Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations

Chandan R. Narayan, and Lily C. McDermott

Citation: The Journal of the Acoustical Society of America **139**, 1272 (2016); doi: 10.1121/1.4944634
View online: https://doi.org/10.1121/1.4944634
View Table of Contents: https://asa.scitation.org/toc/jas/139/3
Published by the Acoustical Society of America

Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations

Chandan R. Narayan[a] and Lily C. McDermott
Speech and Psycholinguistics Laboratory, York University, Toronto, Ontario M3J 1P3, Canada

(Received 31 December 2014; revised 24 January 2016; accepted 27 February 2016; published online 23 March 2016)

The speech rate and pitch (F_0) characteristics of naturalistic, longitudinally recorded infant- and adult-directed speech are reported for three, genetically diverse languages. Previous research has suggested that the prosodic characteristics of infant-directed speech are slowed speech rate, raised mean pitch, and expanded pitch range relative to adult-directed speech. Sixteen mothers (5 Sri Lankan Tamil, 5 Tagalog, 6 Korean) were recorded in their homes during natural interactions with their young infants, and adults, over the course of 12 months beginning when the infant was 4 months old. Regression models indicated that the difference between infant- and adult-directed speech rates decreased across the first year of infants' development. Models of pitch revealed predicted differences between infant- and adult-directed speech but did not provide evidence for cross-linguistic or longitudinal effects within the time period investigated for the three languages. The universality of slowed speech rate, raised pitch, and expanded pitch range is discussed in light of individuals' highly variable implementation of these prosodic features in infant-directed speech.
 [http://dx.doi.org/10.1121/1.4944634]

[LK] Pages: 1272–1281

I. INTRODUCTION

Infant-directed speech (IDS) is the broadly defined linguistic register used by caregivers when interacting with young infants. The acoustic characteristics of IDS [relative to adult-directed speech (ADS)] encompass a variety of segmental and prosodic features (see Cruttenden, 1994, for a review), including higher overall pitch (F_0) (Fernald *et al.*, 1989) and wider pitch range (e.g., Garnica, 1977; Stern *et al.*, 1983; Fernald and Simon, 1984; Fernald *et al.*, 1989), longer pauses between phrases (Stern *et al.*, 1983; Fernald *et al.*, 1989), shorter utterances, slowed speech rate (Fernald and Simon, 1984; Cooper and Aslin, 1990; Tang and Maidment, 1996), expanded vowel space (Kuhl *et al.*, 1997), and less overlap between vowel qualities in $F1 \times F2$ space and less overlap in vowel duration cues in languages with phonological length (Werker *et al.*, 2007). These characteristics have been documented in genetically related and unrelated languages such as French, Italian, German, English (Fernald *et al.*, 1989), Japanese (Fernald *et al.*, 1989; Werker *et al.*, 2007; Martin *et al.*, 2015), Mandarin (Liu *et al.*, 2007), and Thai (Kitamura *et al.*, 2002) to name a few.

A prominent theme in the IDS literature is the issue of acoustic clarity (measured as the amount of separation between speech categories in perceptually relevant acoustic space) in the speech to infants. Many studies show enhancement of segmental differences (e.g., Kuhl *et al.*, 1997; Werker *et al.*, 2007; Lee *et al.*, 2008), which are interpreted as advantageous from a language acquisition perspective (e.g., Karzon, 1985; Liu *et al.*, 2003), while other studies suggest phonetic segments in IDS may not be so clearly produced (e.g., McMurray *et al.*, 2013; Martin *et al.*, 2015). The general prosodic modifications of IDS implicated in the acoustic clarity literature (slowed speech rate, long pauses, expanded pitch contours, etc.) have been linked to language learning and are preferred by infants (Fernald and Simon, 1984; Fernald and Kuhl, 1987; Cooper and Aslin, 1990), with younger infants (under 6 months) showing more attentional response to IDS than do infants at 9 months (Werker and McLeod, 1989). The "prosodic bootstrapping" hypothesis suggests that exaggerated prosodic cues may provide language learners with segmentation information that can serve as a basis for syntactic category development (e.g., Kemler Nelson *et al.*, 1989), or serve as an implicit word teaching strategy (Woodward and Aslin, 1990). There is correlational evidence that the characteristics of this prosodic enhancement serve a facilitatory role in infants' developing grammatical and speech perception systems (Liu *et al.*, 2003), occurring at a time in development when rapid changes are affecting an infant's perceptual system (e.g., Werker and Tees, 1984).

The acoustic clarity literature is difficult to interpret as much of this research presents IDS at unique and often differing moments in an infant's development. For example, mothers in the cross-linguistic cardinal vowel study by Kuhl *et al.* (1997) spoke to 2–5 months old infants, mothers in the English voice-onset time study by McMurray *et al.* (2013), who spoke to 9- to 13-month-olds, and mothers in the comprehensive Japanese vowel and consonant study by Martin *et al.* (2015) who spoke to 18- to 24-month-olds. The literature on the phonetic characteristics of IDS recognized that the register is not rigid, but rather a dynamic phenomenon that changes according to the age and linguistic abilities of the infant (Shockey and Bond, 1980; Stern *et al.*, 1983;

[a]Electronic mail: chandann@yorku.ca

 0001-4966/2016/139(3)/1272/10/$30.00

Soderstrom, 2007). While a number of studies have examined phonetic properties characterizing IDS by collapsing varying time points in an infant's development, fewer have done so from a longitudinal and developmental perspective, thereby capturing the nature of how a caregiver changes her speech to accommodate a developing infant. For example, in a cross-sectional study, Bernstein Ratner (1984) found that English IDS vowel quality in content words showed more clarification (separation between vowel categories in $F1 \times F2$ space) in speech to older children than preverbal or holophrastic-stage children.

The purpose of the present study is to examine the stereotypical prosodic features of IDS (slowed speech rate,[1] raised pitch height and expanded pitch range) from a developmental perspective in a longitudinal design. At the most general level we ask whether these prosodic features of IDS are subject to cross-linguistic variation in their implementation. We also ask whether a caregiver's prosody changes over the course of an infant's first year, and if it does, whether individuals implement the changes in a similar way. We hope to highlight intralanguage variation in the use of speech rate and pitch features in speech addressed to young infants, thereby refining our understanding of prosodic modification in IDS.

Very few studies have directly examined speech rate in IDS, and none have done so in a longitudinal design over the first year of infancy. Fernald and Simon (1984) found that German-speaking mothers addressed their newborn infants significantly more slowly (4.8 syllables/s) than when addressing adults (5.8 syllables/s). In the study by Bernstein Ratner (1985), spontaneous speech was 25% slower in English IDS to 17-month-olds compared with ADS. Similarly, Tang and Maidment (1996) found Cantonese-speaking mothers spoke significantly slower to their 12- to 20-month-old infants than to adults (approximately 40% fewer syllables/min when talking to infants versus adults).

