Journal of Child Language (2024), **51**, 1193–1212
doi:10.1017/S030500092300034X

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

Acoustic cues to phrase and clause boundaries in infant-directed speech: Evidence from LENA recordings

Tianlin WANG[1], Elie ChingYen YU[1], Rong HUANG[1,3] and Jill LANY[2]

[1]University at Albany, State University of New York, USA
[2]University of Liverpool, UK
[3]University of Connecticut, USA
Corresponding author: Tianlin Wang; Email: twang23@albany.edu

(Received 18 July 2022; revised 06 May 2023; accepted 09 May 2023)

Abstract

Infant-directed speech (IDS) produced in laboratory settings contains acoustic cues, such as pauses, pitch changes, and vowel-lengthening that could facilitate breaking speech into smaller units, such as syntactically well-formed utterances, and the noun- and verb-phrases within them. It is unclear whether these cues are present in speech produced in more natural contexts outside the lab. We captured LENA recordings of caregiver speech to 12-month-old infants in daylong interactions ($N = 49$) to address this question. We found that the final positions of syntactically well-formed utterances contained greater vowel lengthening and pitch changes, and were followed by longer pauses, relative to non-final positions. However, we found no evidence that these cues were present at utterance-internal phrase boundaries. Results suggest that acoustic cues marking the boundaries of well-formed utterances are salient in everyday speech to infants and highlight the importance of characterizing IDS in a large sample of naturally-produced speech to infants.

Keywords: Infant-Directed Speech; acoustics; LENA recordings

Acoustic cues to phrase and clause boundaries in infant-directed speech: Evidence from LENA recordings

Across many cultures, the speech that adults use when talking to infants differs, both acoustically and structurally, from the speech they use when talking to other adults and even older children (Fernald & Mazzie, 1991; Hilton et al., 2022; see also Cristià, 2013 for a review). This infant-directed-speech (IDS) is generally higher in pitch than adult-directed speech (ADS), contains more variability in its pitch contours (Stern et al., 1983), and consists of shorter utterances with simpler syntax (Fernald et al., 1989). These features of IDS are likely to serve several important functions across development. For example, it is relatively uncontroversial that IDS is potent in capturing infants' attention: IDS evokes stronger neural responses than ADS (Pena et al., 2003), and seminal work by Cooper and

colleagues showed that infants prefer to listen to IDS over ADS within the first months of life (Cooper & Aslin, 1990; Cooper et al., 1997; Fernald, 1985; Pegg et al., 1992). An international replication across 67 labs recently confirmed that infants prefer to listen to IDS over ADS, and suggests that the preference for IDS increases with age and with the familiarity of the language (Many Babies Consortium, 2020). The attention-getting features of IDS are likely to promote language development by ensuring that infants attend to and encode the speech in their environment.

There is evidence that IDS also serves a communicative, regulatory function in caregiver-infant interactions. IDS conveys affective information in its pitch, and the positive affect within IDS is an important determinant of infants' preference for it (Singh et al., 2002). Across languages, rising contours are consistently used to attract attention and falling contours used to calm or soothe infants. Infants respond differentially to these contours, whether or not they are produced with words from infants' native language, suggesting that IDS may effectively communicate nonsymbolic "meaning" to infants (Fernald, 1992).

Although there is strong evidence that infants preferentially attend to IDS, and that aspects of its exaggerated prosody are communicative, it is less clear whether the acoustic properties of IDS render it a good signal for learning language structure. On the one hand, there is evidence that infants' exposure to IDS is related to their language development. For example, Ramírez-Esparza and colleagues (2014) found that infants who hear more IDS at 11- and 14-months have larger vocabularies, both concurrently and about one year later at age 2. Lany and Shoaib (2020) found relationships between the amount of IDS infants hear and infants' vocabulary size and ability to learn long-distance grammatical dependencies, especially in females. Infants are better able to segment words in fluent IDS than ADS (Thiessen et al., 2005) as well as map them to meaning (Graf Estes & Hurley, 2013; Ma et al., 2011). These findings could suggest that IDS is a superior learning signal, providing clearer input that is more easily encoded and recognized, and which contains more transparent structure for learning. However, these effects could instead arise if infants who hear more IDS simply pay more attention to language. Consistent with the possibility that attention is an important determinant of the connection between IDS and language learning, Nencheva et al. (2021) found that words with a "hill" contour (inverted U-shape) seem to attract infant attention more than words with other contours, and that they tend to learn these words better. Thus, these findings do not rule out the possibility that infants would learn just as well from ADS if they were to pay attention to it more.

Another line of work has approached the question of whether IDS is a good learning signal by quantifying the acoustic properties of IDS. These studies have yielded mixed evidence. For example, one possibility that has received substantial attention is the hyperspeech hypothesis (Cristià & Seidl, 2014), which suggests that acoustic exaggerations in IDS lead to more differentiation among similar vowels, relative to ADS, which may support more accurate speech-sound perception (Cristià, 2008; Liu et al., 2003). This is referred to as hyperarticulation, or a reduction in the acoustic overlap among similar sounds. Several studies of infant-directed speech produced in the lab, in which mothers were asked to use a set of preselected target words, reported that there is less overlap between some vowels in IDS relative to ADS (Burnham et al., 2002; Kuhl et al., 1997; Liu et al., 2003). However, more recent studies, especially those using large datasets recorded in more naturalistic contexts either failed to find evidence that phonemes are hyperarticulated in IDS, or found such evidence in relatively limited contexts, such as in words that receive prosodic focus (Adriaans & Swingley, 2012; Cristià & Seidl, 2014; Martin et al., 2015).

The focus of this paper is the extent to which IDS contains cues that might facilitate learning about simple syntactic structure by facilitating breaking speech into smaller

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

units. There is a rough correspondence between prosodic and syntactic structure in adults' speech. In particular, there are often acoustic changes at the boundaries of syntactic units such as clauses and phrases that could lead these units to be perceptually grouped (Nespor & Vogel, 1986; Selkirk, 1980). For example, across many languages, the ends of utterances, which often correspond to the end of a clause, are generally characterized by long pauses, lengthening of syllables (particularly vowels), and pronounced pitch changes. Some of these acoustic characteristics may derive from the biomechanics of speech production. For example, pitch is affected by changes in the muscles that control exhaling while speaking (Honda, 2004), and there may also be a motor basis for segment-final lengthening (Cho, 2016; Paschen et al., 2022). The upper body movements used in gestures that are temporally coordinated with speech can also impact the acoustic characteristics of boundary locations (Pouw et al., 2020).

