RUNNING HEAD: Acoustic-Phonetic Differences in IDS and ADS

Acoustic-Phonetic Differences between Infant- and Adult-directed Speech: The Role of Stress and Utterance Position

Yuanyuan Wang
Linguistics
Purdue University,

Amanda Seidl
Speech Language & Hearing Sciences,
Purdue University,

Alejandrina Cristia
Laboratoire de Sciences Cognitives et Psycholinguistique,
CNRS,
IEC-ENS,
EHESS,

Corresponding author:
Amanda Seidl
Speech Language & Hearing Sciences,
Purdue University,
500 Oval Dr.,
West Lafayette,
IN 47907
aseidl@purdue.edu

Abstract

Previous studies have shown that infant-directed speech (IDS) differs from adult-directed speech (ADS) on a variety of dimensions. The aim of the current study was to investigate whether acoustic differences between IDS and ADS in English are modulated by prosodic structure. We compared vowels across the two registers (IDS, ADS) in both stressed and unstressed syllables, and in both utterance-medial and -final positions. Vowels in target bisyllabic trochees in the speech of twenty mothers of four- and eleven-month-olds were analyzed. While stressed and unstressed vowels differed between IDS and ADS for a measure of F0, and trended in similar directions for vowel peripherality, neither set differed in duration. These profiles held for both utterance-medial and -final words.

Keywords: IDS, ADS, acoustic cues, stress, utterance-position

Acoustic-Phonetic Differences between Infant- and Adult-directed Speech: The Role of Stress and Utterance Position

Introduction

When interacting with infants, caregivers use a unique speech register known as infant-directed speech (IDS; e.g., Snow, 1977). This register is distinct from adult-directed speech (ADS). Over the past 40 years, a rather extensive body of research has investigated the acoustic-phonetic differences between IDS and ADS instrumentally. In a recent systematic review of this literature, 85% of studies reported slower tempo and/or longer vowel duration; 92% found higher fundamental frequency (F0); and 82% documented larger vowel peripherality (Cristia, 2013). Recent research has begun to explore whether IDS and ADS always differ across the board, or rather variation across the registers can be modulated by other factors. In particular, one recurrent hypothesis is that all IDS-ADS differences could potentially be reduced to a few factors, circumventing the need to postulate that speakers specifically aim to produce each and every acoustic-phonetic characteristic mentioned above. For example, caregivers produce shorter utterances with slower rate when speaking to small infants and tend to place target words at the ends of utterances (Albin & Echols, 1996; Aslin, Woodward, LaMendola, & Bever, 1996). *Ceteris paribus*, the presence of shorter utterances produced at a slower-rate entails more words will be aligned with boundaries, which are thus subject to boundary strengthening effects; a higher proportion of words in focus; and monosyllabic words of longer duration (McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013). Notice that boundary strengthening effects result in longer vowel duration and greater

hyperarticulation (Turk & White, 1999, Wightman, Shattuck-Hufnagel, Ostendorf, & Prince, 1992). Moreover, expanded pitch range is expected in utterances having a greater proportion of focused items and stressed syllables. Thus, the acoustic-phonetic consequences of a single goal (shorter utterances at a slow speech rate) could be far-reaching: they come to have precisely the characteristics observed in IDS (longer vowel duration, higher F0, greater vowel peripherality). In this context, work assessing the modulation of IDS as a function of prosodic boundaries and lexical factors provides a much needed insight into infants' input to language. Thus, the primary purpose of this paper is to explore whether IDS and ADS differences are modulated by prosodic structure, specifically prosodic boundaries and lexical stress. Before turning to our contribution to this question, we briefly summarize previous work in this research line.

Some extant findings are consistent with the reductionist predictions made above. Fernald and Mazzie (1991) studied the use of focus by comparing the speech of mothers telling a story to a fourteen-month-old and an adult. They found that mothers consistently placed focused words with pitch peaks in utterance-final position, whereas prosodic emphasis in ADS was more variable. Kondaurova and Bergeson (2011) compared pitch, pause, and duration cues to large prosodic boundaries in speech to six-month-olds and adults, and found that most of these cues were exaggerated in vowels located in pre-boundary position (utterance-final) as compared to those in post-boundary (utterance-initial) position, suggesting that *final* strengthening was boosted in IDS. More specifically, vowel duration was longer in IDS than ADS for pre-boundary, but not for post-boundary words. Moreover, other studies also provide evidence that vowels do not differ in duration across the registers in utterance *initial* or *medial* position. For example,

Church (2002) found that when the lengthened final syllable was excluded from calculations of duration, IDS was *not* significantly slower than ADS, suggesting that the slower rate of IDS was primarily due to lengthened final syllables.

In contrast, other work suggests that the picture is more complex. For example, Swanson (1990) investigated fifteen English-speaking mothers reading five stories to both their toddlers ranging from eighteen to twenty-eight months of age, and to an adult. Vowel duration was longer in IDS in both final and non-final positions. Similarly, Albin and Echols (1996) found that both stressed and unstressed syllables in IDS (to six- and nine-month-olds) were longer in duration and higher in amplitude as compared with ADS acoustics, in both utterance-medial and utterance-final position. These outcomes do not support the reductionist view that IDS-ADS differences arise from the higher incidence in IDS of boundaries and words in focus.

Contradictory results could be due in part to different methodologies employed. For example, Swanson (1990) used prepared texts for the mothers, but did not control for the use of focus, in a way similar to Fernald and Mazzie (1991); Albin and Echols (1996) explicitly asked caregivers to label specific objects provided, whereas the speech in Kondaurova and Bergeson (2011) and Church's (2002) sample was not oriented to a specific experimental goal. Moreover, inconsistent results could also be partially due to changes in IDS as a function of the child's age and development. Notice that the ages studied varied from six months (in Kondaurova & Bergeson, 2011, and half of the infants in Albin & Echols, 1996) to twenty-eight months (in Swanson, 1990). It has sometimes been reported that IDS changes across the first two years (e.g., Bernstein, 1986; Kitamura & Burnham, 2003; Kitamura, Thanavishuth, Burnham, Luksaneeyanawin, 2002;

Kondaurova & Bergeson, 2011; Liu, Tsao, & Kuhl, 2009; Shute & Wheldall, 1989; Stern, Speiker, Barnett, & Mackain, 1983; Thanavisuth & Luksaneeyanawin, 1998). For example, Stern et al. (1983) found that IDS addressed to four-month-old infants is higher-pitched than the IDS addressed to neonates, and to one- and two-year-olds. In contrast, others report no consistent changes in pitch height between six and twelve months (Kondaurova, Bergeson, & Xu, 2013). Nonetheless, it is clear that the question of whether IDS acoustic cues are exaggerated because of the higher incidence of prosodic boundaries remains open.

