and appearance of visual stimuli) and 300 ms, there was no difference in proportion looking at target and distractor. Starting from a 301–400 ms window, however, infants looked at the target object significantly more than the distractor (by subjects:  $t(15)=2.36, p < 0.05$ , by items:  $t(11)=2.99, p < 0.05$ ), and the difference remained highly significant until the end of the trial. For the Fast-IDS stimuli, infants were less accurate in identifying the target word, and it also took them longer to recognize the target word; infants looked more at the distractor than the target soon after the word offset [0–101 ms: by subjects:  $t(15)=-1.82, p=0.09$ , by items,  $t(11)=-2.24, p < 0.05$ ; 101–200 ms: by subjects:  $t(15)=-2.24, p < 0.05$ , by items:  $t(11)=-1.77, p=0.10$ ]. It was not until the 800–900 ms window that proportional looking to the target was significantly larger than proportional looking to the distractor, by subjects:  $t(15)=4.40, p < 0.001$ , by items:  $t(11)=3.72, p < 0.01$ .<sup>2</sup>

## 2. Vowel space condition

We now examine the time course of infants' looking at target and distractor in the vowel space condition. A com-

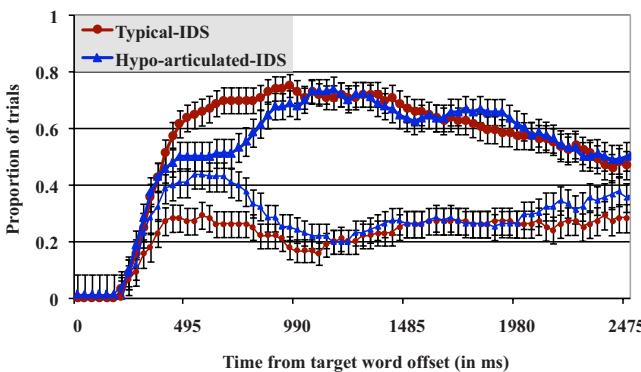


FIG. 3. (Color online) Vowel space condition: Proportion of trials where infants ( $n=16$ ) looked to the target and the distractor over time. Error bars represent standard errors.

parison of Figs. 2 and 3 shows the striking similarity between the speaking rate condition and the vowel space condition between approximately 0–990 ms; when infants listened to Fast-IDS or Hypo-articulated-IDS, the difference in looking proportion between the target and a distractor emerged later than for Typical-IDS.

The analyses on 100 ms bins showed that a significant difference in the proportion of looking to target vs. distractor emerged during the period of 301–400 ms for the Typical-IDS stimuli both by subjects [ $t(15)=3.06, p < 0.01$ ] and by items [ $t(11)=2.99, p < 0.05$ ]. The difference remained significant for the duration of the trial. In contrast, infants were slower to identify the target word when they listened to Hypo-articulated-IDS. The point of divergence did not occur until the 801–900 ms window, by subjects:  $t(15)=3.50, p < 0.01$ , by items:  $t(11)=3.72, p < 0.01$ . Furthermore, unlike the Typical-IDS stimuli, where the difference persisted throughout the test trial, the difference in looking proportion between the target and distractor disappeared around 2100 ms after the word offset when infants listened to the Hypo-articulated-IDS stimuli. In sum, although the single, summary scores based on total looking time or latency of the first look did not capture an effect of vowel hyper-articulation, time course analyses showed that hyper-articulation of vowels facilitated infants' word recognition particularly at the beginning of the word recognition process.