The acoustic cues to the boundaries of syntactic units are not perfectly reliable, as multiple factors influence the duration and pitch of a given syllable (e.g., its phonetic and lexical content, as well as discourse structure), and these influences vary across languages. Likewise, when pauses occur, they often do so at utterance boundaries, which do not always correspond to phase or clause boundaries (Goldman-Eisler, 1972). And, many phrases and clauses are not followed by a pause. For example, in English, utterance-initial noun phrase subjects, like "she" or "they", are unlikely to be followed by pauses. Furthermore, the cues marking phrase boundaries are generally weaker than those marking clause boundaries (Cooper & Paccia-Cooper, 1980). Nonetheless, even imperfect cues could potentially help infants to identify linguistically-relevant groupings, a hypothesis referred to as Prosodic Bootstrapping (Gleitman & Wanner, 1982). Utterances in IDS often consist of a single, very short clause (e.g., Kaye, 1980). Thus, sensitivity to utterance-final acoustic features could help infants to break speech down into smaller chunks that roughly correspond to simple, syntactically-well-formed units.

There is evidence that these acoustic features of boundary locations can lead infants to perceive distinct units within the speech stream. For example, English-learning 9-month-olds preferred to listen to passages of IDS in which pauses naturally occurred at clause and phrase boundaries over passages in which pauses were artificially inserted at non-boundary locations (Hirsh Pasek et al., 1987; Jusczyk et al., 1992). Critically, the natural phrase and clause boundary locations were characterized by larger pitch changes and greater syllable lengthening than non-boundary locations. Infants failed to show such a preference when the same materials were produced in ADS (Kemler Nelson et al., 1989). These results suggest that infants prefer speech in which pauses coincide with larger pitch changes and greater lengthening, and that these cues may be especially salient in IDS (Jusczyk et al., 1992; see also Seidl, 2007). Dutch-learning 6-month-olds and German-learning 8-month-olds also appear to be sensitive to the presence of these cues in their native language (Johnson & Seidl, 2008; Wellmann et al., 2012).

Nonetheless, this evidence that there are cues marking syntactic structure in IDS, and that infants are sensitive to them, comes from studies using carefully designed materials read aloud that are not representative of the IDS that infants hear in their everyday experience. Thus, it is important to determine whether these cues are present in the IDS that caregivers use spontaneously. Fisher and Tokura (1996) made a start at addressing this by recording 3 mothers who were native speakers of American English, and 3 who were native speakers of Japanese, speaking to their 13-14-month-old infants. These recordings primarily took place in a laboratory setting. Fisher and Tokura found that in both languages, vowel lengthening and pitch changes were more pronounced in syllables that occurred at syntactically well-formed utterance boundaries, relative to syllables that occurred within utterances (Fisher & Tokura, 1996). Likewise, pauses at

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

well-formed utterance boundaries were significantly longer than those found within utterances in both English and Japanese. Note that these effects were detected without considering the identity of vowels, or whether the vowels occurred in stressed vs. unstressed syllables, suggesting that these cues are likely to be salient even to infants who are not yet able to accurately identify speech-sounds or to isolate lexical stress.

Fisher and Tokura (1996) also found evidence that there are acoustics cues at utterance-internal phrase boundaries, but they appeared to be less robust. For example, in English, only syllable lengthening was more pronounced at boundaries between noun and verb phrases than at nonboundary locations (Fisher & Tokura, 1996), and only within sentences that contained more than one word in the noun phrase. In Japanese, such boundaries were marked by pitch changes, rather than by lengthening. These results highlight the fact that pauses and pitch and duration changes occur at both utterance and phrase boundary locations in many languages, but there are cross-linguistic differences in how pronounced and reliable these individual cues are (see also Johnson & Seidl, 2008).

In a more recent systematic review, Ludusan et al. (2016) found some evidence that these same acoustic cues marking prosodic boundaries are exaggerated in IDS relative to ADS, but noted that most of the included studies used a very small sample of speakers and/or utterances, and focused on American English. Ludusan et al. also tested whether such differences exist in a larger set of recordings of Japanese IDS in which mothers (N = 22) were recorded in play sessions that took place in a lab setting. They found evidence that syllable lengthening and pause duration at boundaries were present and potentially useful to learners, though pitch changes were not.

In sum, evidence from recordings (largely of American English) made in the lab suggests that utterance, clause, and phrase boundaries are characterized by long pauses, syllable lengthening, and pronounced pitch changes, to varying degrees (see Ludusan et al., 2016 for a review). Infants' sensitivity to these acoustic cues may play a critical role in language learning by helping them to break speech down into units that correspond to syntactic units, especially clauses that constitute entire utterances, but potentially also phrases (Hirsh Pasek et al., 1987; Jusczyk et al., 1992). However, it is less clear whether IDS produced in interactions outside the lab contains similar cues to language structure. Recent research with North American samples suggests that several important measures of speech to infants in lab experiments and play sessions can differ substantially from measures made from recordings of more naturalistic interactions at home (e.g., Bergelson et al., 2018; Tamis-LeMonda et al., 2017). For example, mothers talk more and use a richer vocabulary in structured play sessions than in recordings of typical interactions at home involving feeding, bathing, and both joint and solitary play (Tamis-LeMonda et al., 2017). Bergelson et al. (2018) reported a similar difference for speech quantity in daylong, unstructured audio recordings vs. hour-long video-recorded play sessions. Likewise, there are potentially important differences between lab-based speech and IDS produced in more unconstrained settings captured in daylong recordings of typical daily activity. For example, in lab-based recordings, the adult and infant are typically positioned face-to-face, with the infant in a seat or carrier. Moreover, steps are often taken to ensure that background activity is reduced, and thus there is typically very little to distract the infant. Additionally, the adults generally try to keep their baby engaged in a communicative interaction, rather than feeding, dressing, changing the infant, etc. All of these characteristics of everyday interactions could lead to differences in the manifestation of the IDS caregivers use within them, in comparison to lab-based interactions.