Although less commonly invoked, an enhancement of lexical stress alone could also explain much, if not most, of the acoustic phonetic characteristics commonly attributed to IDS, at least in English. Like many other languages (e.g., Dutch, Spanish, Russian), English has variable lexical stress, such that the same sequence of sounds can have different meanings dependent only on the location of stress (e.g., reCORD-REcord). Nonetheless, most English words begin with a stressed syllable, and most bisyllables are trochees (stressed-unstressed). Previous work suggests that American English-learning infants know the difference between stressed and unstressed syllables (Jusczyk & Thompson, 1978); they encode stressed syllables better (Houston, Santelman, & Jusczyk, 2004); and they use these stress differences to segment the speech stream (Curtin, Mintz, & Christiansen, 2005). In English ADS, stressed syllables differ from unstressed syllables in one or more of the following ways: higher amplitude, longer duration, higher F0, and fuller vowel quality (Fry, 1958; Lehiste, 1970). Notice that the last three characteristics rely on the same acoustic dimensions along which IDS and ADS differ, and thus it becomes relevant to investigate whether lexical stress interacts with register. Specifically,

the first question is whether the registers differ once any enhancement in lexical stressed syllables has been accounted for. This is relevant because Albin and Echols (1996) indicated that stressed and final unstressed syllables differed to a greater extent in IDS than ADS, regardless of their position in the utterance (see also Jacobson, Ward, & Sundara, 2011, for evidence on the acoustic cues to stress in Spanish IDS). Thus, even once the confound with higher incidence of boundaries has been addressed, it remains possible that much of the differences between the registers could be due to the enhancement of stress specifically.

Thus, our goal was to contribute to the understanding of whether IDS and ADS differences are modulated by prosodic structure, especially lexical stress and prosodic boundaries, and how this modulation (if any) changes across development. This is an important question since it may serve to inform us about why IDS, in general, displays exaggerated acoustic characteristics. Researchers have come to believe IDS plays important role in the course of infants' communication and language acquisition, such as to direct and engage infants' attention (Papousek, Papousek, & Symmes, 1991; Stern, MacKain, & Spieker, 1982), to communicate affect and facilitate social interaction (Fernald, 1985), to isolate words from continuous speech (Culter & Norris, 1988; Thiessen, 2005), and to cue the grammar of utterances (Fernald & Mazzie, 1991; Kaplan, Bachorowski, Smoski, & Hudenko, 2002). For example, infants of depressed mothers who use little IDS-like exaggeration learn less from their mother's speech than those of mothers' whose IDS shows exaggerated properties usually associated with IDS (Kaplan et al., 2002).

Much of the speculation on how IDS facilitates infants' language acquisition is tied to the possibility that the exaggerated prosodic and segmental information in IDS increases the saliency of acoustic cues for word recognition (Fernald & Mazzie, 1991; Fernald & Simons, 1984; Gleitman et al., 1988; Peters, 1983; Thiessen, 2005). Thus, if IDS exaggerates acoustic characteristics only in linguistically salient positions, such as in stressed syllables or at prosodic boundaries, particularly if infants and children attend mainly to these salient units, learners may then be processing a very different input than would be heard in typical ADS, which may potentially facilitate their language acquisition. However, if IDS-ADS differences are not modulated by lexical stress or prosodic boundaries, then the unique IDS characteristics that we see may instead be the byproduct of caregiver's expression of intense emotion or affect. Thus, work along these lines may facilitate our understanding of the potential role of IDS during the course of language acquisition.

Given these goals, in an object categorization task, we collected IDS and ADS from two groups of participants at two crucial, yet comparable, points in development. One group was comprised of infants who were about four months of age (still not experts in phonology and not word learners, but showing strong social engagement with caregivers), and another of infants who were about eleven months of age (more proficient word learners and more socially engaged). First, we asked whether IDS exhibits longer duration, higher F0 and expanded vowel peripherality than ADS in stressed syllables, on the one hand, and unstressed syllables, on the other. Larger vowel peripherality would reflect better contrast in the articulatory space, and thus better intelligibility (e.g., Jakobson, 1941, Bradlow, Torretta, & Pisoni, 1996). Second, we asked whether this

difference (if present) is more salient in utterance-final position as opposed to utterance-medial position. Third, we asked whether this highlighting in IDS (if present) is dependent upon the age of the infant.

We predict that if one of the goals of IDS is to present preverbal infants with the prosodic patterns of their ambient language, then caregivers may highlight acoustic cues in prosodically salient positions, i.e., stressed syllables and utterance-final positions, when speaking to their infants. Predictions regarding the role of age were more tentative. Our intuition was that in speech to younger infants prosodic boundaries would explain more of the IDS-ADS differences, since at this age infants are cracking their prosodic structure (Seidl, 2007); whereas stress would be a major factor for the eleven-month-olds, who have mastered the use of metrical and other cues for word segmentation (Jusczyk, Houston, & Newsome, 1999). However, if IDS does not specifically aiming at presenting linguistically relevant units to each age group, then it is less likely that IDS and ADS differences will be distinct between the two age groups.

Methods

Participants

The twenty participants in this study were ten mothers of eleven-month-old infants (*M* = 0;11.40, range: 0;11.12 – 0;12.01) and ten mothers of four-month-old infants (*M* = 0;4.35, range: 0;3.95 – 0;4.99), whose information was obtained through birth announcements in the local newspaper. The mothers selected were primary caregivers, who were native speakers of American English from a small Midwestern town and had no history of hearing, speech, or language disorders. Infants were healthy full terms with typical

development and no known history of hearing impairment. All dyads participated in the study voluntarily and were given informed consent; each infant was compensated with a book or a toy for their participation.

Data Collection

Participants were recorded in two sessions that took place on the same day. In the first, the IDS session, each participant was asked to interact with her four- or eleven-month-old infant. In the second, the ADS session, they interacted with an experimenter and an undergraduate confederate. In this session each mother was asked whether/how she had described the objects to her infant in the IDS session. Both sessions took place in a sound attenuated room. The mothers were fitted with a Lavalier microphone (AKG WMS40), whose signal was recorded onto a Marantz Professional Solid State Recorder (PMD660ENG). For each session mothers were given a container filled with bags containing object/picture sets and were then left alone with their child for about 20-30 minutes; afterwards, they interacted with the experimenter and confederate for 20-30 minutes.

Participants were told that we were interested in how caregivers talk to their infants about categories. Thirty-five object sets were used to assist caregivers in producing target labels in IDS and ADS, such as *pegboard, beetle, bacon, basil*, etc. Before recording, mothers were familiarized with the equipment, and the procedures for the study were explained. For each target-word bag, there were 2 similar objects or pictures (e.g., 2 wooden pegboards of different sizes) and an oddball object/picture (a wooden rod), which were in a bag. The objects used to elicit target labels were monosyllabic, (e.g., *shoes, sheep*, etc.), bisyllabic (e.g., *teddy, bacon*, etc.), or greater than two syllables (e.g.,

bassinet, dictionary, etc.); however, only bisyllabic trochees were used in the analyses given that a trochee is the most common prosodic pattern found in speech addressed to preverbal English-learning infants. There were a total of 23 target trochaic words included in the final analyses. Examples of speech samples from an IDS and an ADS session are provided in Appendix A. It should be noted that vowel qualities in target words differ in stressed and unstressed syllables. For this reason analyses were confined to stressed or unstressed syllables and did not compare across these two syllable types.