Prosodic modifications involving elevated pitch height and expanded pitch range have been studied in a variety of languages. In a cross-sectional design examining speech to older English-hearing infants and children, Garnica (1977) found that mothers' speech to their 2-year-olds had a higher mean pitch and wider pitch range than mothers speaking to their 5-year-olds. Stern *et al.* (1983) examined the prosodic features of English IDS in a longitudinal design (mothers speaking to their newborn, 4-, 12-, and 24-month-old infants) and found that pitch range was greatest in speech to 4-month-olds relative to newborns, 12- and 24-month-olds. Grieser and Kuhl (1988) showed that Mandarin IDS to 2-month-olds had a higher mean pitch and wider pitch range compared to Mandarin ADS. In their longitudinal examination of the development of Thai and Australian English IDS pitch characteristics, Kitamura *et al.* (2002) found that mean pitch was higher in IDS than ADS, but pitch range was not different between the two registers for both languages. The authors found that mean pitch in both Australian English and Thai IDS followed a quadratic developmental trajectory, increasing until infants were 6 (English) or 9 (Thai) months, then falling towards the end of the infant's first year. Although prosodic modification in IDS is not universal (Bernstein Ratner and Pye, 1984), there is some evidence that languages that do exhibit these features in IDS vary in their implementation of raised pitch height and expanded pitch range. Fernald *et al.* (1989) found less pitch range expansion in Japanese IDS to 10- to 14-month-old infants than in the IDS of Germanic languages.

The present study examines the *developmental* nature of three prosodic features of IDS, namely, speech rate, mean pitch, and pitch range from both a cross-linguistic and longitudinal perspective. We examine these prosodic features in IDS relative to ADS at distinct times over the course of an infant's development in speakers of Sri Lankan Tamil, Tagalog, and Korean, highlighting individual variation in the implementation of these features both within and between languages. Our study is differentiated from many acoustic descriptions of IDS in that we present age-yoked comparisons of individuals' IDS and ADS at each moment in an infant's development, rather than a comparison of a singular ADS end state (e.g., Fernald *et al.*, 1989; Kitamura *et al.*, 2002). Last, the present study features IDS from genetically diverse languages that are under-represented in the IDS literature. This study aims to present a comprehensive description of these prosodic features of IDS as they unfold over the course of an infant's first year of life.

II. METHODS

A. Recordings

All speech samples in the present study were taken from the Cross-Linguistic Corpus of Infant-Directed speech (CCIDS). The CCIDS contains high-quality audio recordings of 16 mothers [5 Sri Lankan Tamil (SLT), 5 Tagalog, and 6 Korean] interacting with their infants in their own homes over the course of the first year of the infant's development. Mothers were recruited for participation in the year-long study through university student groups in Toronto. They were informed of the purpose of the corpus and consent was obtained according to university standard procedures. A questionnaire was administered to each participant before acceptance into the CCIDS recording program. The questionnaire assessed language use and socio-economic measures (education, employment, etc.) for each household. Each participant was between the ages of 25 and 35 at the time of recording and a relatively new immigrant to Canada, having lived in Toronto for less than 5 years and having previously lived in Sri Lanka, Korea, or the Philippines. Participants were native speakers of the target languages and conversed with their spouses, family members, and close friends in the target language on a daily basis. The five Filipino participants in the corpus used the standard variety of Tagalog spoken in Manila, Philippines. Three of these speakers were trilingual (along with English, two spoke Tagalog and Cebuano, and one spoke Tagalog and Kampampangan) but had spent the majority of their lives in Manila and spoke exclusively in Tagalog with friends and family. All 6 Korean participants spoke the Seoul variety. Participants who dropped out of the study before the third recording were replaced. The CCIDS project set out to record participants once a month for 12–14 months beginning when the infant was 4 months old. Because of the sensitive nature of in-home visits and unforeseen circumstances (i.e., travel, illness, etc.) not every participant was recorded monthly, but at least once during any two consecutive months until the infant was 15–16 months old. No participant missed two consecutive recording sessions.

Recording sessions lasted approximately 1 h and generally commenced in the late afternoon (for older children) or late morning (for younger children) after the child had napped and had been fed. Participants were outfitted with a wireless lapel microphone (Line 6 XD-V30L), which transmitted its signal to a portable digital audio recorder (Zoom H4n). Recordings were made in .wav format at a sampling rate of 44.1 kHz.

The mothers in the study were instructed to interact with their infant in any way that felt natural and comfortable to them while trying to keep auditory distractions (e.g., television, radio, noisy toys, etc.) to a minimum. The naturalistic aspect of these instructions led to a wide variety of input quantity, as some mothers spoke more than others. No special effort was made to minimize momentary sounds such as a ringing phone or whistling teapot. The mother-infant interaction generally involved feeding/nursing, joint attention to toys, describing books without scripts (provided to the participant by the researcher), dressing, and more unstructured forms of play (such as tickling). A research assistant and recordist were in a different room, out of sight from the participant and infant. Infant-directed speech was recorded in this way for 45 min during each session. Each IDS session was followed by a 10 min ADS recording session, wherein the research assistant, herself a native speaker of either SLT, Tagalog, or Korean would engage with the participant in small talk concerning the infant. In general, participants interacted with one or two native-language research assistants over the course of the 12-month recording period. Not all participants completed the entire recording period, and some participants started recording sessions when their infant was 6 months old.[2] Participants were compensated $35 CDN for each recording session.

B. CCIDS languages

The languages of the CCIDS are genetically diverse: SLT (Dravidian), Tagalog (Austronesian), and Korean (Isolate). All three languages are syllable-timed. While there have been a few studies of Korean IDS (Lee *et al.*, 2008; Narayan and Yoon, 2011), SLT (and continental Tamil), and Tagalog are underrepresented in the IDS literature. Notable features of the CCIDS languages germane to the present study are the nature of their syllable structure and unique aspects of intonation and tone. Tamil has contrastive vowel length and obligatory onsets in syllables. Syllables in Tamil are minimally consonant (C) vowel (V), and maximally CVVCC (Christdas, 1988). Word-level F_0 prominences in Tamil (in the form of rising F_0 contours) can be considered pitch accent (Keane, 2006). Tagalog syllables have a CV(C) structure, with most roots being disyllabic (Zuraw, 2000). Korean is a pitch accent language, where syllables are minimally V and maximally CVC (Yoon, 1995). The CCIDS languages are non-tonal, though Seoul Korean (the variety spoken by the Korean mothers in the corpus) is undergoing a sound change whereby the three-way laryngeal contrast is implemented as a tonal distinction (Kang and Han, 2013).