A handful of North American studies have characterized IDS produced outside the lab, and they provide some evidence that the IDS produced in lab contexts bears a general similarity to more naturalistic IDS. For example, Stern et al. (1983) found that

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

fundamental frequency, pause duration, repetitiveness, and MLU were similar in an analysis of 6 recordings of mothers talking to their 4-month-olds in a lab setting and at home. Soderstrom et al. (2008) tested whether the acoustic cues typically associated with syntactic boundaries were present in IDS produced by 2 English-speaking mothers in an everyday interaction at home captured when their infants were 9 months of age. They found clear evidence for pronounced pauses and duration and pitch changes at utterance boundaries, as well as utterance-internal clause boundaries, consistent with results from the lab-based recordings of Fisher and Tokura (1996). Also similar to Fisher and Tokura, they did not find evidence of these cues at utterance-internal phrase boundaries when considering all utterances, but these cues did occur at phrase boundaries in Yes/No questions, which contain more than one word in the noun phrase.

These results provide preliminary evidence that IDS used in everyday interactions contains acoustic cues that could facilitate grouping speech into syntactically-relevant units. However, the evidence comes from recordings of only 2 dyads, using small samples of speech, and it is important to test whether these cues are also present in a dataset containing a larger number of adult speakers. Thus, we built on this work by testing whether acoustic cues that could help infants to identify well-formed syntactic units, such as clauses or phrases, are present in IDS produced in a more naturalistic recording context – infants' homes.

Method

Participants

A total of 57 12-month-old infants (29 females) and their families participated in the study. All infants were the youngest family member and had an average of one sibling in the house. Eight participants were later excluded due to 1) a parental report of developmental delay in hearing, vision, or language development (N = 6), 2) being born before 36 gestation weeks or weighing less than 5lbs 5oz at birth (N = 1), or 3) being exposed to languages other than English for more than 15 hours a week (N = 1). Therefore, recordings from a total of 49 infants were included for analysis.

Participant families were mostly white (96%) and monolingual, and English was the primary language spoken in the homes of all of the infants. The mother of each infant was identified as the primary caregiver. On average, the mothers obtained a 4-year college degree. The household income ranged from $50,000 to $75,000. Recruitment and data collection were done in a mid-size Midwestern city in the U.S.

Procedure

Following other studies using multi-day recording procedures (e.g., Weisleder & Fernald, 2013), parents were asked to record eight hours of continuous audio on each of two consecutive days, for a total of 16 hours. In order to capture spontaneously produced speech in the home environment, infants wore a special vest with a built-in Language Environment Analysis (LENA) system (LENA Foundation, 2018). In the vest, there is a pocket on the chest that fits the digital language processor (DLP). The DLP records the audio within an infant's language environment and can hold 16 hours of recordings.

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

Some measures of IDS from these recordings were reported in Lany and Shoaib (2020). In that work, utterances were sampled and identified as IDS or ADS, and the quantity of IDS was related to measures of artificial language learning and language development. Note that in that study, the method of sampling utterances was substantially different, and specific acoustic characteristics of IDS were not investigated. The utterances included in the present study were acoustically coded for the analyses reported in this study, and these measures have not been reported elsewhere.

The audio recordings from each participant were pre-processed using LENA software. The software automatically excludes crying, whining, laughing, and ambient noise. The report provided an estimate of the adult word count (AWC) in five-minute intervals across the 16 hours of recordings. Because the amount of speech an infant hears can vary substantially from day to day (e.g., from 6,000 to 19,000 words per day, Gilkerson & Richards, 2009), we chose a total of four 30-minute audio clips, with two clips from each day for each participant, in order to obtain a more representative sample. The two 30-minute clips with the highest AWC estimates on each day were selected for transcription and further analysis. One participant recorded the 16-hours of audio over 3 days. For this participant, one 30-minute clip was chosen from each day and the fourth 30-minute clip was the segment with the second highest AWC estimate over the 3 days. One participant recorded over 4 days, and the clip that contained the highest estimate of AWC from each of those days was used. The LENA software was designed to identify and exclude electronic speech (e.g., from television, radio etc.) from word counts during data preparation, However, it does not always do so successfully (Bergelson et al., 2018), and thus we manually examined selected clips for the presence of electronic speech. If any chosen clip consisted of more than 50% electronic speech, it was replaced with one containing the next-highest AWC for that day. A total of 10 clips were manually identified and replaced due to high electronic speech content.

Transcription and Utterance Coding

The LENA2CHAT command was used to convert LENA files for further transcription and coding. Each 30-minute-long clip was then transcribed in CLAN following the CHAT guidelines (MacWhinney, 2000). Four CLAN files were thus generated for each participating infant. The LENA software automatically produced lines in the CLAN files that corresponded to vocalization in the LENA recording. The LENA-generated speaker codes were also imported into CLAN using LENA2CHAT. When working with the resulting CLAN files, human coders found many of the LENA codes to be incorrect (e.g., background noise was sometimes incorrectly classified as *adult speaker,* while speech was sometimes missed and untagged). Human coders therefore manually checked each line in the CLAN files and transcribed all speech into text. In addition to the utterances identified by the LENA algorithms, coders used syntax (where the utterance is grammatically different from its context) to identify utterances. These initial coding steps are detailed in Thompson (2019). Specifically, the speech from all selected audio clips was transcribed by trained research assistants. Based on the speaker codes generated by LENA and imported into CLAN, assistants then identified the speaker of each utterance as a male adult near the infant, male adult far from the infant, female adult near the infant, female adult far from the infant, the target infant, another child near the infant, another child far from the infant, or noise or electronic speech. Note that these codes do not specify who exactly was speaking (e.g., if it was the primary caregiver or another adult). Thus, the

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

research assistants used a log provided by parents that detailed what occurred during the recording period (e.g., went to friend's house, played with sister), and who was present during the events, and their experience with that family's audio files, to replace the LENA-generated speaker codes with new codes that provided a more detailed identity of the speaker (e.g., mother, father, brother, sister, grandmother, unknown female adult, unknown male adult, unknown female child, unknown male child, etc.). Once the transcription was complete, a second person reviewed it for errors and met with the original coder to discuss and resolve any disagreements.

Similarly, two research assistants coded each utterance to determine whether it was directed towards the infant or to someone else using the log as well as the audio content. The coding utilized the GEM feature in CLAN, which allows specific sections of the transcript to be tagged. In this case, the audio corresponding to each utterance was tagged using the GEM feature to show if the speech was overheard by the infant or directed towards the child. The research assistants independently coded whether the speech was directed to the infant, and then compared coding results to discuss and resolve any disagreements. Only speech directed toward the infant is included in the subsequent acoustic coding and analyses.