Speech Sampling and Coding

The recordings of both IDS and ADS were stored in 20 separate files (one audio file for each participant) at 44.1 kHz. All coding and analyses were done using Praat (Boersma & Weenink, 2011). For each file, a waveform and a spectrogram of the utterances were displayed on the screen of the computer to facilitate coding.

Four tiers were created in a text file, in which target words, vowels, register and utterance-position were coded, as shown in Figure 1. In the target word tier, all target words were annotated except those overlapped with toy noise, infant vocalization, conversation with the experimenter, or produced with whispered or glottalized speech.

In the vowel tier, stressed and unstressed syllables from the target words were tagged. For example, for the target word Pepsi, [ɛ] is the stressed vowel, and thus was tagged as stressed; similarly, [i] was tagged as unstressed. Highly trained coders marked the vowel onset and offset for all of the syllables. The onset of the vowel was defined as the first upward crossing after the onset of periodicity following the burst or fricative release of the preceding consonant. The offset of the vowel was determined as an abrupt attenuation

of energy, evident in both the waveform and the spectrogram. Coders could also used the auditory signal to assist in coding.

In the register tier, the files were divided into two parts depending on whether they were from the IDS or ADS session. In the utterance-position tier, for each target word, subjective judgments of utterance positions were made by three research assistants with training in phonetics. They were instructed to annotate the utterance position of each target word as either utterance-medial or utterance-final. If there was a disagreement regarding the utterance position, the sample was discarded. (Words in utterance-initial position or in isolation were not included in the analyses since there were too few of these tokens to provide meaningful data.)

A Praat script was written to extract the following acoustic measures in all the tagged vowels (both stressed and unstressed): (a) mean duration in seconds; (b) mean F0 in ERB; and (c) F1 and F2 in Bark. All of these measurements were extracted in a completely automatized fashion (with no human intervention), using routines available in Praat. For example, formants are identified from the peaks drawn from a linear prediction analysis of the acoustic signal. Since formant measurements are greatly affected by a setting for the maximum frequency at which formants can be found, and given that the optimal setting varies with speaker and vowel identity, we followed Escudero, Boersma, Rauber, and Bion (2009) and estimated formants using a range of maximum frequency settings. We then selected the setting that minimized variance within speaker and vowel. More information on this process can be found in the supplementary materials available on https://www.dropbox.com/sh/wi2g5rsgwax51cg/RlzQ8nXcnL.

To estimate vowel peripherality, we first retrieved the average F1 and F2 of the vowels /ɑ, i, u/ (point vowels) as instantiated in the target words *shopping* cart, *sheep*, and *shoes* produced by each mother in the IDS and ADS sessions (had we restricted our estimations to /ɑ, i, u/ tokens produced only in the IDS session, a further 5 caregivers would have had to be excluded).

We then calculated an average across all three vowels, which we subsequently used as center of the vowel space for each mother. Finally, for each vowel in our stressed and unstressed syllables, a measure of vowel peripherality was calculated as the Euclidean distances from that center, $d= \sqrt{(\mu F1 - F1i)^2 + (\mu F2 - F2i)^2}$, where 'μ' indicates the average and 'i' the token under examination. The larger the d value, the more distant a given token is from that vowels center of mass for that speaker (e.g., Bradlow et al., 1996). Peripherality was calculated for each participant and within each register because it was expected that different participants would have inherently different vowel peripherality, which may vary across register. Three caregivers did not have tokens for at least one of the vowels and thus could not be included in the peripherality analyses.

Statistical analyses

Due to the fact that the data examined have different vowel qualities in stressed and unstressed syllables, and vowel quality can impact F0 and duration, analyses were conducted within each prosodic type, i.e., we ran statistical analyses on stressed and unstressed syllables separately.

We analyzed these data in several different ways. For these spontaneous, and therefore unbalanced datasets containing structured variance, a mixed model analysis performed on transformed data (such that residuals would be normal) is most appropriate

(Baayen, 2008). Nonetheless, mixed models are a relatively new statistical tool in psycholinguistic research and thus more traditional analyses facilitate integration with previous studies. We used three of these more traditional measures (*Cohen's d, Binominal distribution*, and *ANOVA*) on *untransformed* data. The results from the mixed model were checked against an ANOVA where the outcome was the median for the relevant cue within each talker for a single selected vowel, declaring register and sentence position as within-participant, and age as between participants. Results from the effect size and binomial analyses were similarly replicated in the same dataset (without controlling for sentence position). The specific vowel was /ɛ/ for the stressed syllables (589 tokens). Eight caregivers of four-month-olds and 9 caregivers of eleven-month-olds had data for both registers and sentence position and thus could be included in the ANOVA. For unstressed syllables, the vowel /i/ was selected on the same criterion (349 tokens). Seven caregivers of four-month-olds and 6 caregivers of eleven-month-olds could be included as they had complete data for the ANOVA. Our general discussion will only dwell on results that are stable across all *four* analyses. Interested readers are invited to visit the project website (Cristia, 2014), where the original dataset, scripts, and all intermediate analyses are available for download.

Mixed-effects models. Six mixed models were fit as the crossing of three acoustic correlates (duration in s, F0 in ERB, and vowel peripherality in Bark) in the two prosodic types (stressed and unstressed). Within each model, we declared two within-participant fixed factors: register (IDS vs. ADS) and position (utterance-medial vs. utterance-final), one between-participant fixed factor: age (speech to four- vs. eleven-month-olds), as well as two random factors: talker (maximally 20 levels) and word (maximally 23 levels for

stressed, 11 levels for unstressed). It should be noted here that the difference in the number of levels of vowels between stressed and unstressed syllables is due to the fact that we excluded all the unstressed syllables with a schwa (e.g., *tend**er***) or possible syllabic consonants (e.g., *basi**l**, baco**n***). Notice particularly that we include random intercepts for words (and therefore the vowels contained in them) to control for the fact that different target words/vowels could have intrinsic effects on acoustic characteristics (for example, vowels could vary intrinsically in F0). One set of analyses compared IDS and ADS along each of the 3 acoustic cues within stressed syllables; the second set of analyses did the same for unstressed syllables. These analyses were implemented using the lmer function, part of the lme4 package (Bates & Sarkar, 2007), in R. (The full model, fitted with the complete structure, thus translates to *lmer(mydata[,outcome] ~ age.code* sentPosition.code* register.code + (1|talker)+(1|word), data=mydata).)*

This type of statistical model is comparable to ANOVAs, in that it relies on Gaussian distributions. However, it is clear that duration, F0, and vowel peripherality measures do not follow a normal distribution (for example, they are bounded at zero). Therefore, we inspected the distributions of our outcome measures prior to model fitting using histograms and QQ plots, and the residuals of the full models. Predictably, it was often the case that the distributions of the dependent variables were right-skewed. When so, we z-scored (converting the values by subtracting the mean and dividing the difference by the standard deviation), and removed items more than 2.5SD off the mean. In most cases, this was sufficient to yield quasi-normal distributions; we point out below when this was not the case. As a result of these transformations and the complex model structures used, the estimated betas are difficult to interpret, and thus not reported here.