C. Measurements

Each hour-long IDS/ADS recording was divided into ten 5-min samples. Researchers trained in phonetics and acoustic analysis methods in Praat (Boersma and Weenink, 2009) selected single phrase utterances that were approximately 5 s long without silences more than 300 ms [$n = 1603$, $M = 4.68$ s, standard deviation (SD) $= 1.24$ s, Coefficient of variability (CV) $= 0.26$] devoid of infant sounds, maternal singing, or non-speech intrusions (phone ringing, toy sounding, etc.), from each audio sample for speaking rate calculation. Utterance duration between the two registers was equated to the extent possible ($M_{\text{IDS}} = 4.33$ s, $M_{\text{ADS}} = 4.58$ s, $t = -0.42$, NS) to avoid utterance length effects in the calculation of speech rate as IDS is often characterized by overall shorter utterances relative to ADS (see Soderstrom, 2007, for review). Ten minutes of audio from a given recording session (two 5-minute samples) were first manually coded by a trained phonetician, whose syllable count was then compared to the automatic detection. Syllable count was then extracted automatically using a speech-rate script for Praat (de Jong and Wempe, 2009). The speech-rate script detected syllabic nuclei according to changes in intensity (dB) of vocalic energy. Using two Dutch corpora, de Jong and Wempe (2009) showed correlations between automatic speech rate extraction and human measurements of 0.71–0.88. For the CCIDS, if reliability between human and automatic detection was less than 75%, the dB threshold parameter of the script was changed to increase reliability to at least 85%. These parameters were then applied to the entire set of single-utterance audio files from a particular recording session for syllable count extraction. The syllable detection script has three parameters: silence threshold, minimum dip between peaks, and minimum pause duration. Silence threshold and minimum pause duration remained at default settings of −25 dB and 300 ms, respectively, for the entire data processing procedure. The minimum-dip-between-peaks parameter was adjusted on a recording-by-recording basis. The average minimum dB dip for syllable nucleus detection was 1.3 dB in the CCIDS corpus. In general, the default dB dip parameter (2 dB) under reported the true number of syllables in speech files. In the cases where automatic syllable counts yielded outliers (greater than or less than 3 standard deviations from the mean), those audio files were examined by a trained phonetician who manually counted the number of syllables in the file (12% of single utterance files).

Pitch (F_0) measurements were likewise extracted from Praat using an autocorrelation method. As with the speech rate measurements, pitch measurements were made on the entirety of each utterance 5-s file of connected speech and included periodic information from both sonorant consonants and vowels. The default pitch floor and ceiling settings of 75 and 500 Hz, respectively, were often raised to accommodate higher pitched individuals in the corpus and to remedy halving (where two periods are treated as one) and

TABLE I. Mean speech rates (syllables/second) for IDS and ADS by language.

Language	Register	Speech rate (SD, CV)
Tagalog	ADS	5.26 (0.75, 0.14)
	IDS	4.45 (0.95, 0.21)
SL Tamil	ADS	5.89 (0.78, 0.13)
	IDS	5.62 (1.17, 0.21)
Korean	ADS	5.66 (0.76, 0.13)
	IDS	4.64 (1.01, 0.22)

doubling errors. Pitch floor settings ranged from 75 to 150 Hz and pitch ceiling settings ranged from 500 to 700 Hz. All other pitch parameters were left at their default settings.[3] In approximately 5% of the adult-directed speech samples, speakers produced creaky voice (as expected at the ends of phrases). These brief portions of the speech sample were excluded from analysis. The automatically extracted pitch measurements considered for analysis included: mean pitch and pitch range. Samples were binned into six time epochs representing infant ages of 0;4–0;5.30, 0;6–0;7.30, 0;8–0;9.30, 0;10–0;11.30, 1;0–1;1.30, 1;2–1;3.30 (ages represent years;months.days). This was done to normalize the time dimension (into 2-month epochs), as there was variability in the times at which mothers were recorded.

III. RESULTS

A. Speech rate

Table I gives mean speech rate in the different languages for the two registers across time. SLT-speaking mothers had a faster speaking rate ($M = 5.76$ syllable/s, $SD = 0.53$, $CV = 0.09$) than either Tagalog-speaking ($M = 5.10$ syllable/s, $SD = 0.65$, $CV = 0.13$) or Korean-speaking ($M = 4.91$ syllable/s, $SD = 0.59$, $CV = 0.12$) mothers, and across all three languages, ADS ($M = 5.57$ syllable/s, $SD = 0.52$, $CV = 0.09$) was faster than IDS ($M = 4.9$ syllable/s, $SD = 0.7$, $CV = 0.14$).

Means of multiple measurements at each time epoch are given per individual participant in Fig. 1. Time epochs represent measurements made over the course of 2 months.

Mixed-effects linear regression models were fit to the longitudinal data using the *lmer* package in the R statistical programming environment. Mixed-effects regression models are considered best suited for these data as they accommodate unbalanced observations between subjects as well as variability in the time epochs, restrictions that would violate basic assumptions of repeated measures designs. Repeated measures models assume compound symmetry where all variances and covariances are equal and are generally less valid in longitudinal designs were subject measurements might not be taken on a strict schedule. As speakers varied with respect to the schedule on which recordings were made as well as their total number of recordings, another variable (Occasion) was created. The Occasion variable quantified the total number of observations recorded for a given speaker at a particular epoch. For example, speaker one may have had 15 observations from time 1, while speaker 2 may have had 25. The Occasion variable thereby accounted for correlations among observations within any speaker at a particular time. Speech rate was first modeled in a fully crossed linear mixed-effects model with fixed predictors Register (IDS, ADS), Language (SLT, Tagalog, Korean), and Time, and random effects (intercepts) of Speaker and Occasion. A Wald test revealed that removing the three-way interaction term in the model did not significantly affect the overall fit. The resulting model, with Korean and ADS as the reference language and register, is shown in Eq. (1) below, where u

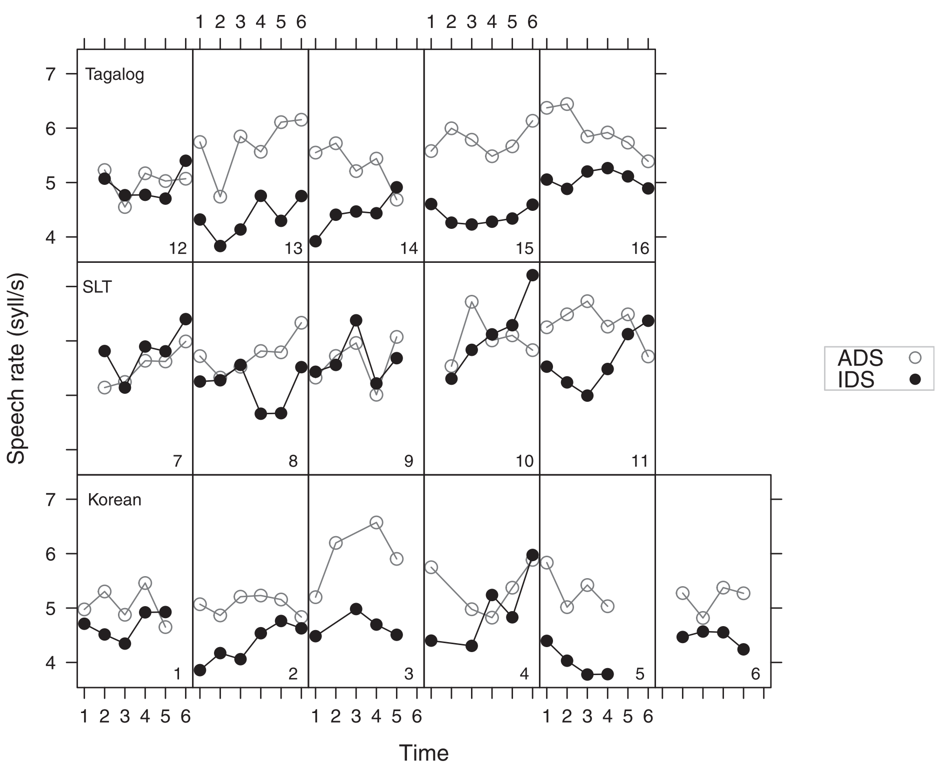

FIG. 1. Individual speaking rates (syllables/s) in both IDS and ADS across infant development. Six time epochs encompass infant ages from 4 to 15 months. Rows represent languages. Individual speaker numbers are given in lower right corner of each plot. Note that not all speakers made recordings at every time epoch.