We next followed the syntactic coding schemes reported in Fisher and Tokura (1996) to determine what constitutes an utterance. We first transcribed the speech and identified segments that should be separated from the next word by a comma or a full stop. This allowed us to classify utterances into types, such as full clauses versus fragments, and to identify major syntactic units within them. Following Fisher and Tokura's (1996) approach, only full clauses were included for later analyses.

We first coded full clauses as either declaratives, Yes/No questions, or WH-questions. Two coders identified the utterance-type categories independently at first, and then discussed and resolved any discrepancies that occurred until 100% agreement was achieved. We then followed Fisher and Tokura's coding of the *first major phrase* and identified it as the subjects of declarative sentences and Yes/No questions, as in Example 1 below, the fronted WH-phrases in WH-questions, as in Example 2, as well as initial locative phrases, as shown in Example 3 (italicized words mark first major phrases in all three examples). The utterances in examples 1-3 come from separate recordings and were not spoken sequentially. Following their approach, we also excluded imperatives from phrase boundary coding. As a result, each utterance in the present study had a maximum of one coded internal phrase boundary. Twenty-five percent of utterances were randomly selected for reliability coding by two coders. There was a 99.46% agreement rate on the identification of words preceding the subject-verb boundaries. The subject-verb boundaries occur at the end of the italicized speech in each example.

1a. Do you see a bridge?
1b. The polar bear wants to slide.
2a. What's in there?
2b. Where'd you get that little ouchie on your head?
2c. Who do you see?
3a. There you go.
3b. There's plenty of different things.
3c. Here comes the door.

Fisher and Tokura (1996) found evidence of phrase-final prosodic bracketing only in utterances with more complex first major phrases (i.e., instances in which there was more

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

than one syllable before subject-verb boundary), and thus we further coded for the complexity of internal syntactic structure using the same criterion. In our study, when there was only one syllable before the subject-verb boundary, these syllables always corresponded to monosyllabic words. For utterances with more complex first major phrases (i.e., 2 or more syllables prior to the subject-verb boundary), there were multiple words before the boundary. Thus, for each utterance we coded whether the first Noun Phrase (NP) contained one or more words preceding the subject (NP>0). As Fisher and Tokura pointed out, pronouns (e.g., "She handed you Snoopy") and pro-locatives (e.g., "Here's your milk") are unlikely to be stressed and therefore tend not to precede the boundary of a prosodic unit within an utterance (e.g., Gee & Grosjean, 1983; Gerken et al., 1994; Read & Schreiber, 1982, as cited in Fisher & Tokura, 1996). Within this subset, Yes/No questions such as Example 4, declaratives such as Example 5, WH-questions such as Example 6, and utterances with an initial conjunction such as Example 7, were included.

4. *Do you* see a bridge?
5. *The polar bear* wants to slide.
6. *What time* did he wake up?
7. *Therefore it* changes from under me.

Acoustic Coding

In preparation for acoustic coding, we first segmented the audio recordings of all speech into separate files, with each file containing one utterance. To accurately identify the utterance-final position and its related acoustic features, including pause length, each file included a segment starting at the beginning of one utterance and ending at the beginning of the following utterance. The beginning and end of each utterance, as well as utterance-final pauses, can therefore be accurately identified.

Next, we used PRAAT software (version 6.1; Boersma, 2001) for acoustic coding. For each utterance, three positions were first identified and marked in PRAAT. These positions were utterance-final, phrase-final, and nonfinal. Within an utterance, each syllable was first identified auditorily (i.e., marked off as a single syllable) and then categorized as utterance-final (the syllable preceding the end of an utterance), phrase-final (the syllable at the end of the first major phrase), or nonfinal (all other syllables). All syllables were then coded for vowel length, pitch range, and pause duration. Although amplitude was reported in Fisher and Tokura (1996), it did not differ as a function of utterance or phrase position. In addition, in the naturalistic settings in which these recordings were made, the presence of ambient background noise rendered the recordings too noisy to provide a reliable amplitude measurement. We therefore excluded amplitude from coding or analysis. Note that ambient background noise did not pose problems for identifying pauses or vowel lengths based on the spectrograms.

For all acoustic measures, two trained research assistants worked with a lead researcher, all of whom had a background in linguistics. One research assistant coded the measures independently first, and a second research assistant checked all coding and marked out tokens on which a differing opinion occurred. These tokens were then checked by either the lead researcher or resolved by the two research assistants after a discussion.

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

Following the rationale provided in Fisher and Tokura (1996), vowel duration (rather than syllable duration) was measured because it is most strongly related to how adult listeners perceive prosodic boundaries in English (Wightman et al., 1992). For example, "father" produces two vowel duration entries (/ɑ/ and /ɚ/). Soderstrom et al. (2008) used the same approach in their analysis of acoustic cues to syntax in IDS. Measuring vowel duration in this way has the advantage of not presupposing that infants have access to other linguistic information that they could use to perceive differences that take vowel identity and stress into account. Based on the criteria in Mack (1982), and the procedure used in Geffen et al. (under review), a trained coder marked the vowel portions by hand by examining the spectrogram and waveform and listening to the corresponding audio. Once the portion was marked, the duration as well as F0 information were automatically calculated in PRAAT.

Pitch range was calculated by subtracting the lowest fundamental frequency f (F0min) from the highest (F0max) within a syllable. Pause duration was quantified as the duration of any non-sounding interval between syllables (phrase final or nonfinal) or between sentences (utterance-final pause) that is longer than 10ms using the To TextGrid (Silences) function in PRAAT. Silences that occurred within a syllable were not considered in our analyses. The 10ms window was a first pass that allowed PRAAT to highlight all potential pauses. Fisher and Tokura (1996) reported that the means of nonfinal and phrase-final pauses (which are much shorter than utterance final pauses) were 37ms and 26ms (pp. 3201), for which the 10ms preset of PRAAT offers a low enough threshold to detect as a first pass. We then manually checked each pause to code the presence and duration of all pauses of interest. For example, segments of periods of silence may be marked within a word by PRAAT, but they were not coded as a pause. Only those appearing at the nonfinal, phrase-final, and utterance-final positions were coded as pauses of interest and categorized as such.

In total, 1465 well-formed utterances were coded across all participants, with an average of 29.94 (SD = 10.74) sentences for each participant. In contrast to Fisher and Tokura (1996), in which only the mother's language input for each participant was included (one mother was recorded at home and the other two speakers were recorded in laboratory settings), our initial sample consisted of speech that was produced by multiple speakers (n = 127), both male and female, in the infants' immediate environment. We included only female speakers (n = 97) in our final dataset to provide a clearer comparison with Fisher and Tokura's (1996) analyses and results. There were a total of 1295 well-formed utterances from female speakers.