In this mixed model analysis, p-values were estimated using pvals.fnc from languageR (Baayen, 2008), which relies on Markov Chain Monte Carlo sampling (N=1,000). Since the data for duration, F0, and peripherality in stressed and unstressed vowels are drawn from the same speech samples, they constitute 6 repeated measures. Therefore, we used a Bonferroni correction (multiplied p-values by 6) to establish significance levels. We use the subscript in $p_{MCMC\text{-}B}$ to indicate these are p-values from MCMC, Bonferroni-corrected. Since there is controversy regarding how degrees of freedom for mixed models should be calculated (see, for example, the explanation by the co-creator of the lmer package in R regarding why this analysis should not give degrees of freedom: https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html), we provide the number of datapoints (N), talker and vowel clusters for each model.

Cohen's d. Our second reported measure is Cohen's d, which is calculated from the division of two terms. The numerator is the mean difference scores across the two registers, with each register being represented by the median values of untransformed measures within individual talkers. This mean difference score was divided by the standard deviation of the difference scores over talkers. Thus, effect size indicates, in a single number, how large the difference between registers is, and how consistent this difference is across participants. As a result, effect size is widely viewed as a "common currency" (e.g., Lipsey & Wilson, 2001), greatly facilitating integration with previous and future work. Cohen's d of about 0.3 are described as "small", about 0.5 "medium", and greater than 0.8 as "large".

Binomial distributions. Third, we report a non-parametric description of the individual data, namely what proportion of talkers have higher median value for a given measure in

IDS as compared to ADS. This test has the advantage of not assuming that the underlying distributions have any specific shape, and it further uniquely illustrates whether an effect is widespread across the population (regardless of whether the *size* of the difference between registers is constant across participants).

ANOVA on the most common vowel. Finally, we fit an Analysis of Variance where the outcome is the median for the relevant cue within each talker for a single selected vowel, declaring register and sentence position as within-participant, and age as between participants. For stressed syllables, the vowel /ɛ/ was selected because it was the most commonly spoken (589 tokens). Eight caregivers of 4-month-olds and 9 caregivers of 11-month-olds had data for both registers and sentence position and thus could be included in the analysis. For unstressed syllables, the vowel /i/ was selected on the same criteria (349 tokens). Seven caregivers of four-month-olds and 6 caregivers of eleven-month-olds could be included as they had complete data for the ANOVA. ANOVAs, as the mixed models applied above, assume normality of residuals, an assumption that was not met for most of our measures when unnormalized. Thus, results from this analysis should be interpreted with caution, and need not be viewed as more valid or meaningful than the mixed models. As with the mixed model, we multiplied p-values by 6 to apply the Bonferroni correction (indicated as p_B).

Results

There were a total of 1,365 trochaic words across all the 20 participants. After removing words whose pronunciation as trochees or iambs was ambiguous (e.g., *bamboo*, *baboon*), and removing vowels whose target was a schwa or could be absent (e.g., *tender*, *basil*),

1273 tokens remained for the analyses of stressed syllables, and 636 tokens remained for unstressed syllables.

Each mother contributed a different number of tokens for both IDS and ADS in different positions. The average number of target word tokens per participants included in the analyses is shown in Table 1. As evident in this table, register and position are not independent of each other in both age groups. Indeed, mothers of the four-month-olds were as likely to produce words in utterance-final as in utterance-medial positions in IDS, whereas they had many more utterance-medial than utterance-final tokens in ADS (for example in the dataset for stressed syllables, χ^2 (1, N = 676) = 21.09, $p < .001$, but this difference was less marked in mothers of eleven-month-olds, χ^2 (1, N = 597) = 7.84, p = .005). The total number of times each target word was used varied. Furthermore, the frequency of each target word in IDS and ADS sessions was different as shown in Table 2, (e.g., for the stressed dataset, χ^2 (22, N=1273) = 41.76, p = .007). The inclusion of word as a random factor in the mixed model analysis allows us to control for differences across the registers that are related merely to differences in frequency of occurrence of target words (and the vowels they contain).

We also investigated whether any difference could be due to caregivers producing more tokens of the same target word, and thus more repetition. By the second and subsequent repetitions, a word is more predictable (as it is the established topic), and it can consequently be hypo-articulated. The median number of repetitions per word type averaged across talkers was 1.68 (SD = 0.52) for ADS and 2.2 (SD = 1.02) for IDS, a difference that met the typical threshold for statistical significance in a two-tailed paired comparison $t(19)$ = 2.17, p = 0.04. Notice, however, that this predicts *less*

hyperarticulation in IDS than ADS, counter to what others find. Moreover, this effect cannot affect stressed and unstressed syllables differently, and is thus orthogonal to the research question at hand.

Means and standard deviations per register are reported in Table 3, for stressed vowels, and Table 4, for unstressed vowels. Main effects and/or interactions with register, specifically, according to each of our four statistical methods are summarized in Table 5. We report in detail only the mixed model because it is the only multifactorial method and it was run on transformed data, thus making it necessary to convey the transformations applied to the data.

Vowels in Stressed Syllables

Duration. The logarithm was applied following previous phonetic work (e.g., Escudero et al., 2009). The resulting distribution was normal and no further transformations or exclusions were necessary (N = 1,273; 23 word clusters, 20 talker clusters). This mixed model analysis revealed no main effects or interactions (the largest t = -2.27, $p_{MCMC\text{-}B}$ = .13, corresponded to sentence position). Further inspection of Table 5 reveals that results of the four analyses concur: none of them gives strong evidence that vowel duration in stressed syllables varied with register.

Fundamental frequency. Inspection of the distribution of this measurement revealed it was right-skewed. Therefore, we applied z-scoring and trimmed values more than 2.5SD (N = 26) to improve the residual distribution. The residuals from a mixed model fit on these data neared a normal distribution. Results (N = 1,247; 23 word clusters, 20 talker clusters) revealed only a main effect of register, t = 3.99, $p_{MCMC\text{-}B}$ = .006, due to the F0 of

stressed vowels being higher in IDS than that in ADS. Inspection of Table 5 confirms that the effect of register was large and stable across all analyses.