TABLE II. Linear mixed-effects model of speech rate.

Predictor	β	*SE*	df	*t*	*p*
Intercept	5.31	0.22	1515	23.73	<0.0001
Register	−1.12	0.16	1515	−6.84	<0.001
Lang_{SLT}	0.36	0.28	13	1.25	0.23
$\text{Lang}_{\text{Tagalog}}$	0.40	0.29	13	1.41	0.18
Time	−0.01	0.06	65	−0.10	0.92
Reg × Lang_{SLT}	0.55	0.14	1515	4.00	0.0001
Reg × $\text{Lang}_{\text{Tagalog}}$	−0.21	0.14	1515	−1.47	0.14
Reg × Time	0.09	0.04	1515	2.35	0.02
Lang_{SLT} × Time	0.06	0.07	65	0.93	0.36
$\text{Lang}_{\text{Tagalog}}$ × Time	−0.01	0.07	65	−0.21	0.84

and ε, represent the random effects term and the error term, respectively,

$$E(y) = \beta_0 + \beta_{1(\text{Reg})} + \beta_{2(\text{LangSLT})} + \beta_{3(\text{LangTagalog})} + \beta_{4(\text{Time})} + \beta_{5(\text{Reg}\times\text{SLT})} + \beta_{6(\text{Reg}\times\text{Tagalog})} + \beta_{7(\text{Reg}\times\text{Time})} + \beta_{8(\text{SLT}\times\text{Time})} + \beta_{9(\text{Tagalog}\times\text{Time})} + \mu + \varepsilon. \quad (1)$$

The results of the regression model for speech rate are given in Table II.

The model showed a significant main effect of Register, which can be interpreted in conjunction with its significant interactions with Language (SLT) and Time.[4] Figure 2 shows the significant Register × Time interaction, which indicates that the difference in IDS and ADS speech rate across the three languages decreases as time increases. The effect size for the interaction is small (cf. $\beta = 0.09$)[5] and may be driven by Register × Language interaction indicating that the IDS-ADS difference is less in SLT than either Korean and Tagalog (see below). The effect sizes of the register differences quantify the interaction, indicating less of a register effect with increasing infant age ($d_{\text{Time1}} = 0.69$;

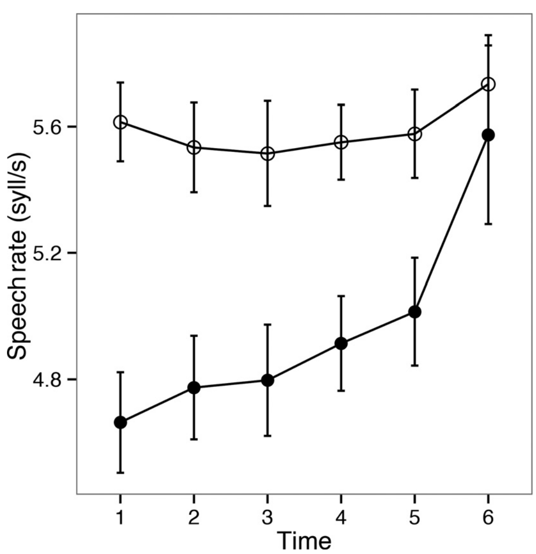

FIG. 2. Mean speech rate (syllables/s) in IDS and ADS across the first year of development (infant ages 4-15 months). Error bars represent standard error.

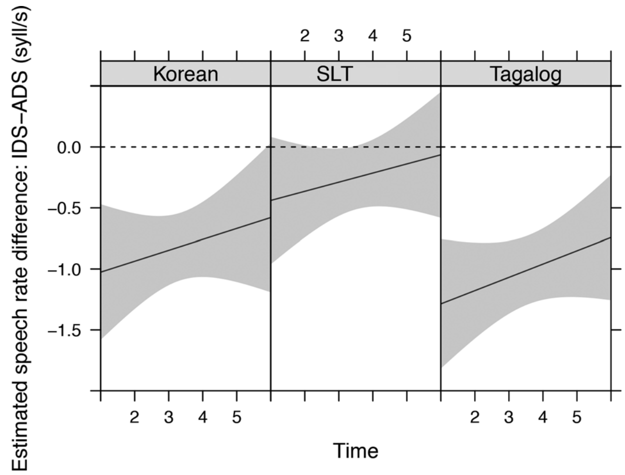

FIG. 3. Estimated difference in speech rate between IDS and ADS across time based on prediction from a mixed-effects linear regression model. Evidence of a predicted difference between IDS and ADS speech rate is taken where Scheffé confidence bands do not cross the horizontal axis.

$d_{\text{Time2}} = 0.56$; $d_{\text{Time3}} = 0.56$; $d_{\text{Time4}} = 0.51$; $d_{\text{Time5}} = 0.40$; $d_{\text{Time6}} = 0.25$).

The Register × $\text{Language}_{\text{SLT}}$ interaction indicates that implementation of register effects on speech rate in SLT is different from both Korean and Tagalog. This becomes clear when the model is used to predict difference estimations between IDS and ADS speech rate in the three languages (Fig. 3). The model predicts that IDS speech rate is very similar to ADS speech rate in SLT across time. The model's prediction is reinforced by individual participants' mean speech rates in Fig. 1, where three of the five SLT speakers (seven, nine, ten) show very little speech rate difference between the IDS and ADS registers. Across the three languages, the model predicts that difference between IDS and ADS speech rate decreases by 0.082 syllables/s per unit of Time [standard error $(SE) = 0.04$, $df = 1513$, $t = 2.12$, $p < 0.05$]. IDS speech rate increased an average of 0.13 syllables/s per unit of Time. Just noticeable differences for speech tempo have been reported between 4.43% (Eefting and Reitveld, 1989) and 10% (Quené, 2007). The IDS difference between Time 1 to Time 5 is 6% and 14% from Time 1 to Time 6.

B. Pitch

1. Mean pitch

Mean pitch was higher in IDS ($M = 266.61$ Hz, $SD = 57.87$, $CV = 0.22$) than in ADS ($M = 237.45$, $SD = 35.24$, $CV = 0.15$), a difference of 1.98 semitones ($SD = 0.45$, $CV = 0.23$). Figure 4 presents individual speakers' mean IDS and ADS pitch at six time epochs.

Table III gives mean pitch measurements according to language and register.