Results

We first report the descriptive statistics for the acoustic measures observed at different utterance positions. We then examine the role of acoustics in bracketing using statistical analysis. For all acoustic measures, raw values of all well-formed utterances by female speakers were aggregated and reported below.

Descriptive Statistics

Table 1 shows the untransformed value of acoustic measures for all well-formed utterances (AU) and those with NP>0. The overall mean length of utterance (MLU) was 5.79 morphemes. Although the vowel lengths across the three positions were comparable to

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

Table 1. Acoustic Measures by Position within Utterances

	Nonfinal		Phrase final		Utterance final		Overall	
	AU	NP>0	AU	NP>0	AU	NP>0	AU	NP>0
Pause duration in ms								
Mean	90.71	93.32	65.17	53.22	1825.64	1748.56	1022.31	968.77
SD	120.29	118.50	147.99	50.43	1779.49	1707.20	1019.73	953.81
n	680	255	373	142	1233	453	2286	850
Range	(13.50–2063.33)	(13.50–1385.06)	(0.12–2106.15)	(1–400)	(4.66–6442.51)	(11.88–11840)	0.12–16442.51	1–1840
F0 range in hertz								
Mean	81.57	65.02	49.46	46.60	110.93	113.85	80.71	75.22
SD	336.89	254.19	69.08	72.15	112.36	128.07	172.01	151.57
n	1262	473	1277	471	1285	474	3824	1418
Vowel duration in ms								
Mean	101.04	97.22	100.52	96.70	226.99	220.58	143.17	138.16
SD	51.39	51.60	60.90	62.31	151.09	141.54	88.06	85.15
n	1268	474	1291	474	1294	474	3853	1422

those reported in Fisher and Tokura (1996), the pause durations observed in the present study were about 2- to 3-times larger than those reported in their study. Pitch changes in the current dataset were similar to those reported in Fisher and Tokura's study at the phrase- and utterance-final positions but were about 2 times larger at the non-final position compared to the value in their study. For both pause duration and pitch change, the current dataset also exhibits much larger standard deviations across the three positions than those reported in Fisher and Tokura (1996).

The distribution of utterance types (shown in Table 2) is very similar to that reported in previously studies (e.g., see review in Fernald & McRoberts, 1996; also data from Gleitman et al., 1984), though the current dataset contains a higher percentage of Yes/No questions relative to WH-questions when compared to the pattern reported in Fisher and Tokura (1996) (Y/N questions account for 31.66% in current study and 14% in Fisher & Tokura; WH-questions consist of 13.12% in present study and 36% in Fisher & Tokura, 1996). Overall, the utterance-type distribution of IDS utterances taken from daylong home recordings is largely similar to the distribution observed in speech produced in a more artificial and constrained lab setting.

In order to compare the values of acoustic measures across speakers, all values of each measure were standardized by speaker within each infant based on Fisher and Tokura (1996).

Figure 1 shows the mean standardized value of each acoustic measure broken down by utterance position. The patterning of values is similar to that in Fisher and Tokura (1996, see Figure 1, p. 3201). Utterances with NP>0 also exhibit a similar pattern (Fig. 2).

Acoustic Cues at Utterance Boundaries

One of our main questions was whether IDS within these naturalistic recordings contains utterance-final lengthening of pauses and vowels, as well as exaggerated pitch changes, as reported in Fisher and Tokura (1996). To test this question, for each infant, all acoustic measures were aggregated across the speakers to reflect the range of IDS heard by that infant. For some infants, multiple speakers contributed utterances, and the final aggregation was done across all utterances. In doing so, a speaker whose utterances accounted for a greater percentage of the total input is proportionally weighted. We also conducted the analyses with measures aggregated at the speaker level. Results based on values at the participant level or the speaker level yielded identical patterns of significance for all subsequent analyses reported.

To determine whether similar patterns held in our LENA recording as did in Fisher and Tokura (1996), we first ran ANOVAs using RStudio statistical software (version 1.2.5019; RStudio Team, 2019) with the Stats package (version 4.1.2). Results showed that

Table 2. Frequency of Utterance Types in Female Utterances

Full sentence utterance type	Frequency (%)
Declarative	715 (55.21%)
WH-question	170 (13.12%)
Yes-No question	410 (31.66%)
Total	1295 (100%)

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

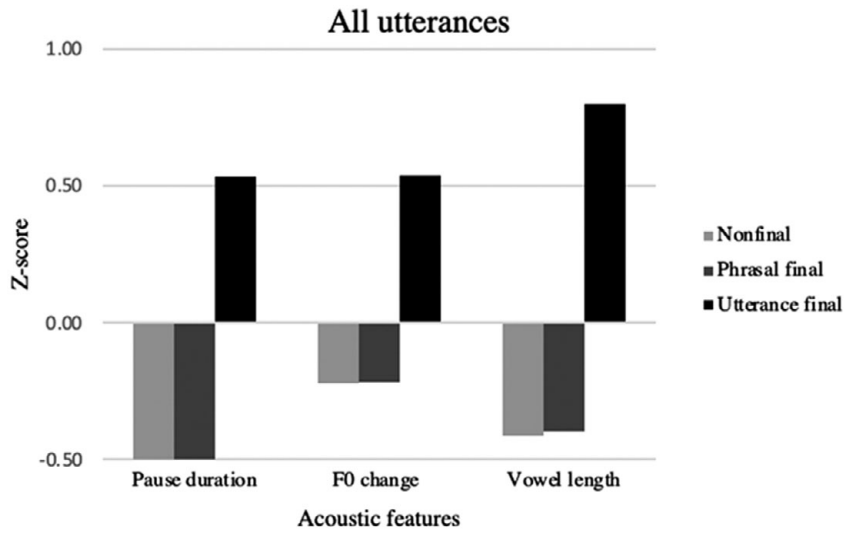

Figure 1. Mean Standardized Acoustic Measures by Position in All Utterances (Female Speakers Only).