Peripherality. This variable was also right-skewed. Residuals from a mixed model on the z-scored and trimmed distribution (excluded N = 124) did not greatly deviate from a normal distribution. The mixed model on the transformed peripherality data (N = 1,149, 23 word clusters, 17 talker clusters) revealed no significant main effects or interactions (the largest $t = 1.42$, $p_{MCMC\text{-}B} = .972$, corresponded to the position by register interaction). Inspection of Table 5 suggests that there is a trend for greater peripherality in IDS, but it is not markedly large or stable across participants. Given that the effect size for four-month-olds classifies as large, and that most caregivers showed it, it becomes relevant to report in full the analyses focused on the vowel /ɛ/. In the ANOVA, the F for register was $F(1,12) = 0.1$; the effect size for four-month-olds was 0.43 and for eleven-month-olds was 0.12; 5 out of 8 caregivers of four-month-olds and 4 out of 9 caregivers of eleven-month-olds had greater peripherality in IDS than ADS stressed vowels.

Vowels in Unstressed Syllables

Duration. As with stressed syllables, we applied the logarithm, and the residuals of the mixed model applied on these data were fairly normal. The mixed model (N = 636, 11 word clusters, 20 talker clusters) revealed a main effect of age, $t = 3.52$, $p_{MCMC\text{-}B} = .006$, a main effect of position, $t = -3.52$, $p_{MCMC\text{-}B} = .012$, and an age by position interaction, $t = -3.00$, $p_{MCMC\text{-}B} = .006$. Since none of these effects involves register (our main interest in this paper), we do not dwell on them. Inspection of Table 5 confirms that register did not have a stable effect, reflected in small effect sizes, and this trend was not consistent

across talkers. Thus, in unstressed as well as in stressed vowels, the registers did not diverge in terms of duration.

Fundamental frequency. The distribution of F0 was fairly normal after z-scoring and trimming values more than 2.5SD from the mean (N = 16). The mixed model applied to these data (N = 620, 11 word clusters, 20 talker clusters) revealed no significant main effects or interactions (the largest t =1.90 $p_{MCMC\text{-}B}$ = .348, corresponded to an age by register interaction). Inspection of Table 5 suggests that these mixed model results do not match up perfectly with the non-parametric analyses, which seem to reveal a larger effect of register. Analyses ran on a single vowel /i/ suggested that the effect size observed in speech to four-month-olds is reduced when target vowel is controlled for. The effect of register estimated in the ANOVA was $F(1,11) = 4.41$, $p_B = 0.358$; the effect sizes were 0.49 in four-month-olds and 0.29 in eleven-month-olds; 7 (out of 10) and 6 (out of 10) caregivers in the respective groups exhibited greater F0 in IDS than ADS unstressed vowels. Therefore, we conclude that the impact of register on unstressed vowels is not particularly large or stable across talkers once possible confounds are controlled for, though it is certainly not negligible.

Peripherality. The departures from normality were also present in this distribution, but the distribution of residuals after z-scoring and trimming (excluded N = 49) seemed fairly normal. Results (N = 587, 11 word clusters, 17 talker clusters) showed an effect of register $t = 2.57$, $p_{MCMC\text{-}B}$ = .024. Effect size estimations and binomial distributions, reported in Table 5, concurred; however, the subset analyses on a single vowel in ANOVA does not give a significant register effect. The effect of register estimated in the ANOVA was $F(1,11) = 6.51$, $p_B = 0.161$; the effect sizes were 0.98 in four-month-olds

and 0.64 in eleven-month-olds; 7 (out of 8) and 6 (out of 9) caregivers in the respective groups showed greater peripherality in IDS than ADS unstressed vowels. The latter may be due to the use of untransformed dimensions, and thus the lack of normality in the data – violating one of the assumptions of ANOVA. All things considered, it appears that unstressed vowels are more peripheral in IDS than ADS.

Discussion

During the first few months of life, the primary source of language input to infants is from caregivers. Clearly, IDS plays an important role in language acquisition by attracting infants' attention, showing affect and assisting in social interaction. In this paper we have explored the ways in which IDS may further modulate crucial linguistic input to shape infants' budding language. While a great deal of work suggests that IDS vowels on average have a special acoustic profile, less is known about whether this is a side effect of a simple goal: shorter utterances produced at a slower speech rate. If this hypothesis is true, then the acoustic-phonetics differences between IDS and ADS should be modulated by prosodic boundary alignment and lexical stress. Conversely, if all the acoustic-phonetic differences between IDS are modified across the board, a Reductionist explanation is not supported.

To fill this gap, we collected spontaneous speech in the context of an object categorization task. The first question we asked was whether the established differences between the registers were equally marked in stressed and unstressed vowels; our second and third questions related to a modulation of these effects depending on alignment to an utterance boundary and the addressee's age.

Notice that, given the structure of our corpus, we could only compare the acoustic measures in syllables matched in stress. Since we did not compare *between* stressed and unstressed syllables, we cannot statistically demonstrate whether there is any interaction of register with stress. As the general trend is similar in both types of syllables, we conservatively conclude that IDS modification is *not greatly* modulated by lexical stress for all the three measures (duration, F0, and vowel peripherality).

Extending previous work, vowels in both stressed and unstressed syllables exhibited higher F0 and wider vowel peripherality in IDS as compared to those in ADS. It should be noted that the effect of register for F0 is strong in stressed syllables, but weaker in unstressed syllables. In contrast with previous work our analyses of duration did not yield any significant difference between IDS and ADS in either stressed or unstressed syllables. The latter result is somewhat surprising. We mentioned in the introduction that varied results could be due in part to different methodologies employed, thus we restrict our comparison with prior studies using similar tasks. Recall that both Swanson (1990) and Albin and Echols (1996) found that vowel duration was longer in IDS as compared with ADS; both of these two studies were conducted in a somewhat controlled experimental environment, e.g., they provided either prepared text, or specific objects to assist in elicitation and are thus similar to our study. Different results among these studies may originate from different ages of our addressees. Furthermore, different statistical analyses may also serve to explain variance in results. Given that we adopted the mixed model analyses which controlled for vowel quality together with three more traditional statistical methods, we believe our model has eliminated potential confounds in the measure of duration that may have impacted previous work.

Most importantly, our study also addressed the possibility that such register changes are more marked in salient positions of the utterance, which predicts greater IDS-ADS differences utterance-finally than utterance-medially and also that IDS quality may be due primarily to the frequency of utterance boundaries. Recall that in our sample, caregivers were telling their infants how they would categorize objects or pictures, and thus were incidentally teaching the target words. A preliminary description of our corpus revealed that caregivers tend to produce target words in different prosodic positions depending on the interlocutor, with a higher rate of utterance-final occurrences of target words in speech addressed to the infant. Although this fits well with previous reports on active word teaching to toddlers (Aslin et al., 1996), this is relatively surprising given the fact that the infant addressees in our sample were so much younger. Moreover, there seemed to be a trend for developmental changes. In IDS to four-month-olds, target words occurred in utterance-medial position as frequently as they did in utterance-final position. In ADS only about a quarter of target words occurred in the latter position. IDS to eleven-month-olds was somewhere in between, with about a third of target words occurring in the salient, utterance-final position. This incidental finding could be verified in larger samples, where the source of this variation may also be elucidated. We hypothesize that this could relate to sentences becoming longer in speech to older infants, but it is also possible that caregivers favor different focus strategies at different ages.