Following Kitamura *et al.* (2002), who found a quadratic effect of time on mean pitch in Australian English and Thai IDS, we first modeled mean pitch with a quadratic Time term in Eq. (1). The quadratic model was followed by a model with a linear Time term. The linear model fit the data better than the quadratic model ($\text{AIC}_{\text{quad}} = 16\,757.85$, $\text{AIC}_{\text{lin}} = 16\,736.50$,

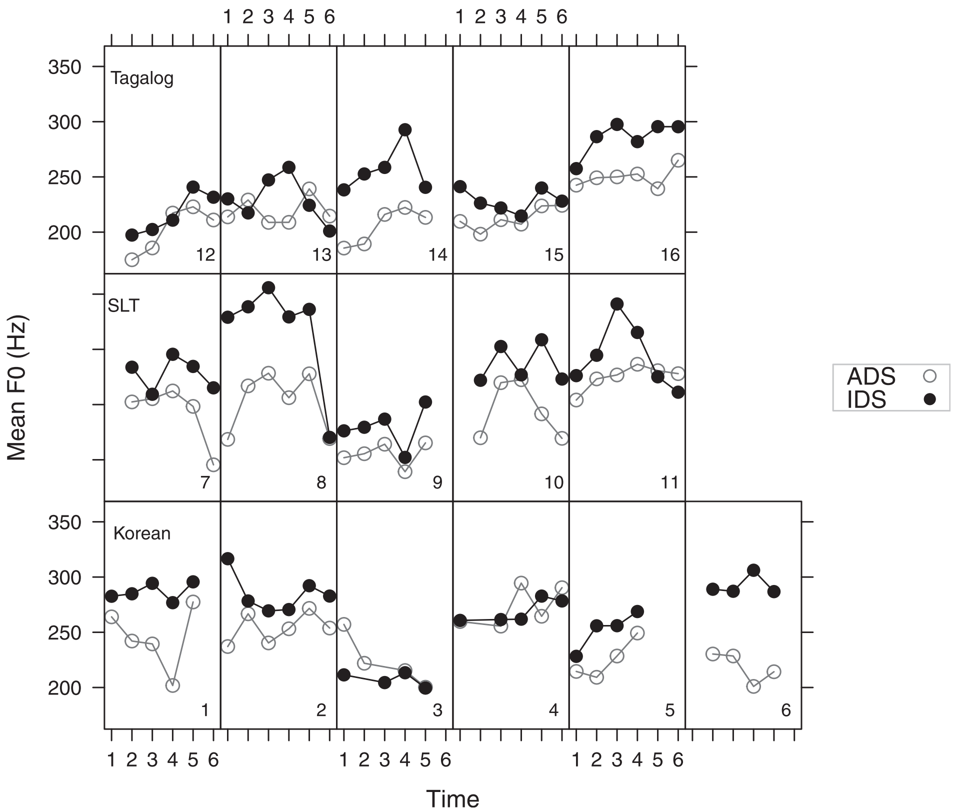

FIG. 4. Individuals' mean pitch (Hz) in both IDS and ADS across infant development. Time epochs encompass infant ages from 4 to 15 months. Rows represent languages. Individual speaker numbers are given in lower right corner of each plot. Note that not all speakers made recordings at every time epoch.

LRT = 31.35, $p < 0.001$). Using the linear model ($R^2 = 0.40$) speakers' mean pitch showed only a main effect of Register ($\beta = 34.74$, $SE = 7.63$, $t = 4.55$, $p < 0.001$). No interaction terms were significant in the prediction indicating similar implementation of high pitch for IDS regardless of language and age of infant. The size of the Register effect was large across the longitudinal sample ($d_{Time1} = 0.88$; $d_{Time2} = 1.06$; $d_{Time3} = 0.86$; $d_{Time4} = 0.84$; $d_{Time5} = 0.88$; $d_{Time6} = 1.3$). Figure 5 presents speakers' averaged mean pitch and standard errors across the six time epochs.

2. Pitch range

Across all speakers and utterances, pitch range (Hz) showed considerable variability ($M = 199.53$, $SD = 109.87$, $CV = 0.55$). Pitch range was greater and showed more variability in IDS ($M = 208.90$, $SD = 116.95$, $CV = 0.56$) than in ADS ($M = 166.73$, $SD = 71.34$, $CV = 0.43$). Figure 6 presents individual speakers' pitch range for IDS and ADS at six time epochs.

As with mean pitch, pitch range was modeled both with quadratic and linear Time terms in Eq. (1). The linear model produced a better fit to the data ($AIC_{quad} = 19\,114.13$, $AIC_{lin} = 19\,108.95$, $LRT = 21.17$, $p < 0.01$). The linear model ($R^2 = 0.30$) showed only a main effect of Register ($\beta = 40.58$, $SE = 19.27$, $p < 0.05$). As with the mean pitch data, the size of the Register effect was large across the longitudinal sample ($d_{Time1} = 0.82$; $d_{Time2} = 0.60$; $d_{Time3} = 0.93$; $d_{Time4} = 0.61$; $d_{Time5} = 0.89$; $d_{Time6} = 0.11$). No other main effects or interactions achieved significance. Figure 7 presents speakers' mean pitch ranges and standard errors across the six time epochs.

TABLE III. Mean pitch (Hz) for IDS and ADS by language.

Language	Register	Mean pitch (SD, CV)
Tagalog	ADS	239.31 (31.38, 0.13)
	IDS	269.88 (50.53, 0.19)
SL Tamil	ADS	248.30 (40.02, 0.16)
	IDS	286.26 (62.45, 0.22)
Korean	ADS	221.88 (26.61, 0.12)
	IDS	246.14 (52.33, 0.21)

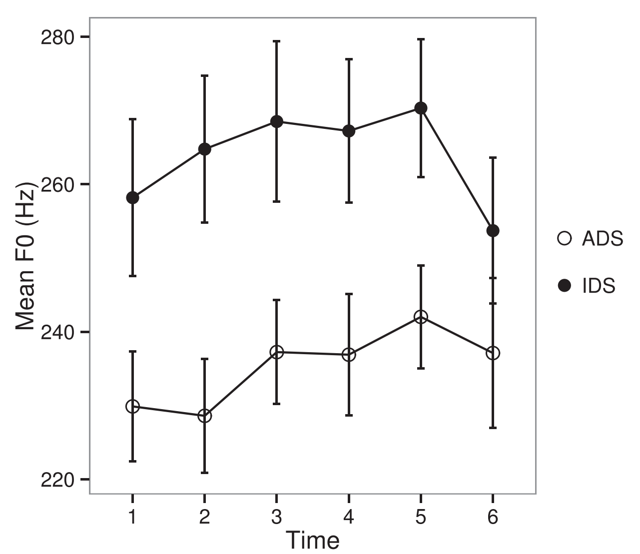

FIG. 5. Averaged mean pitch (Hz) in IDS and ADS across the first year of development (infant ages 4-15 months). Error bars represent standard error.