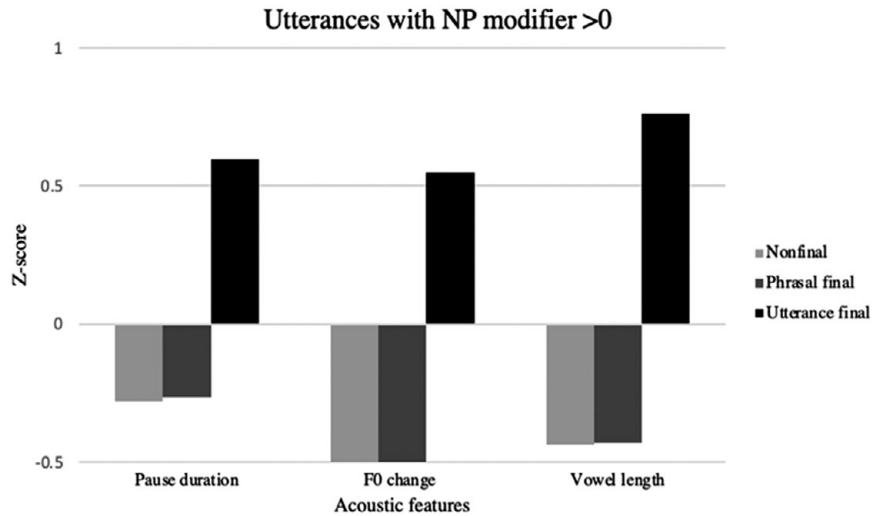

Figure 2. Mean Standardized Acoustic Measures by Position in Utterance with 1 or More NP Modifiers.

there were significant differences in all three measures as a function of position; F0 change, $F(2, 144) = 147.70$, $p < .001$, $\eta^2 = .67$; vowel length, $F(2, 144) = 1409.81$, $p < .001$, $\eta^2 = .95$; and pause duration, $F(2, 136) = 895.8$, $p < .001$, $\eta^2 = .929$.

Follow-up multiple pairwise comparisons with Bonferroni adjustments showed identical patterns across all three measures to those reported in Fisher and Tokura (1996). Table 3 summarizes the results of pairwise comparisons with Bonferroni adjustments, and Figure 3 contains a visualization of the results. As shown in Table 1, utterance-final syllables contained vowels that were more than twice as long as non-final syllables, and

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

Table 3. Pairwise Comparisons with Bonferroni Adjustment

Acoustic Measures	Group1	Group2	n1	n2	Statistic	df	*p*	*p*.adj	*p*.adj. signif
F0 Change	nonF	PF	49	49	−0.17762	48	0.86	1.00	ns
	nonF	UF	49	49	−18.5619	48	<.001	<.001	***
	PF	UF	49	49	−17.2107	48	<.001	<.001	***
Pause Duration	nonF	PF	49	49	2.487351	42	0.02	0.05	ns
	nonF	UF	49	49	−43.0801	46	<.001	<.001	***
	PF	UF	49	49	−42.7182	42	<.001	<.001	***
Vowel Length	nonF	PF	49	49	−0.59071	48	0.56	1.00	ns
	nonF	UF	49	49	−39.1313	48	<.001	<.001	***
	PF	UF	49	49	−35.4404	48	<.001	<.001	***

Note. PF = phrasal final, nonF = non-final, UF = utterance final.
***p < .001.

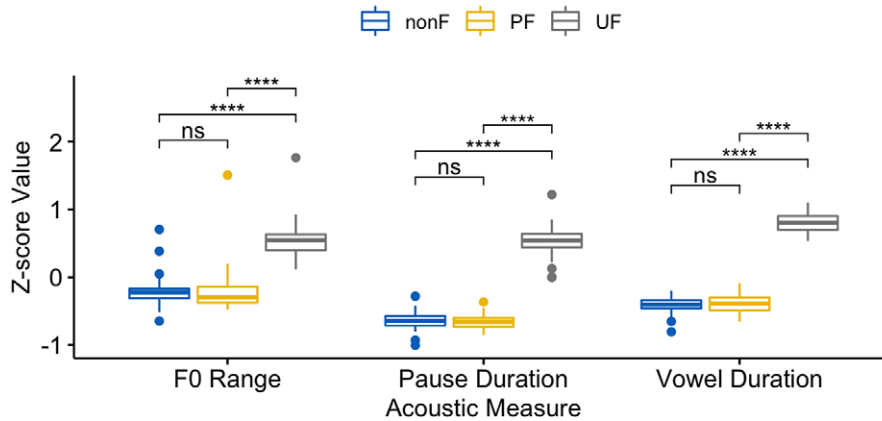

Figure 3. Pairwise Comparisons of Final and Nonfinal Values for F0 Range, Pause Duration, and Vowel Duration across All Utterances.
Note. PF = phrase-final, nonF = non-final, UF = utterance final.
**** p<.0001.

contained about 1.5 times more pitch change (226.99 ms vs. 101.04 ms, and 110.93 Hz vs. 81.57 Hz, respectively). Pauses at the utterance-final position were on average about 20 times longer than those at the phrase-final and non-final positions (1825.64 ms vs. 65.17 ms and 90.71 ms).

Acoustic Cues to Phrase Boundaries

Fisher and Tokura (1996) reported that for utterances with more than one syllable preceding the boundary between the subject and the verb of the sentence, phrase-final

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

vowels were longer than their preceding non-final vowel within the same phrase, and that this vowel lengthening was more pronounced compared to the differences between two consecutive non-final vowels. To determine whether the same was true in our data, we ran a repeated measures ANOVA with vowel duration as the repeated measure, the position of the second syllable (phrase-final vs. non-final) as a between-items variable, and vowel length as the dependent variable. Across all participants in our sample, 522 phrase-final syllables were preceded by at least one non-final syllable (36% of all utterances). Since only adjacent syllables were analyzed, raw values were used, following Fisher and Tokura's approach. An interaction between position and the repeated measures of vowel duration would indicate that the amount of acoustic difference between a syllable and the one preceding it differed depending on position.

Results showed that there was no significant main effect of position, $F(1, 136) = 1.087$, $p = .299$, $\eta^2 = .002$, no main effect of the repeated measures of vowel length, $F(1, 886) = .245$, $p = .621$, $\eta^2 = .000$, nor was there a significant interaction of the repeated measures with position, $F(1, 886) = .483$, $p = .487$, $\eta^2 = .000$.