## 3. Pitch range condition

Analyses 1 and 2 showed that looking time and latency to the target were not affected by pitch range. The time course analysis was consistent with this, showing no difference in infants' behavior as a function of pitch range (see Fig. 4). For the Typical-IDS stimuli, the analyses on the 100 ms window showed that a significant difference in the proportion looking to the target and a distractor emerged during 201–300 ms in the by-subjects analysis,  $t(15)=3.17, p < 0.01$ , and during 901–1000 ms in the by-items analysis,  $t(11)=2.82, p < 0.05$ . Infants showed similar behavior for Monotonous-IDS stimuli. In the by-subject analysis, proportion looking to target and distractor differed from a 201–300 ms window,  $t(15)=2.52, p < 0.05$ . In the by-item analysis, the difference was significant during a 201–300 ms window,  $t(11)=2.51, p < 0.05$ , and during a 601–700 ms window,

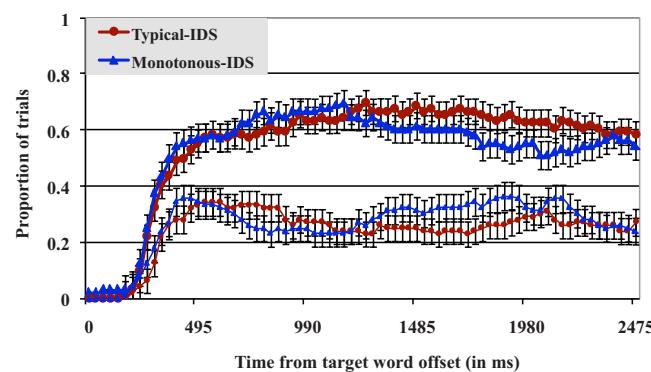


FIG. 4. (Color online) Pitch range condition: Proportion of trials where infants ( $n=16$ ) looked to the target and the distractor over time. Error bars represent standard errors.

$t(11)=2.38$ ,  $p<0.05$ , and onwards. In both by-subject and by-item analyses of the Monotonous-IDS stimuli, the difference was not present during 1801–2200 ms.

To summarize, 19-month-olds were about 500 ms faster to look proportionately more at the target than the distractor when they listened to Typical-IDS compared to Fast-IDS or Hypo-articulated-IDS. Furthermore, as shown in Figs. 2 and 3, proportional looking to the target was overall higher for Typical-IDS compared to Fast-IDS or Hypo-articulated-IDS, suggesting that slow speaking rate and hyper-articulation of vowels increased accuracy in identifying the target. In contrast, infants' ability to recognize words was not affected by pitch range.

Recall that our Typical-IDS stimuli were naturally produced, whereas the Modified-IDS stimuli were manipulated in some way. One might then wonder if this was a potential confounding factor affecting the results. As a way to address this issue, we conducted naturalness ratings on the Typical-IDS and Fast-IDS stimuli that were used in the experiment. We focused on these stimuli because we found the largest difference in infant behavior in the speaking rate condition. The main finding was that there is no significant difference in ratings between naturally spoken Typical-IDS stimuli and manipulated Fast-IDS stimuli. Below we lay out the detailed procedure and the results.

The participants in the naturalness ratings were 12 college students with normal hearing. During the ratings, students were presented with 24 pairs of stimuli. Each pair contained the same sentence (e.g., *Where's the book?*), but one sentence was naturally produced and the other was manipulated. Half of the pairs (12 pairs) consisted of naturally spoken Typical-IDS and manipulated Fast-IDS. These stimuli were the ones originally used in the study. The other 12 pairs were “control stimuli” and they consisted of naturally spoken fast speech and manipulated slow speech. The control stimuli, which were not used in the original study, were included here to prevent the possibility that the subjects would always judge the fast speech as manipulated if only naturally spoken Typical-IDS and manipulated Fast-IDS pairs were presented.

Twenty-four pairs of stimuli were randomly mixed and presented. For half of the pairs, participants listened to the manipulated stimuli first. In the other half, naturally spoken stimuli were presented first. After listening to each pair, participants were asked to judge whether the first or the second stimulus had been manipulated. They were also asked to indicate their level of confidence by choosing between 1 (very confident) and 5 (not at all confident).