At first glance, the higher occurrence rate of target words in utterance-final position in IDS seems to reinforce the view that utterance position could play a non-negligible role in explaining cross-register differences. As noted in the introduction, a reductionist view can be put forward, speculating that much of the differences between the registers

could be due, quite simply, to the higher incidence of prosodic boundaries in IDS, where strengthening might occur. In contrast, a more functional view of IDS would hold that IDS-ADS differences do not arise from the higher incidence of boundaries and words in focus. In our study, the analyses of the acoustic measurements between utterance-medial and -final vowels provide no evidence in support of the reductionist view. When position was entered as a factor, it did not interact significantly with register.

This finding is somewhat surprising given that prior work has sometimes suggested that durational differences in utterance-final position should be more prominent and salient than those in utterance-medial position (see Albin & Echols, 1996; Church, 2002; Kondaurova & Bergeson, 2011). Similarly, although interactions between register and utterance position have been less frequently explored in F0, Kondaurova and Bergeson (2011) report that F0 range differences are more salient in utterance-final position. We speculate that our distinct pattern of results might be due to the nature of the pragmatic situation in which our target words are spoken. Specifically, we used objects to elicit target words from participants, like Albin and Echols (1996) and unlike both Church (2002) and Kondaurova and Bergeson (2011). Given that our study obtained results similar to Albin and Echols (1996), who also used a similar elicitation situation, it is possible that the lack of interaction may due to the pragmatic context. It may be the case that mothers emphasized the target words in both utterance-medial and -final position so to introduce new objects to both their infants and the experimenter.

It should be noted that differences between IDS and ADS cannot be reduced merely to the number of repetitions of target words. An incidental finding in our data was that parents produce more repetitions of the target words in conversation with their infant.

Bortfeld and Morgan (2010) have investigated the acoustic effects of repetition on words pronounced in IDS. They report that first mentions of words are longer, higher pitched and have greater pitch range than do second and subsequent mentions. Thus, as there are more and more repetitions in IDS, words come to resemble the typical profile for IDS. Thus, these differences in number of repetitions cannot explain differences between IDS and ADS.

The third question we set out to answer is whether IDS-ADS modulations varied with infant age, for which we compared speech to four- and eleven-month-olds. Analyses revealed that the duration, F0, and vowel peripherality was *not* dependent on infant age in either stressed or unstressed syllables. In other words, IDS and ADS differences were *not* modulated by age in our sample. This is not necessarily counter-evidence to studies in which age-specific effects are found. For example, IDS-related changes linked to age could be primarily explained by the caregivers' concepts of their child's phonological and lexical development. Thus, it is possible that in the sample we recorded, the caregivers of both four- and eleven-month-olds were persuaded that neither group of infants was learning in a substantially different way. This is likely given that the two age groups in our study are young and primarily preverbal. Future work could better test age-related predictions by actually measuring caregiver's beliefs about their infants' competence through questionnaires, or possibly by modifying it; for example, a subgroup of parents could be informed that their child is indeed learning words.

The current study is limited in several respects. First, as noted above, vowel qualities were not matched in stressed and unstressed syllables, which prevented a direct comparison. Second, focus and repetition could play a role in the present results. An ideal

follow-up would undertake a more informed descriptive analysis of the types of focus structures found and whether they change with infant development. This project would furthermore connect with classical work on communicative intent in IDS (e.g., Papousek & Papousek, 1979). Finally, our selection of target items included only trochees, such that the stressed syllable was also the first syllable of the word. It would be interesting for future work to replicate the present exploration in words with other stress patterns (e.g., iambs).

In summary, the present corpus is consistent with prior studies showing higher F0 and wider vowel peripherality in IDS than in ADS; moreover, this study has revealed that IDS-ADS differences are *not* modulated by prosodic boundaries, lexical stress, or addressee's age. Thus, infants are receiving slightly different input directly addressed to them regardless of whether or not they are attending primarily to prosodically prominent positions.

References

Albin, D. D., & Echols, C. H. (1996). Stressed and word-final syllables in infant-directed speech. *Infant Behavior and Development, 19*, 401-418.

Aslin, R. N., Woodward, J., Lamendole, N., & Bever, T. G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan & K. Demuth (Eds.), *Signal to Syntax: Boostrapping from Speech to Grammar in Early Acquisition (117-134)*. Mahwah, NJ: Erlbaum.

Baayen, R. H. (2008). languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". R package version 0.953.

Bates, D., & Sarkar, D. (2007). lme4: Linear mixed-effects models using S4 classes.

Bernstein, R. N. (1986). Durational cues which mark clauses boundaries in mother-child speech. *Journal of Phonetics, 12(2*), 303-309.

Boersma, P., & Weenink, D. (2011). *Praat: doing phonetics by computer (Version 5.2.28)* [Computer program]. Retrieved 11 July, 2011, from http://www.praat.org/.

Bortfeld, H., & Morgan, J. L. (2010). Is early word-form processing stress-full? How natural variability supports recognition. *Cognitive Psychology, 60*(4), 241-266.

Bradlow, A. R., Torretta, G. M., & Pisoni, D. B. (1996). Intelligibility of normal speech I: Global and fine-grained acoustic-phonetic talker characteristics. *Speech Communication, 20*(3–4), 255-272.

Church, R. (2002). *Prosodic modification in infant-directed speech*. (Master Thesis). University of British Columbia, Canada.

Cristia, A. (2013). Input to language: The phonetics and perception of infant-directed speech. *Language and Linguistics Compass, 7*(3), 157-170.

Curtin, S., Mintz, T. H. & Christiansen, M. H. (2005). Stress changes the representational landscape: Evidence from word segmentation. *Cognition, 96,* 233-262.

Cutler, A. & Norris, D.G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113-121.

Escudero, P., Boersma, P., Rauber, A. S. & Bion, R. A. H. (2009). A cross-dialect acoustic description of vowels: Brazilian and European Portuguese. *Journal of the Acoustical Society of America, 126*, 1379-1393. doi: 10.1121/1.3180321

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development, 8*, 181-195.

Fernald, A. (2000). Speech to infants as Hyperspeech: Knowledge driven processes in early word recognition. *Phonetica, 57*, 242-254.

Fernald, A., & Mazzie, C. (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology, 27*, 209-221.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20,* 104-113.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1,* 126-152.

Gleitman, L. R., Gleitman, H., Landau, B., & Wanner, E. (1988). Where learning begins: Initial representations for language learning. In F.J. Newmeyer (Ed.), Linguistics: *The Cambridge Survey: Vol. 3.* Language: Psychological and biological aspects. Cambridge, England: Cambridge University Press.