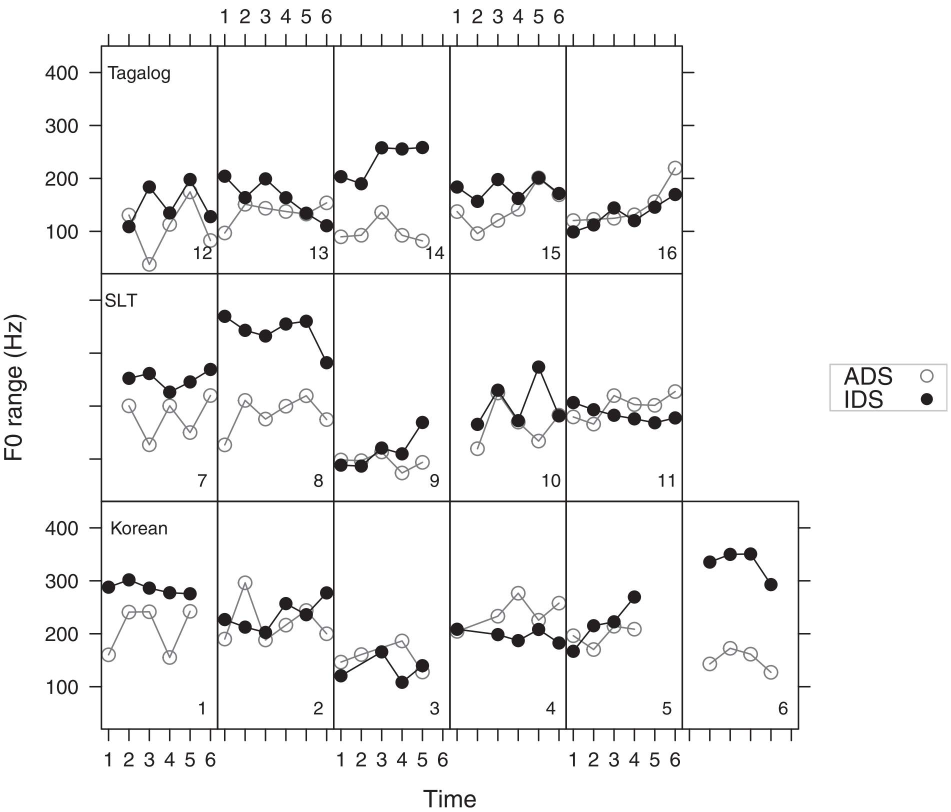

FIG. 6. Individuals' pitch range (Hz) in both IDS and ADS across infant development. Time epochs encompass infant ages from 4 to 15 months. Rows represent languages. Individual speaker numbers are given in lower right corner of each plot. Note that not all speakers made recordings at every time epoch. Table IV gives mean pitch measurements separated by language and register.

IV. DISCUSSION

To summarize our results, our models of speech rate, mean pitch, and pitch range in the CCIDS suggest that prosodic modifications in IDS are consistent with previous research findings, where caregivers speak to young infants slowly, with raised pitch, and with wide pitch excursions. Only speech rate showed a significant longitudinal trend in our analysis, suggesting speakers slowly increase their speech rate as infants develop over the course of 12 months between the ages of 4 months and 16 months (though see below). The model also showed that the difference between IDS and ADS speech rate is variable according to language, with SLT mothers showing less of a difference than either Tagalog- or Korean-speaking mothers. This language difference must be tempered, however, given the high variability in individuals' implementation of the speech rate feature (discussed in Sec. IV C below). Pitch characteristics did not significantly change over the course of infants' development in the CCIDS. Across speakers of all three languages in the CCIDS, mean pitch when speaking to infants was on average higher than ADS, and pitch range was likewise greater in IDS across languages and infant development. Below we discuss the nature of prosodic modifications in the CCIDS with an eye toward arguments of enhancement. We also discuss the individual variation in prosodic modifications seen in the CCIDS and its implications for our understanding of general characteristics of the IDS register.

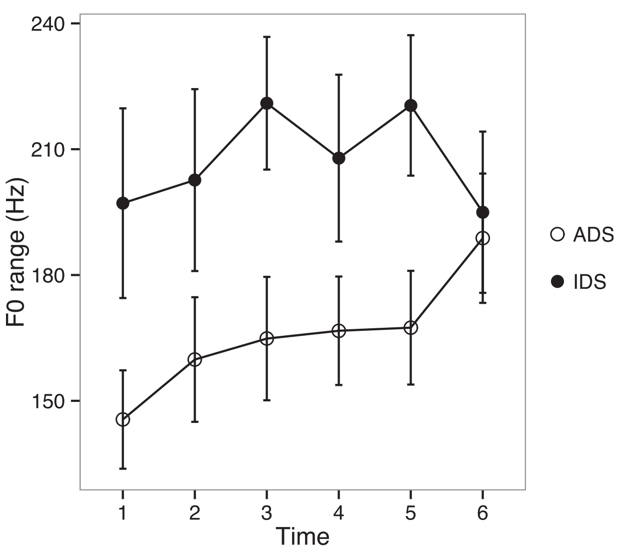

FIG. 7. Averaged pitch ranges (Hz) in IDS and ADS across the first year of development (infant ages 4-15 months). Error bars represent standard error.

A. Speech rate in IDS

Our longitudinal analysis of speech rate is novel in that it presents evidence that mothers' infant-directed (ID) speech rate approximates adult-directed (AD) rates before infants' second year. While the mixed-effects model showed a small, but significant longitudinal effect on speech rate, the

TABLE IV. Mean pitch ranges (Hz) for IDS and ADS by language.

Language	Register	Pitch range (SD, CV)
Tagalog	ADS	199.43 (67.06, 0.34)
	IDS	242.89 (123.44, 0.51)
SL Tamil	ADS	165.74 (76.87, 0.46)
	IDS	219.99 (121.62, 0.55)
Korean	ADS	131.54 (48.67, 0.37)
	IDS	170.89 (94.51, 0.55)

results must be tempered by the unbalanced nature of the data. Six of the 16 speakers did not have recordings at time 6, the endpoint where the model showed a dramatic decrease in between the two registers. Although mixed-effects models are generally considered robust with longitudinal data with missing values (cf. Krueger and Tian, 2004; Verbeke and Molenberghs, 2009, p. 215), we nonetheless entertain the possibility that the significant effect of time in the model is an artifact of the attrition of speakers at the last epoch. Without considering the unbalanced attrition at time 6, the decrease in the effect size of the register difference in speech rate (Sec. III A) is consistent with theoretical explanations of ID speech as subject to developmental constraints—with speech to infants becoming more adult-like as infant themselves become more linguistic. Does the decreasing difference between ID and AD speech rate suggest that speech to younger infants is acoustically enhanced or clearer than speech to older infants? The literature on perceptual advantages afforded to listeners (both infant and adult) in slow speaking rate conditions is unclear, showing differing effects on overall intelligibility and segmental contrast acoustics. In general, intelligibility and word recognition is improved in slow speech rate conditions. Research in the clear speech literature has shown increased intelligibility of slowed speech (one characteristic of clear speech on average) to adult listeners (Liu *et al.*, 2003). There is also some evidence that slowed speech rate improves word recognition in children. Zangl *et al.* (2005) found that infants 12- to 31-months old recognized unaltered words better than time-compressed words that were twice as fast. Song *et al.* (2010) found that slowed speech rate significantly enhanced 19-month-olds' ability to recognize words. They found infants correctly looking at target words 60% of the time when presented with typical IDS speech rate relative to 40% of the time in a fast IDS condition.