The results showed that there was no significant difference in naturalness ratings between naturally spoken Typical-IDS stimuli and manipulated Fast-IDS stimuli. On average, participants were able to correctly determine the manipulated stimuli only in 51% of the trials. A one-sample t-test indicated that this value was not significantly different from the chance value of 50%,  $t(143)=0.17$ ,  $p=0.87$ . There was no item effect; effects held equally well for all 12 pairs. Finally, a paired t-test showed that there was no significant difference in confidence ratings between incorrectly determined pairs

(2.76) and correctly determined pairs (3.16),  $t(10)=-1.84$ ,  $p=0.10$ . This suggested that the subjects were equally confident of their correct or incorrect decisions.

In sum, adult subjects were not able to determine which was manipulated and which was naturally spoken between Typical-IDS and Fast-IDS. This suggests that the difference in effect between Typical-IDS and Fast-IDS is due to their difference in speaking rate, but not due to the fact that one is naturally spoken and the other is manipulated. Although infants may have different sensitivities than adults, the results from the naturalness ratings suggest that the Typical-IDS and Fast-IDS stimuli in the present study sounded equally natural.

In addition, unlike many eye-tracking studies, measurements were made from the offset of the target word in this study, rather than the target word onset. This was also when the pictures of target and distractor were presented. We have chosen this method because we wanted to compare infants' ability to recognize words within the same time window, but the durations of the same target word were different between Typical-IDS and Fast-IDS stimuli. However, since infants had already seen the pictures during familiarization and salience trials, which were prior to the actual test trial, it is possible that infants might have started orienting to the monitor where the target picture had been on even before the target word ended. This is a particularly interesting question given that studies have shown that the speed and efficiency of word recognition increases over the course of the development, and that by 2 years of age, infants can orient to the target picture before the end of the spoken word (Fernald *et al.*, 1998).

However, the time course analyses shown in Figs. 2–4 demonstrated that this was not the case in the present study. The number of trials in which infants were already looking at the target or the distractor at the word offset was very small for both the Typical-IDS (5 out of 288 trials in three conditions) and Modified-IDS (7 out of 288 trials in three conditions). This suggests that infants' gaze was at the same starting point for measurements when listening to both Typical-IDS and Modified-IDS. Recall that the center light was on while the audio stimulus was played during the test trial, and the pictures of target and distractor were presented at the target word offset. We speculate that this is probably why infants in the current study started orienting to the target only after the target word ended, unlike findings from previous studies. In this regard, it would be useful to compare our results to those of Zangl *et al.* (2005). When the target picture was presented at the onset of the target word in their study, infants' response time to the target was shorter for ‘time-compressed’ stimuli than for ‘unaltered’ stimuli. This reflected the fact that the information specifying the target word was available earlier in ‘time-compressed’ condition. In contrast, the accuracy in identifying the target was greater when infants listened to ‘unaltered’ stimuli, suggesting that slower speaking rate improved infants' accuracy.

#### IV. DISCUSSION

The present study examined the individual roles of three typical acoustic characteristics of IDS in 19-month olds' rec-

ognition of familiar words: slow speaking rate, hyper-articulation of vowels, and wide pitch range. The results showed that slow speaking rate significantly enhanced 19-month-olds' ability to recognize words. This suggests that the slow speaking rate characteristic of IDS effectively accommodates the perceptual needs of infants, who are not yet as efficient as adults in processing speech.

Although infants showed no difference in either total looking time or latency of the first look to the target as a function of vowel space, time course analysis suggested that 19-month-olds' word recognition might also benefit from hyper-articulated vowels to some degree. That is, the difference in proportion looking to target vs. distractor emerged significantly earlier when infants listened to Typical-IDS than Hypo-articulated-IDS, suggesting that they were faster to recognize the target word when they listened to the target word with hyper-articulated vowels as compared to hypo-articulated vowels. This result is important because it provides the first empirical evidence that infants use the hyper-articulation cue in word recognition.

Exaggerated pitch range is known to be effective in drawing infants' attention to speech (Fernald and Kuhl, 1987). However, the findings of the current study showed that wide pitch range did not necessarily enhance 19-month-olds' ability to detect the sound patterns of words in the word recognition task. In sum, slow speaking rate and hyper-articulation of vowels facilitated 19-month-olds' ability to recognize words, whereas wide pitch range did not.