Houston, D., Santelmann, L. & Jusczyk, P. (2004). English-learning infants' segmentation of trisyllabic words from fluent speech. *Language and Cognitive Processes, 19,* 97-136.

Jacobson, S., Ward, N., & Sundare, M., (2011). Acoustic correlates of stress in Spanish infant-directed speech. Journal of Acoustic Society of America, *130*(4), 2428-2428.

Jakobson, R. (1941). Kindersprache, Aphasie, und allgemeine Lautgesetze. Mouton: The Hague.

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The Beginnings of Word Segmentation in English-Learning Infants. *Cognitive Psychology, 39*(3–4), 159-207. doi: http://dx.doi.org/10.1006/cogp.1999.0716

Jusczyk, P., & Thompson, E. (1978). Perception of a phonetic contrast in multisyllabic utterances by 2-month-old infants. *Perception and Psychophysics, 23(2),* 105-109.

Kaplan, P. S., Bachorowski, J. A., Smoski, M. J., & Hudenko, W. J. (2002). Infants of depressed mothers, although competent learners, fail to learn in response to their own mothers' infant-directed speech. *Psychological Sciences, 13*(3), 268-271.

Kitamura, C., & Burnham, D. (2003), Pitch and Communicative Intent in Mother's Speech: Adjustments for Age and Sex in the First Year. *Infancy, 4(1),* 85-110.

Kitamura, C., Thanavishuth, C., Burnham, D., 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 Behavior and Development, 24(4),* 372-392.

Kondaurova, M. V., & Bergeson, T. R. (2011). The Effects of Age and Infant Hearing Status on Maternal Use of Prosodic Cues for Clause Boundaries in Speech. *Journal of Speech, Language and Hearing Research (Online), 54*(3), 740-754A.

Kondaurova, M. V., Bergeson, T. R., & Xu, H. (in press). Age-Related Changes in Prosodic Features of Maternal Speech to Prelingually Deaf Infants with Cochlear Implants. *Infancy*.

Lehiste, I. (1970). *Suprasegments*. Cambridge, MA: MIT Press.

Lipsey M. W., & Wilson, D. B. (2001). *Practical meta- analysis.* Thousand Oaks, CA: Sage.

Liu, H., Tsao, F. M., & Kuhl, P. K., (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: a longitudinal study. *Journal of Child Language, 36(4),* 909-922.

McMurray, B., Kovack-Lesh, K. A., Goodwin, D., & McEchron, W. (2013). Infant directed speech and the development of speech perception: enhancing development or an unintended consequence? *Cognition, 129*(2), 362-378. doi: 10.1016/j.cognition.2013.07.015

Papousek, H., & Papousek, M. (1979). Early ontogeny of human social interaction: It's biological roots and social demensions. In *Human ethology: Claims and limits of a new discipline*, edited by M. Von Granach, K. Foppa, W. Lepenies, & D. Ploog. Cambridge: Cambridge University Press.

Papousek, M., Papousek, H., & Symmes, D. (1991). The meanings and melodies in motherese in tone and stress languages. *Infant Behavior and Development, 14,* 415-440.

Peters, A.M. (1983). *The units of language acquisition.* Cambridge, England: Cambridge University Press.

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language, 57,* 24-48.

Shute, B., & Wheldall, K. (1989). Pitch alterations in British motherese: some preliminary acoustic data. *Journal of Child Language, 16*(03), 503-512.

Snow, C. E. (1977). Mothers' speech research: From input to interaction. In C. E. Snow & C. A. Ferguson (Eds.), *Talking to children: Language input and acquisition*.Cambridge, England: Cambridge University Press.

Stern, D. N., MacKain, K.., & Spieker, S. (1982). Intonation contours as signals in maternal speech to prelinguistic infants. *Developmental Psychology, 18*, 727-735.

Stern, D. N., Speiker, S., Barnett, R. K., & Mackain, K. (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language, 10*, 1-15.

Swanson, L. A. (1990). *Vowel duration in mothers' speech to young children*. (Doctoral disseartation). Retrieved fromU.M.I Database. (AAT 9104712)

Thanavisuth, C. Luksaneeyanawin, S. (1998). Acoustic Quality of Thai Infant Directed Speech and Adult Directed Speech. In Robert Mannell and J. Robert-Ribes (eds), *Proceedings of the 5th International Conference on Spoken Language Processi*ng-ICSLP 98, 445-448.

Thiessen, E.D. (2005). Infant-directed speech facilitates word segmentation. *Infancy, 7*(1)*,* 53-71.

Turk, A. E., & White, L. (1999). Structural influences on accentual lengthening in English. *Journal of Phonetics, 27(2),* 171-206.

Wightman, C. W.Shattuck-Hufnagel, S., Ostendorf, M, & Prince, P. J.(1992). Segmental durations in the vicinity of prosodic boundaries. *Journal of Acoustic Society of American, 91(3),* 1707-1717.

Appendix A

Speech samples from IDS and ADS session (M: mother; E: experimenter).

IDS session:

M: *Look at these! These're measuring spoons. This one. We have some teaspoons. We have two teaspoons and one tablespoon. See that how they can all these three of them be measuring spoons. But one is different. The two teaspoons are smaller and the tablespoon isn't. We've gotten a half teaspoon, and one teaspoon and one tablespoon. Yes, that one is a little bigger.*

ADS session:

E: How about this one?

M: *Teaspoons. We talked about the measuring spoons. There are two different kinds of measuring spoons, teaspoons and tablespoon; and the tablespoon is bigger.*

E: *Did you talk about the two different sizes?*

M: *I think I did. I think I said there were half teaspoon, and one teaspoon, and a tablespoon.*

E: *Great!*

Table 1

Average number of target word tokens per participants (10 four-month-olds and 10 eleven-month-olds) included in the acoustic analyses in both stressed and unstressed syllables.

Syllable Type	Age	Register	Utterance-final	Utterance-medial
Stressed	Four-month-old	ADS	9.3	20.3
		IDS	18.6	19.4
	Eleven-month-old	ADS	8.1	23.3
		IDS	10.3	18
Unstressed	Four-month-old	ADS	3.7	10.9
		IDS	10.1	10.4
	Eleven-month-old	ADS	4	11.7
		IDS	3.9	8.9

Table 2

Frequency of each target word tokens in ADS and IDS speech in stressed and unstressed syllables