Although Fisher and Tokura (1996) found a significant main effect of the repeated duration measures, with 76% of their phrase-final syllables being longer than the preceding non-final syllable, the rate was 57% in our sample. Phrase-final syllables were not reliably longer than their preceding syllable. In addition, they reported a significant interaction showing that the difference in the duration of phrase-final vowels and their preceding non-final vowels (59ms) was larger than that between non-final vowels and their preceding non-final vowels (3ms). However, there were no such differences in our sample: vowels in phrase-final syllables were of similar length (M = 97.34 ms, SD = 66.31) to those in the adjacent preceding non-final syllables (M = 98.84 ms, SD = 97.01). Nonfinal syllables (M = 108.47 ms, SD = 70.38) were on average 5 ms longer than their adjacent preceding non-final syllables (M = 103.66 ms, SD = 65.81). Thus, the phrase-final position in multisyllabic subject phrases was not marked by longer vowels compared to the non-final position.

Taken together, these patterns are partially consistent with the results of Fisher and Tokura (1996). When considering all well-formed utterances, vowels at the utterance-final position consistently showed larger pitch changes and longer lengths than those in non-final position or phrase-final positions. Results from utterances with multisyllabic subject phrases did not match those reported in Fisher and Tokura. They reported vowel lengthening at the phrase-final position compared to the non-final position – however we did not find that vowels were longer at the phrase-final boundary in infant-directed speech that was produced in naturalistic environments.

Discussion

Several studies suggest that IDS contains acoustic cues that could facilitate perceptually grouping speech into units that roughly correspond to syntactic units such as clauses or phrases. For example, syllables in well-formed utterance and phrase-final position contain exaggerated vowel lengthening and pitch changes, and are followed by longer pauses, relative to syllables in non-final positions (Fisher & Tokura, 1996; Ludusan et al., 2016; Soderstrom et al., 2008). However, this evidence largely comes from analyses of IDS produced in relatively constrained laboratory settings, or from a small number of speakers, and thus it is unclear whether these cues are present in IDS that is more representative of the speech infants hear in their everyday lives. Thus, we investigated

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

whether these acoustic cues to syntactic structure in English are also present in IDS produced in more naturalistic settings. Specifically, we used LENA recordings to capture at-home language input to 49 infants. Like Fisher and Tokura, we found that syllables in the final positions of well-formed utterances were characterized by greater vowel lengthening and pitch changes, and were followed by longer pauses, in comparison to syllables that did not occur in utterance final position. Although Fisher and Tokura (1996) found that utterances with multisyllabic subjects had longer vowels at the phrase-final position compared to the non-final position, we found that vowel length at these two positions did not differ significantly.

Our results suggest that the ends of syntactically well-formed utterances are acoustically distinctive in speech that caregivers use in everyday interactions with their infants, exhibiting more pitch change and vowel lengthening, as well as long pauses, in comparison to other parts of the utterance. These acoustic characteristics may function as important cues for language learning. Most obviously, these features could facilitate segmenting speech into discrete utterances, which tend to contain short, simple, syntactically-well-formed units (e.g., Kaye, 1980). Indeed, several studies suggest that these acoustic cues contribute to perceiving units within ongoing speech, or to segmenting the speech stream into prosodically, and often syntactically, well-formed units. Infants under a year of age often prefer to listen to prosodic units in which edges are marked by pauses, pitch changes, and vowel lengthening (Hirsh-Pasek et al., 1987; Jusczyk et al., 1992; but see Seidl & Cristià, 2008, for evidence for a novelty preference at 4 months of age).

Not only do infants detect these acoustic features that occur at a boundary or edge, and prefer listening to speech sequences bracketed by such cues, but infants also appear to have better memory for specific sequences of words within vs. across these boundaries (e.g., Nazzi et al., 2000; Seidl, 2007; Soderstrom et al., 2003). For example, infants are better able to recognize a sequence of words they were just familiarized with (i.e., *rabbits eat leafy vegetables*) when it is a well-formed clause "…rabbits eat leafy vegetables." than when the sequence spans a clause boundary "…rabbits eat. Leafy vegetables…" (Seidl, 2007). Thus, bracketing cues potentially facilitate encoding and remembering well-formed speech sequences, including their phonetic, phonotactic, prosodic, and lexical content. By enhancing memory for the contents of speech, bracketing could also facilitate distributional analysis, or tracking co-occurrence relationships among words within these sequences (e.g., Gerken, 1996).

The presence of all three cues at utterance boundaries – syllables with more exaggerated vowel-lengthening and pitch changes, and which are followed by a pause – could also facilitate learning language-specific cue weightings. For example, 4-month-old English-learning infants appear to require all 3 cues to bracket speech (Seidl & Cristià, 2008). However, by 6 months English-learning infants appear to rely most strongly on pitch changes, which tend to be especially strong in English, as long as they are accompanied by either a pause or a duration change (Seidl, 2007). By 6-8 months, German-learning infants can use a combination of pitch change and lengthening, even in the absence of pauses, which tend to be unreliable in German (Wellmann et al., 2012). In contrast, Dutch-learning infants rely more heavily on pauses than pitch or duration, also likely reflecting the extent to which the cues are present and reliable in infants' native language (Johnson & Seidl, 2008).

In addition, experience with strong cues at utterance boundaries corresponding to clauses could help infants detect these cues at phrase boundaries, which tend to be weaker (Cooper & Paccia-Cooper, 1980). Indeed, we failed to find evidence of these cues at phrase

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

boundaries in IDS to 12-month-olds. In lab studies, infants tend to show evidence of using these prosodic cues to group and recognize clauses by 6 months but use them to group sequences into phrases only by 9 months (e.g., Jusczyk et al., 1992). However, Soderstrom et al. (2003) found that when phrases are marked by strong prosodic cues, even 6-month-olds appear to use them to segment and remember the speech stream. We suggest that the very strong cues at the end of well-formed utterances (i.e., clause boundaries) could be used to bootstrap sensitivity to smaller and more subtly marked units, when these cues do occur.

The acoustic exaggerations we observed in everyday IDS could also impact word segmentation, recognition, and learning. Vowel space expansion (including vowel lengthening) and pitch changes characteristic of IDS can facilitate developments in infants' speech perception, and specifically perceptual narrowing that leads to more accurate phoneme perception (Liu et al., 2003). Thus, the presence of longer vowels with larger pitch changes at utterance boundaries may facilitate infants' encoding and discrimination of speech at the end of an utterance. Because the end of an utterance is also nearly always the end of a word, these acoustic features could facilitate processing speech that occurs at the end of an utterance. Indeed, infants tend to segment (Seidl & Johnson, 2006) and recognize words better when they occur in utterance-final position (Fernald et al., 2001).