While intelligibility and word recognition may be enhanced in slow speech rate conditions, concomitant effects on phonetic segments may not necessarily be enhancing. While we know that infants normalize speech rate in a way that preserves segmental discriminability (Eimas and Miller, 1980; Miller and Eimas, 1983), the literature has yet to provide clear evidence that speaking slowly to infants necessarily results in their enhanced, less effortful, or more efficient speech perception of segments. While we might predict that by simply presenting *more* spectral information (as in the case of stressed vowels of slow speech; Gay, 1981), infants will have more evidence for spectrally determined categories, research suggests that the slowed speech of IDS results in greater spectral variability resulting in more overlap between phonetic categories than in ADS. McMurray *et al.* (2013) found that English IDS to 9- to 13-month olds increased variability in tongue height ($F1$), resulting in more overlap between front vowels than in ADS. Consonant contrasts that rely on timing relationships are likewise affected by the slowed speech rate of IDS. Consistent with earlier reports on voice-onset time (Englund, 2005), McMurray *et al.* (2013) showed that voice-onset times (VOT) for both voiced and voiceless stops are lengthened (not necessarily enhancing the phonetic contrast), and that the effect is due to the slowed speech rate of IDS. By the time infants are 15–16 months, however, English IDS shows less overlap between voicing categories along VOT than ADS (Malsheen, 1980). Given that the developmental trajectory of speech rate in the CCIDS occurs at a time when infants' perceptual categories are taking shape (e.g., Werker and Tees, 1984), we consider the small but steady increases in IDS speech rate over time as potentially having concomitant beneficial effects from the standpoint of phonetic category formation.

B. Pitch

The pitch modifications made by mothers in the CCIDS are consistent with previous reports of high mean pitch and expanded pitch range in IDS relative to ADS. The overall difference between the mean pitch of IDS and ADS in the CCIDS was approximately 30 Hz (or two semitones). While it has been suggested that pitch modifications in IDS begin to resemble ADS patterns towards the end of the infant's first year (Fernald, 1992, p.65), the present data do not show such a trend on average for speakers in the CCIDS. Kitamura *et al.* (2002) showed an effect of time on pitch range and mean pitch in Thai and Australian English IDS, though it is unclear whether the effect is evident within infancy or results from their including the ADS sample as the end point of their longitudinal continuum. That is, the current study assessed IDS characteristics relative to ADS at distinct time points in development and did not consider an adult-directed sample as the last moment of IDS in the longitudinal model.

Similar to the effects of speech rate on language acquisition processes, pitch modifications have variable effects on segmental perception and word recognition. The perceptual effect of high pitch may be an impediment to segmental discrimination. Trainor and Desjardins (2002) showed that 6- to 7-month-old infants' discrimination of the English tense/lax high-front vowel distinction was diminished in a high pitch (340 Hz) relative to a lower pitch (240 Hz) condition. Although the difference between the two conditions (six semitones) in the Trainor and Desjardins (2002) study is greater than the natural differences represented in the CCIDS languages, it is likely the case that elevated pitch does not, at the least, facilitate discrimination at the segmental level. The exaggerated pitch contours of IDS (here proxied as pitch range), however, may indeed be perceptually advantageous for the infant. In the same study by Trainor and Desjardin (2002), the front vowels were more discriminable to infants when presented with a falling contour of 200 Hz/s. Word recognition may not be affected by increased pitch range, however. Song *et al.* (2010) found no difference in 19-month-old infants' recognition of target words in typical-IDS and monotone-IDS conditions.

C. Individual variation in prosodic modifications in IDS

Perhaps the most striking aspect of the CCIDS data is the lack of either speech rate or pitch modifications in some speakers of IDS across the three languages (e.g., speakers 9, 12, and 16). Previous research describing the prosodic modifications that are characteristic of IDS have often failed to present individual data and especially data with time-yoked comparisons to an individual's ADS. The speakers in the present study show considerable variation both within and between languages. For example, speakers 7, 9, and 12 showed practically no periods of reduced speech rate in IDS, while speakers 4, 9, 12, and 15 showed little to no mean pitch increase. Speakers 2, 3, 4, 5, 9, 11, and 16 showed little to no pitch range expansion in their IDS. Curiously, while some speakers implement one type of prosodic modification over another in their IDS, there are some speakers who showed little or no prosodic modifications of any kind in their IDS (e.g., speakers 9 and 12), and these tendencies seem to be consistent across their infants' development. The individual variation in the implementation of these features is comparable to other linguistic registers with stereotypical acoustic characteristics. For example, Ferguson and Kewley-Port (2007) showed that the general slowed speech rate feature in the English clear speech register is highly variable, with some speakers showing little if any effect and others exhibiting dramatic slowing.

While the issue of universality of prosodic modifications in IDS has been addressed in previous research (e.g., Bernstein Ratner and Pye, 1984), prosodic exaggerations in the speech to infants nonetheless remain an oft-mentioned characteristic of IDS. The CCIDS data suggest that individual variation within languages encompasses a variety of prosodic modifications in naturalistic IDS and that prosodic modifications do not necessarily characterize the register. We believe that the individual variation in CCIDS prosody is representative of IDS prosody variation in general and related to the varieties of social-emotional affect (most often communicated with modifications in pitch and tempo, see Scherer, 1986) that caregivers display in interacting with infants.

V. CONCLUSION

Our longitudinal examination of various prosodic characteristics of IDS is revealing in light of the general literature on the acoustics of IDS. While certain features, such as speech rate, generally conformed to a developmental pattern found with other phonetic features (like voice onset time) (and the commonsensical notion that one speaks differently to a 4-month-old than to a 16-month-old), becoming more like ADS over the course of 12 months of development, pitch features in our study did not exhibit a similar longitudinal change. The study also indicates that speech rate modifications in IDS may have language-specific instantiations, with SLT-speaking mothers in the corpus exhibiting less of a difference between their speaking rates to infants versus adults.

The goal of our study was to describe the developmental complexity of IDS prosodic features with special attention to the variability exhibited by individual speakers. While gross patterns emerged from our models, individual behavior also suggests that prosodic modifications to IDS are not a general characteristic, but rather indicative of individualized predispositions caregivers might have in the presence of infants.

ACKNOWLEDGMENTS

This research was supported by a grant from the Social Science and Humanities Research Council of Canada (SRG#4487010) and a research grant from York University. We thank Evguenia Ignatova, Hugh McCague, and Georges Monette for their invaluable assistance with this project.

[1]The literature distinguishes between "speech rate" and "articulation rate," the former being a count of some prosodic unit (usually syllables) over a general period of time. Thus, speech rate usually includes moments of silence or pauses between utterances, while articulation rate refers to the computation of number of syllables over the course of phonation (cf. Goldman-Eisler, 1961). In this paper, we use "speech rate" to refer to articulation rate, that is, speech rate is computed only over stretches of connected speech and does not include pauses between utterances.

[2]Speakers 1-6 spoke Korean, 7-11 spoke SLT, and 12-16 spoke Tagalog. Speaker 3 had corrupt IDS files at 6 months, corrupt ADS files at 8 months, and dropped out of the recording procedure after her infant turned 14 months. Speaker 4 missed recording sessions at the 6-month period. Speaker 5 dropped out of the study when her infant reached 12 months. Speakers 6, 7, 10, and 12 began their recording sessions at 6 months. Speakers 1, 6, 9, and 14 dropped out of the study when their infants reached 14 months.

[3]Maximum number of candidates = 15; silence threshold = 0.03; voicing threshold = 0.45; octave cost = 0.01; octave-jump cost = 0.35; voiced/unvoiced cost = 0.14.

[4]Effect size with mixed-effects models is an unresolved issue in the literature, as there is not a clear-cut method for including and decomposing variance from random effects in the model. However, following a suggestion from Bates (2010) on the Mixed Effects Models mailing list, we take a coefficient of determination (correlation between fitted and observed values) as an overall effect size for the model. The mixed-effects model of speech rate has an R^2 of 0.37.