Target word	Transcription	Stressed		Unstressed	
		ADS	IDS	ADS	IDS
bacon	ˈbe.kən/ˈbe.kn	30	26	-	-
basil	ˈbe.səl/ˈbe.sl	18	19	-	-
beetle	ˈbi.təl/ˈbi.tl	26	65	-	-
benji	ˈbɛnʤi	29	25	29	25
dancer	ˈdæn.sɚ	27	25	-	-
daycare	ˈde.kɛr	25	9	-	-
decker	ˈdɛ.kɚ	26	25	-	-
disney	ˈdɪz.ni	39	35	39	35
pansy	ˈpæn.zi	36	31	36	31
paper	ˈpe.pɚ	25	25	-	-
pedal	ˈpɛ.dəl/ ˈpɛ.dl	34	32	-	-
pegboard	ˈpɛg.bɔrd	15	18	15	18
pencil	ˈpɛn.səl/ ˈpɛn.sl	24	35	-	-
pendant	ˈpɛn.dənt/ˈpɛn.dnt	22	23	-	-
pepsi	ˈpɛp.si	30	28	30	28
pesto	ˈpɛs.to	31	46	31	46
picnic	ˈpɪk.nɪk	21	18	21	18
piglet	ˈpɪg.lɪt	15	32	15	32
shopping	ˈʃɑ.pɪŋ	21	15	21	15
tassel	ˈtæ. səl/ˈtæ.sl	22	20	-	-
teaspoon	ˈti.spun	31	28	31	28
teddy	ˈtɛ.di	35	57	35	57
tender	ˈtɛn.dɚ	28	26	-	-
total	23	611	663	303	333

Table 3

Means (Standard Deviations) for each acoustic dimension within each age group, prosodic position, and register, for stressed vowels.

	Four-month-olds				Eleven-month-olds			
	Utterance-finally		Utterance-medially		Utterance-finally		Utterance-medially	
	ADS	IDS	ADS	IDS	ADS	IDS	ADS	IDS
Duration	0.082 (0.028)	0.078 (0.035)	0.067 (0.019)	0.070 (0.020)	0.072 (0.020)	0.087 (0.031)	0.060 (0.011)	0.063 (0.015)
Pitch	5.099 (0.551)	5.939 (0.629)	5.342 (0.488)	5.995 (0.785)	4.925 (0.900)	5.773 (0.581)	5.382 (0.375)	5.761 (0.426)
Peripherality	1.597 (0.508)	1.720 (0.545)	1.367 (0.228)	1.535 (0.330)	1.366 (0.401)	1.606 (0.461)	1.439 (0.332)	1.554 (0.410)

Table 4

Means and Standard Errors for each acoustic dimension within each age group, prosodic position, and register, for unstressed vowels.

	Four-month-olds				Eleven-month-olds			
	Utterance-finally		Utterance-medially		Utterance-finally		Utterance-medially	
	ADS	IDS	ADS	IDS	ADS	IDS	ADS	IDS
Duration	0.196 (0.092)	0.172 (0.066)	0.080 (0.017)	0.086 (0.016)	0.114 (0.039)	0.132 (0.050)	0.083 (0.027)	0.075 (0.018)
Pitch	4.859 (0.518)	6.125 (1.358)	5.047 (0.766)	5.845 (1.002)	5.435 (1.043)	5.007 (1.318)	5.041 (0.453)	5.642 (0.696)
Peripherality	2.086 (0.702)	2.089 (0.328)	1.378 (0.343)	1.849 (0.335)	1.940 (0.547)	2.193 (0.671)	1.833 (0.451)	2.110 (0.497)

Table 5

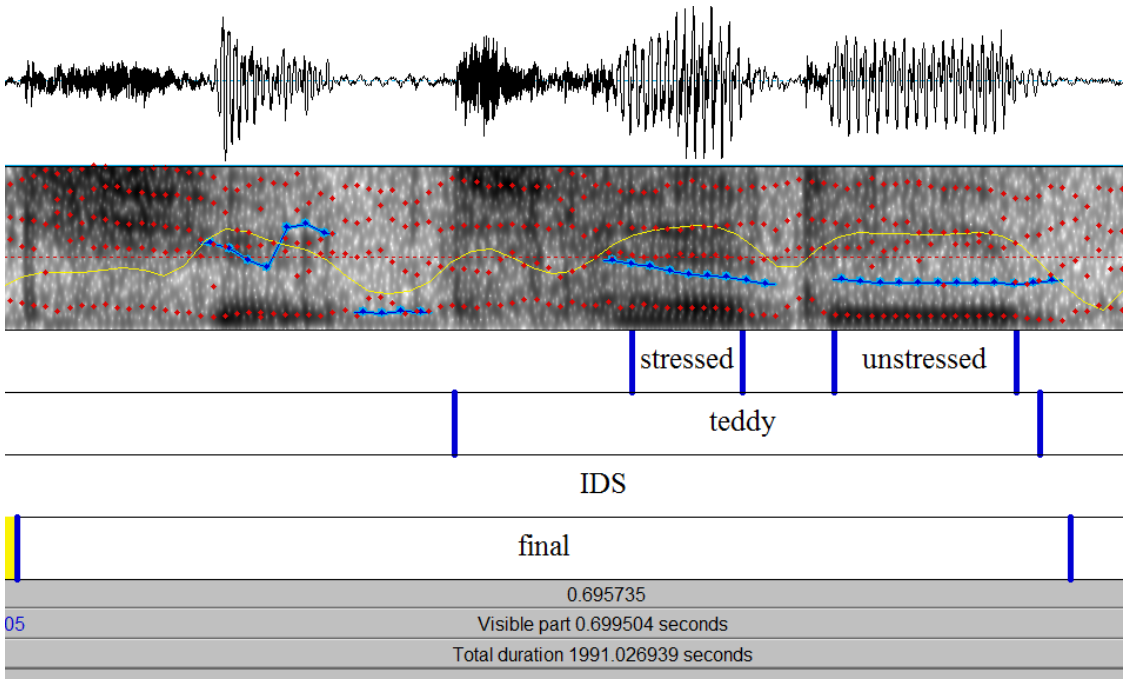
Effect of register according to four complementary analyses: a mixed model on normalized data which controls for addressee age, sentence position, talker, and lexical item; an ANOVA on the untransformed data from single vowel, where age, sentence position and talker are declared; the effect size on the data from speech to 4- and 11-month-olds (left and right respectively); and the number of caregivers showing the expected pattern (in the same order). Sections in bold indicate results that are stable across analyses.

Syllable Type	Analyses	Duration	Pitch	Peripherality
Stressed	Mixed model	**None**	**Main effect**	**None**
	ANOVA	**None**	**Main effect**	**None**
	Effect size	**0.28 0.30**	**2.08 0.83**	**0.76 0.37**
	Binomial	**7/10 5/10**	**10/10 8/10**	**7/8 6/9**
	Generalization	***No/small effect***	***Large effect***	***Small effect***
Unstressed	Mixed model	**None**	None	**Main effect**
	ANOVA	**None**	None	**None**
	Effect size	**0.18 -0.29**	1.02 0.66	**0.99 0.45**
	Binomial	**7/10 4/10**	9/10 8/10	**7/8 7/9**
	Generalization	***No/small effect***	*Medium effect*	***Large effect***

Figure Captions

Figure 1. Coding samples in Praat.

Figure 1


stressed
unstressed
teddy
IDS
final
0.695735
05
Visible part 0.699504 seconds
Total duration 1991.026939 seconds