The distinction between English vowels typically involves contrasts in duration, as well as formant frequencies (e.g., as in *beat* and *bit*). In this respect, it is interesting that 19-month-olds were particularly sensitive to speaking rate and vowel clarity, which might affect duration and formant frequencies. Slow speaking rate and hyper-articulation of vowels may also provide more detailed linguistic information to infants: slow speaking rate helps to separate phonemes from the surrounding ones. Similarly, hyper-articulation of vowels is known to have the effect of making vowels more distinct from one another.

Speech rate and vowel clarity carry important cues to linguistic contrasts. On the other hand, pitch range is typically independent of vowel quality in a non-tonal language like English. Rather, pitch range is one of the typical acoustic correlates of emotion or attitude (Scherer, 1986). Thus, changes in pitch range might be less linguistically informative for English-learning infants. We might then predict that learners of a tonal language like Chinese may be more sensitive to the changes in pitch than those of a non-tonal language like English.

These results are broadly consistent with previous studies showing that infants gradually learn to disregard linguistically irrelevant acoustic variation in the word recognition process. For example, 7.5-month-old infants failed to recognize the same word when it differed in pitch level or vocal affect, or when it was spoken by talkers of different genders. However, by the end of the first year, infants were able to recognize the same word independent of these non-linguistic changes in the acoustic signal (Singh *et al.*, 2004; Singh *et al.*, 2008; Houston and Jusczyk, 2000). These findings,

together with results from the current study, suggest that infants' ability to recognize the same words, as well as their efficiency in recognizing target words, is affected only by lexically relevant acoustic cues.

The findings of the present study therefore suggest that IDS facilitates infants' word recognition by providing better perceptual cues and well-specified linguistic information compared to ADS, and not simply by drawing their attention to speech by means of exaggerated pitch range characteristics. As Soderstrom (2007) points out in a recent review of IDS, we have a very limited understanding of how infants' attention to IDS changes over time. Although a number of studies have demonstrated infants' strong listening preferences for IDS over ADS, infants in these studies were mostly under 9 months (Cooper and Aslin, 1990; Fernald, 1985; Werker and McLeod, 1989). More recent studies have shown that infants' attentional preference for IDS actually decreases with age. For instance, Newman and Hussain (2006) showed that 9-month-olds did not show a listening preference for IDS over ADS. Infants also begin to tune into the details of their native language structure around 9 months (Werker and Tees, 1984). Thus, it is likely that 19-month-olds are more sensitive to linguistically relevant acoustic characteristics than simply pitch in IDS.

Given that wide pitch range is effective in holding younger infants' attention in particular, it is possible that wide pitch range may enhance younger infants' word recognition, although it did not improve 19-month-olds' word recognition. In this case, we would find a discrepancy in behavior between younger infants (especially under 9 months old) and 19-month-olds. Alternatively, perhaps both younger infants and older infants use only lexically relevant acoustic cues when processing words even if younger infants generally pay attention to speech containing wide pitch range. Also, if older infants pay particular attention to the linguistic details contained in speech, then it would be helpful to speculate regarding how infants might use this information in different types of speech processing tasks (e.g., segmenting, disambiguating, learning phonemes), and how this might change over time. The results of present study therefore raise many interesting questions, opening up a discussion which could lead to a deeper understanding of early speech perception and its effects on language development.

A number of studies have shown that slow speaking rate, vowel hyper-articulation, and wide pitch range are typical characteristics of clear speech directed to various adult populations (Smiljanic and Bradlow, 2005). One might wonder how these three acoustic properties affect adult listeners' perception of speech. That is, do infants and adults show similar behavior? Findings on speaking rate are mixed: although slow speaking rate has been often shown to improve speech intelligibility for adult listeners (Bradlow and Bent, 2002), several studies have also shown no association between the two (Bradlow *et al.*, 1996). On the other hand, vowel hyper-articulation has been commonly found to be associated with improved intelligibility (Bradlow *et al.*, 1996). A few researchers have also found a tendency for wide pitch range to correlate positively with higher speech intelligibility scores rated by normal-hearing adult listeners (Bradlow *et al.*,