[5]We take the standardized regression coefficient as an effect size measure for the interaction.

Bates, D. (**2010**). "Mixed effects models mailing list," http://thread.gmane.org/gmane.comp.lang.r.lme4.devel/3281 (Last viewed March 14, 2015).

Bernstein Ratner, N. (**1984**). "Patterns of vowel modification in mother-child speech," J. Child Lang. **11**, 557–578.

Bernstein Ratner, N. (**1985**). "Dissociations between vowel durations and formant frequency characteristics," J. Speech Lang. Hear. Res. **28**, 255–264.

Bernstein Ratner, N., and Pye, C. (**1984**). "Higher pitch in BT is not universal: Acoustic evidence from Quiche Mayan," J. Child Lang. **11**, 515–522.

Boersma, P., and Weenink, D. (**2009**). "Praat: Doing phonetics by computer (version 5.3.75) [computer program]," http://www.fon.hum.uva.nl/praat/ (Last viewed April 30, 2014).

Christdas, P. (**1988**). "The phonology and morphology of Tamil," Ph.D. thesis, Cornell University, Ithaca, NY.

Cooper, R. P., and Aslin, R. N. (**1990**). "Preference for infant-directed speech in the first month after birth," Child Dev. **61**, 1584–1595.

Cruttenden, A. (**1994**). "Phonetic and prosodic aspects of baby talk," in *Input and Interaction in Language Acquisition*, edited by C. Gallaway and B. J. Richards (Cambridge University Press, Cambridge, UK), pp. 135–152.

de Jong, N. H., and Wempe, T. (**2009**). "Praat script to detect syllable nuclei and measure speech rate automatically," Behav. Res. Methods **41**, 385–390.

Eefting, W., and Rietveld, A. C. M. (**1989**). "Just noticeable differences of articulation rate at sentence level," Speech Commun. **8**, 355–361.

Eimas, P. D., and Miller, J. L. (**1980**). "Contextual effects in infant speech perception," Science **209**, 1140–1141.

Englund, K. T. (**2005**). "Voice onset time in infant directed speech over the first six months," First Lang. **25**, 219–234.

Ferguson, S. H., and Kewley-Port, D. (**2007**). "Talker differences in clear and conversational speech: Acoustic characteristics of vowels," J. Speech Lang. Hear. Res. **50**, 1241–1255.

Fernald, A. (**1992**). "Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective," in *The Adapted Mind:*

Evolutionary Psychology and the Generation of Culture, edited by J. H. Barkow, L. Cosmides, and J. Tooby (Oxford University Press, New York), pp. 391–428.

Fernald, A., and Kuhl, P. K. (**1987**). "Acoustic determinants of infant preference for motherese speech," Infant Behav. Dev. **10**, 279–293.

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.

Garnica, O. (**1977**). "On some prosodic and paralinguistic features of speech to young children," in *Talking to Children: Language Input and Acquisition*, edited by C. Snow and C. Ferguson (Cambridge University Press, Cambridge, UK), pp. 271–285.

Gay, T. (**1981**). "Mechanisms in the control of speech rate," Phonetica **38**, 148–158.

Goldman-Eisler, F. (**1961**). "The significance of changes in the rate of articulation," Lang. Speech. **4**, 171–174.

Grieser, D. L., and Kuhl, P. K. (**1988**). "Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese," Dev. Psych. **24**, 14–20.

Kang, Y., and Han, S. (**2013**). "Tonogenesis in early Contemporary Seoul Korean: A longitudinal case study," Lingua **134**, 62–74.

Karzon, R. G. (**1985**). "Discrimination of polysyllabic sequences by one- to four-month-old infants," J. Exp. Child Psychol. **39**, 326–342.

Keane, E. (**2006**). "Rhythmic characteristics of colloquial and formal Tamil," Lang. Speech **49**, 299–332.

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., Thanavishuth, C., Burnham, D., and Luksaneeyanawin, S. (**2002**). "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.

Krueger, C., and Tian, L. (**2004**). "A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points," Biol. Res. Nurs. **6**, 151–157.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevinikova, 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.

Lee, S., Davis, B. L., and Macneilage, P. F. (**2008**). "Segmental properties of input to infants: A study of Korean," J. Child Lang. **35**, 591–617.

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. (**2007**). "Acoustic analysis of lexical tone in Mandarin infant-directed speech," Dev. Psychol. **43**, 912–917.

Malsheen, B. J. (**1980**). "Two hypotheses for phonetic clarification in the speech of mothers to children," Child Phon. **2**, 173–184.

Martin, A., Schatz, T., Versteegh, M., Miyazawa, K., Mazuka, R., Dupoux, E., and Cristia, A. (**2015**). "Mothers speak less clearly to infants than to adults: A comprehensive test of the hyperarticulation hypothesis," Psychol. Sci. **26**, 341–347.

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., and McEchron, W. (**2013**). "Infant directed speech and the development of speech perception: Enhancing development or an unintended consequence?," Cognition **129**, 362–378.

Miller, J. L., and Eimas, P. D. (**1983**). "Studies on the categorization of speech by infants," Cognition **13**, 135–165.

Narayan, C. R., and Yoon, T. J. (**2011**). "VOT and f0 in Korean infant-directed speech," Can. Acoust. **39**, 152–153.

Quené, H. (**2007**). "On the just noticeable difference for tempo in speech," J. Phon. **35**, 353–362.

Scherer, K. R. (**1986**). "Vocal affect expression: A review and a model for future research," Psychol. Bull. **99**, 143–165.

Shockey, L., and Bond, Z. S. (**1980**). "Phonological processes in speech addressed to children," Phonetica **37**, 267–274.

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

Song, J. Y., Demuth, K., and Morgan, J. (**2010**). "Effects of the acoustic properties of infant-directed speech on infant word recognition," J. Acoust. Soc. Am. **128**, 389–400.

Stern, D., Spieker, S., Barnett, R., and MacKain, K. (**1983**). "The prosody of maternal speech: Infant age and context related changes," J. Child Lang. **10**, 1–15.

Tang, J. S., and Maidment, J. A. (**1996**). "Prosodic aspects of child-directed speech in Cantonese," Speech Hear. Lang. **9**, 257–276.

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.

Verbeke, G., and Molenberghs, G. (**2009**). *Linear Mixed Models for Longitudinal Data* (Springer, New York).

Werker, J. F., and McLeod, P. J. (**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. C. (**1984**). "Cross-language speech perception: Evidence for perceptual reorganization during the first year of life," Infant Behav. Dev. **7**, 49–63.

Woodward, J. Z., and Aslin, R. N. (**1990**). "Segmentation cues in maternal speech to infants," in *7th Biennial Meeting of the International Conference on Infant Studies*, April, Montreal, Quebec, Canada.

Yoon, Y. B. (**1995**). "Experimental studies of the syllable and the segment in Korean," Ph.D. dissertation, University of Alberta, Edmonton, Alberta, Canada.

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.

Zuraw, K. R. (**2000**). "Patterned exceptions in Tagalog," Ph.D. thesis, University of California, Los Angeles, CA.