1996). As Smiljanić and Bradlow (2005) point out, however, it is generally not clear if pitch range directly affects speech intelligibility. This suggests that adults and infants as young as 19 months may be sensitive to similar acoustic cues when perceiving speech. Investigating listeners' abilities at various stages of development in various types of perception tasks may enrich our current understanding of the factors that facilitate word recognition. Overall, the findings from the present study demonstrate that IDS is a useful source of information that provides perceptual advantages for infants. In this regard, IDS can be seen as one kind of clear speech that enhances speech intelligibility, even for 19-month-old infants.

## V. CONCLUSION

Researchers have long suggested that the exaggerated acoustic properties of IDS might facilitate infants' language development. However, the details of this relationship have been less clear. The goal of this study was thus to identify the acoustic cues that facilitate 19-month-olds' word recognition, an age at which many new words are being learned daily. The results showed that slow speaking rate and vowel hyperarticulation significantly enhanced infants' ability to recognize words. On the other hand, infants' word recognition was not affected by pitch range, which is a lexically irrelevant cue in English. This suggests that IDS facilitates infants' word recognition by providing better-specified linguistic information, and not simply by attracting their attention to speech by means of wide pitch range. These findings provide a more in depth understanding of the mechanisms by which IDS facilitates infants' word recognition, which is an essential part of the word learning process.

## ACKNOWLEDGMENTS

We thank Sheila Blumstein, Lori Rolfe, Elena Tenenbaum, Yen-Liang Shue, Julie Sedivy, and Melanie Cabral for assistance at the various stages of this work. This work was supported by NSF Grant No. BCS-0544127 and NIH Grant No. R01MH60922 to Katherine Demuth, and NIH Grant No. R01HD32005 to James Morgan.

## APPENDIX A

Word pairs used in test stimuli. The *proportion* indicates the percentage of 16-month-old infants reported to understand the words in the MacArthur CDI.

	<i>Target</i>	<i>Proportion</i>	<i>Distractor</i>	<i>Proportion</i>
1	Bird	79	Dog	88
2	Block	69	Sock	85
3	Bunny	61	Puppy	61
4	Car	93	Ball	93
5	Cat	76	Hat	61
6	Cup	86	Book	90
7	Door	79	Chair	64
8	Flower	68	Apple	74
9	Hand	64	Bed	68
10	Kitty	78	Cookie	83
11	Spoon	75	Balloon	82
12	Truck	67	Duck	79

## APPENDIX B

Formant values of hyper-articulated vowels in Typical-IDS (*ID*), target hypo-articulated vowels (*AD target*), and the actual vowels that were used as Hypo-articulated-IDS stimuli (*AD used*).

Target word	<i>ID</i>	F1		F2		
		<i>AD target</i>	<i>AD used</i>	<i>ID</i>	<i>AD target</i>	<i>AD used</i>
Bird	883	806	705	1847	1926	1926
Block	857	788	744	1669	1797	1728
Bunny	816	758	788	1636	1773	1871
Bunny	318	397	343	3252	2942	2943
Car	1122	979	879	1809	1898	1830
Cat	1404	1183	1014	2095	2105	2066
Cup	834	771	763	1947	1998	1902
Door	830	768	486	1242	1488	1537
Flower	1173	1016	953	1961	2008	2079
Hand	902	820	510	2002	2038	1994
Kitty	690	667	520	2708	2549	2629
Kitty	335	410	454	3071	2811	2725
Spoon	500	529	475	1495	1671	1805
Truck	990	884	706	1624	1764	1867

<sup>1</sup>Calculation done by authors based on the data presented in Cooper and Aslin (1990).

<sup>2</sup>We did not correct for experiment-wise  $\alpha$  in these analyses. Doing so would have shifted the first period in which looking to the target significantly exceeded looking to the distractor to a later window. This would have applied to both stimulus types, and the relative pattern (which is of primary interest here) would have remained the same; significantly greater looking to the target would occur earlier for Typical-IDS stimuli than for Fast-IDS stimuli. Concerns about Type-I errors may be mitigated somewhat by the fact that once looking to the target became significantly greater than looking to the distractor (for either stimulus type), it remained so for all subsequent windows until the end of the trial, except as noted. These comments apply equally to all stimulus type time-course analyses.

- Bernstein Ratner, N. (1984). "Phonological rule usage in mother-child speech," *J. Phonetics* **12**, 245–254.
- Bernstein Ratner, N. (1986). "Durational cues which mark clause boundaries in mother-child speech," *J. Phonetics* **14**, 303–309.
- Bernstein Ratner, N., and Pye, C. (1984). "Higher pitch in babyltalk is not universal: Acoustic evidence from Quiche Mayan," *J. Child Lang* **11**, 515–522.
- Berry, M. D., and Erickson, R. L. (1973). "Speaking rate: Effects on children's comprehension of normal speech," *J. Speech Hear. Res.* **16**, 367–374.
- Boersma, P., and Weenink, D. (2005). PRAAT: Doing phonetics by computer (Version 4.4.07) retrieved from <http://www.praat.org/> (Last viewed 3/1/2006).
- Bradlow, A. R., and Bent, T. (2002). "The clear speech effect for non-native listeners," *J. Acoust. Soc. Am.* **112**, 272–284.
- Bradlow, A. R., Kraus, N., and Hayes, E. (2003). "Speaking clearly for learning-impaired children: Sentence perception in noise," *J. Speech Lang. Hear. Res.* **46**, 80–97.
- Bradlow, A. R., and Pisoni, D. B. (1999). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," *J. Acoust. Soc. Am.* **106**, 2074–2085.
- Bradlow, A. R., Torretta, G. M., and Pisoni, D. B. (1996). "Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics," *Speech Commun.* **20**, 255–272.
- Burnham, D., Kitamura, C., and Vollmer-Conna, U. (2002). "What's new pussycat? On talking to animals and babies," *Science* **296**, 1435.
- Colombo, J., Frick, J. E., Ryther, J. S., Coldren, J. T., and Mitchell, D. W. (1995). "Infants' detection of analogs of "motherese" in noise," *Merrill-Palmer Q.* **41**, 104–113.
- Cooper, R. P., Abraham, J., Berman, S., and Staska, M. (1997). "The devel-

- opment of infants' preference for motherese," *Infant Behav. Dev.* **20**, 477–488.
- Cooper, R. P., and Aslin, R. N. (1990). "Preference for infant-directed speech in the first month after birth," *Child Dev.* **61**, 1584–1595.
- Cooper, R. P., and Aslin, R. N. (1994). "Developmental differences in infant attention to the spectral properties of infant-directed speech," *Child Dev.* **65**, 1663–1677.
- Dale, P. S., and Fenson, L. (1996). "Lexical development norms for young children," *Behav. Res. Methods Instrum. Comput.* **28**, 125–127.
- Dunn, J., and Kendrick, C. (1982). "The speech of two- and three-year-olds to infant siblings: 'Baby talk' and the context of communication," *J. Child Lang.* **9**, 579–595.
- Fernald, A. (1985). "Four-month-old infants prefer to listen to motherese," *Infant Behav. Dev.* **8**, 181–195.
- Fernald, A. (2000). "Speech to infants as hyperspeech: Knowledge-driven processes in early word recognition," *Phonetica* **57**, 242–254.
- Fernald, A., and Kuhl, P. K. (1987). "Acoustic determinants of infant preference for motherese speech," *Infant Behav. Dev.* **10**, 279–293.
- Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., and McRoberts, G. (1998). "Rapid gains in speed of verbal processing by infants in the second year," *Psychol. Sci.* **9**, 228–231.
- Fernald, A., and Simon, T. (1984). "Expanded intonation contours in mothers' speech to newborns," *Dev. Psychol.* **20**, 104–113.
- Fernald, A., Taeschner, T., Dunn, J., Papousek, M., De Boysson-Bardies, B., and Fukui, I. (1989). "A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants," *J. Child Lang.* **16**, 477–501.
- Fernald, A., Zangl, R., Portillo, A. L., and Marchman, V. A. (2008). "Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children," in *Developmental Psycholinguistics: On-Line Methods in Children's language processing*, edited by I. A. Sekerina, E. M. Fernandez, and H. Clahsen (John Benjamins, Amsterdam, The Netherlands), pp. 97–135.
- Gay, T. (1978). "Effect of speaking rate on vowel formant movements," *J. Acoust. Soc. Am.* **63**, 223–230.
- Goldfield, B. A., and Reznick, J. S. (1990). "Early lexical acquisition: Rate, content, and the vocabulary spurt," *J. Child Lang.* **17**, 171–183.
- Gout, A., Christophe, A., and Dupoux, E. (2002). "Testing infants' discrimination with the orientation latency procedure," *Infancy* **3**, 249–259.
- Grieser, D. L., and Kuhl, P. K. (1988). "Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese," *Dev. Psychol.* **24**, 14–20.
- Hirsh-Pasek, K., and Golinkoff, R. M. (1996). *The Origins of Grammar: Evidence From Early Language Comprehension* (MIT, Cambridge, MA).
- Hollich, G. (2003). SUPERCODER (Version 1.5) retrieved from <http://hincapie.psych.purdue.edu/Splitscreen/SuperCoder.dmg> (Last viewed 8/1/2006).
- Houston, D. M., and Jusczyk, P. W. (2000). "The role of talker-specific information in word segmentation by infants," *J. Exp. Psychol. Hum. Percept. Perform.* **26**, 1570–1582.
- Ingram, D. (1995). "The cultural basis of prosodic modifications to infants and children: A response to Fernald's universalist theory," *J. Child Lang.* **22**, 223–233.
- Jacobson, J. L., Boersma, D. C., Fields, R. B., and Olson, K. L. (1983). "Paralinguistic features of adult speech to infants and small children," *Child Dev.* **54**, 436–442.
- Johnson, K. (2004). "Massive reduction in conversational American English," in *Spontaneous Speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, edited by K. Yoneyama and K. Maekawa (The National International Institute for Japanese Language, Tokyo, Japan), pp. 29–54.
- Karzon, R. G. (1985). "Discrimination of polysyllabic sequences by one-to four-month-old infants," *J. Exp. Child Psychol.* **39**, 326–342.
- Kemler Nelson, D. G., Hirsh-Pasek, K., Jusczyk, P. W., and Cassidy, K. W. (1989). "How the prosodic cues in motherese might assist language learning," *J. Child Lang.* **16**, 55–68.
- Kitamura, C., and Burnham, D. (1998). "The infant's response to maternal vocal affect," in *Advances in Infancy Research*, Vol. **12**, edited by C. Rovee-Collier, L. Lipsitt, and H. Hayne (Ablex, Stamford, CT), pp. 221–236.
- Kitamura, C., Thanavishuth, C., Burnham, D., and Luksaneeyanawin, S. (2001). "Universality and specificity in infant-directed speech: Pitch modifications as a function of infant age and sex in a tonal and non-tonal language," *Infant Behav. Dev.* **24**, 372–392.
- Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., and Lacerda, F. (1997). "Cross-language analysis of phonetic units in language addressed to infants," *Science* **277**, 684–686.
- Liu, H.-M., Kuhl, P. K., and Tsao, F.-M. (2003). "An association between mothers' speech clarity and infants' speech discrimination skills," *Dev. Sci.* **6**, F1–F10.
- Liu, H.-M., Tsao, F.-M., and Kuhl, P. K. (2005). "The effect of reduced vowel working space on speech intelligibility in Mandarin-speaking young adults with cerebral palsy," *J. Acoust. Soc. Am.* **117**, 3879–3889.
- Malsheen, B. J. (1980). "Two hypotheses for phonetic clarification in the speech of mothers to children," in *Child Phonology. Volume 2: Perception*, edited by G. Yeni-Komshian, J. Kaveanagh, and C. Ferguson (Academic, New York), pp. 173–184.
- Morgan, J. L. (1986). *From Simple Input to Complex Grammar* (MIT, Cambridge, MA).
- Nelson, N. W. (1976). "Comprehension of spoken language by normal children as a function of speaking rate, sentence difficulty, and listener age and sex," *Child Dev.* **47**, 299–303.
- Newman, R. S., and Hussain, I. (2006). "Changes in preference for infant-directed speech in low and moderate noise by 5-to 13-month-olds," *Infancy* **10**, 61–76.
- Payton, K. L., Uchanski, R. M., and Braida, L. D. (1994). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," *J. Acoust. Soc. Am.* **95**, 1581–1592.
- Picheny, M. A., Durlach, N. I., and Braida, L. D. (1985). "Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech," *J. Speech Hear. Res.* **28**, 96–103.
- Scherer, K. R. (1986). "Vocal affect expression: A review and a model for future research," *Psychol. Bull.* **99**, 143–165.
- Singh, L., Morgan, J., and White, K. (2004). "Preference and processing: The role of speech affect in early spoken word recognition," *J. Mem. Lang.* **51**, 173–189.
- Singh, L., Morgan, J. L., and Best, C. T. (2002). "Infants' listening preferences: Baby talk or happy talk?" *Infancy* **3**, 365–394.
- Singh, L., White, K., and Morgan, J. L. (2008). "Building a word-form lexicon in the face of variable input: Influences of pitch and amplitude on early spoken word recognition," *Lang. Learn. Dev.* **4**, 157–178.
- Sjölander, K., and Beskow, J. (2000). "Wavesurfer—An open source speech tool," in *Proceedings of the ICSLP 2000*, Beijing, China, Vol. **4**, pp. 464–467.
- Smiljanic, R., and Bradlow, A. R. (2005). "Production and perception of clear speech in Croatian and English," *J. Acoust. Soc. Am.* **118**, 1677–1688.
- Soderstrom, M. (2007). "Beyond babyltalk: Re-evaluating the nature and content of speech input to preverbal infants," *Dev. Rev.* **27**, 501–532.
- Stern, D. N., Spieker, S., Barnett, R. K., and MacKain, K. (1983). "The prosody of maternal speech: Infant age and context related changes," *J. Child Lang.* **10**, 1–15.
- Swingley, D., and Fernald, A. (2002). "Recognition of words referring to present and absent objects by 24-month-olds," *J. Mem. Lang.* **46**, 39–56.
- Thiessen, E. D., Hill, E. A., and Saffran, J. R. (2005). "Infant-directed speech facilitates word segmentation," *Infancy* **7**, 53–71.
- Trainor, L. J., and Desjardins, R. N. (2002). "Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels," *Psychon. Bull. Rev.* **9**, 335–340.
- Uchanski, R. M., Choi, S., Braida, L. D., Reed, C. M., and Durlach, N. I. (1996). "Speaking clearly for the hard of hearing IV: Further studies of the role of speaking rate," *J. Speech Hear. Res.* **39**, 494–509.
- Werker, J. F., and McLeod, P. (1989). "Infant preference for both male and female infant directed talk: A developmental study of attentional and affective responsiveness," *Can. J. Psychol.* **43**, 230–246.
- Werker, J. F., Pons, F., Dietrich, C., Kajikawa, S., Fais, L., and Amano, S. (2007). "Infant-directed speech supports phonetic category learning in English and Japanese," *Cognition* **103**, 147–162.
- Werker, J. F., and Tees, R. (1984). "Cross-language speech perception: evidence for perceptual reorganization during the first year of life," *Infant Behav. Dev.* **7**, 49–63.
- Zangl, R., Klarman, L., Thal, D., Fernald, A., and Bates, E. (2005). "Dynamics of word comprehension in infancy: Developments in timing, accuracy, and resistance to acoustic degradation," *J. Cogn. Dev.* **6**, 179–208.