As mentioned, one point of departure from the results in Fisher and Tokura (1996) concerns the presence of acoustic cues to phrase boundaries. In their study, vowel lengthening regularly occurred at the phrase boundaries in utterances that contained multisyllabic noun-phrase subjects. The authors concluded that this phrase-final lengthening, though less robust than the effects at the utterance-final position, allowed infants to detect the general prosodic shape of English utterances. In the current study, however, we did not find differences in vowel length between the phrase-final position and non-final position when considering utterances with multisyllabic subject phrases. These results are consistent with evidence that the prosodic correlates of phrase boundaries are strongest at higher-order syntactic nodes, and may be weak or absent altogether in the short simple constructions characteristic of IDS (Cooper & Paccia-Cooper, 1980). Nonetheless, these cues may be present in more complex speech, such as in speech to older children and adults, which contains longer, more syntactically complex utterances. As discussed above, becoming attuned to the robust cues present at utterance boundaries could help infants and children to detect them at phrase boundaries when they do occur.

It is also possible that the lack of phrase-level bracketing in IDS sampled from daylong recordings in the home is driven by being used in different types of interactions than those within which lab-based IDS is typically produced. Previous studies have shown that caregivers' sensitivity to environment or context can be reflected in their speech, and that they alter their intonation, word use, and sentence structure to reflect the various goals of speech (Rondal, 1980). In a lab-based session, caregivers are primarily motivated to engage their infants, and the predominant acoustic function of their IDS may therefore be attracting and maintaining the infant's attention and engagement. In contrast, input captured from real-world interactions may consist more of urging children to do something or preventing them from engaging in certain behavior, and therefore differ from speech used in lab sessions at a functional level. It is plausible that speech at home and speech in the lab have different acoustic properties as a result of different communicative functions. Relatedly, even though the distribution of utterance types in the current study is similar to that reported in Fisher and Tokura (1996), the specific utterances produced in the two contexts may have served different functional goals and therefore yielded different acoustic patterns.

https://doi.org/10.1017/S030500092300034X Published online by Cambridge University Press

In sum, this study provides an initial step towards characterizing the IDS that infants hear in their everyday lives, and the extent to which it may contain cues relevant for learning syntactic structure. In future studies it will be important to extend this investigation to languages other than American English. Nonetheless, our initial results suggest that the IDS that infants hear in everyday interactions contains cues relevant to bracketing, and that these cues are not identical to those captured in lab recordings, thus underlining the importance of collecting naturalistic data. Our results indicate that there are exaggerations in vowel pitch and duration, and in pause length at the utterance-final position. Naturalistically produced IDS did not contain similar cues at utterance-internal syntactic boundaries, perhaps due to different pragmatic and functional use. Importantly, our replication demonstrates that well-formed utterance boundaries are not only marked by prosodic cues in short interactions in a lab setting, in which caregivers are primarily motivated to engage their infants. The present study provides a foundation for future studies aimed at obtaining a better understanding of speech to infants, and contributes to our knowledge about IDS captured in an unobtrusive long-form recording method (e.g., Bergelson et al., 2018; Ramírez-Esparza et al., 2014, 2017; Weisleder & Fernald, 2013).

Acknowledgements. We thank the members of the Infant Studies Lab at the University of Notre Dame, especially Jayde Homer and Dr. Abbie Thompson, for their contributions to this work. We are also grateful to the families whose participation made this research possible. This research was supported by funds from NSF BCS-1352443.

References

Adriaans, F., & **Swingley, D.** (2012). Distributional learning of vowel categories is supported by prosody in infant-directed speech. *Proceedings of the Annual Meeting of the Cognitive Science Society*, **34**, 72–77. https://escholarship.org/uc/item/7pt904fz

Bergelson, E., **Casillas, M.**, **Soderstrom, M.**, **Seidl, A.**, **Warlaumont, A. S.**, & **Amatuni, A.** (2018). What do North American babies hear? A large-scale cross-corpus analysis. *Developmental Science*, **22**(1), e12724. https://doi.org/10.1111/desc.12724

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*. **5**, 341–345.

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

Cho, Y. Y. (2016). Korean phonetics and phonology. In M. Aronoff (Ed.). *Oxford research encyclopedia of linguistics* (published online). Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.176

Cooper, R. P., & **Aslin, R. N.** (1990). Preference for infant-directed speech in the first month after birth. *Child Development*, **61**(5), 1584–1595. https://doi.org/10.1111/j.1467-8624.1990.tb02885.x

Cooper, R. P., **Abraham, J.**, **Berman, S.**, & **Staska, M.** (1997). The development of infant's preference for motherese. *Infant Behavior and Development*, **20**(4), 477–488. https://doi.org/10.2307/1130766

Cooper, W. E., & **Paccia-Cooper, J.** (1980). *Syntax and speech* (No. 3). Harvard University Press.

Cristià, A. (2008). The impact of caregivers' speech on infants' discrimination of a speech sound contrast. *The Journal of the Acoustical Society of America*, **124**(4), 2437–2437. https://doi.org/10.1121/1.4782531

Cristià, A. (2013). Input to language: the phonetics and perception of infant-directed speech. *Language and Linguistics Compass*, **7**(3), 157–170. https://doi.org/10.1111/lnc3.12015

Cristià, A., & **Seidl, A.** (2014). The hyperarticulation hypothesis of infant-directed speech. *Journal of Child Language*, **41**(4), 913–934. https://doi.org/10.1017/S0305000912000669

Fernald, A. (1985). Four-month-old infants prefer to listen to motherese. *Infant Behavior and Development*, **8**, 181–195. https://doi.org/10.1016/S0163-6383(85)80005-9

Fernald, A. (1992). Meaningful melodies in mothers' speech to infants. In H. Papoušek & U. Jürgensand M. Papoušek (Eds.), *Nonverbal vocal communication: Comparative and developmental approaches* (pp.262–282). Cambridge University Press.

Fernald, A., & **Mazzie, C.** (1991). Prosody and focus in speech to infants and adults. *Developmental Psychology*, **27**(2), 209. https://doi.org/10.1037/0012-1649.27.2.209

Fernald, A., & **McRoberts, G.** (1996). Prosodic bootstrapping: A critical analysis of the argument and the evidence. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (pp. 365–388). Lawrence Erlbaum Associates.

Fernald, A., **McRoberts, G. W.**, & **Swingley, D.** (2001). Infants' developing competence in recognizing and understanding words in fluent speech. In J. Weissenborn & B. hOBLE (Eds.), *Approaches to bootstrapping: Phonological, lexical, syntactic and neurophysiological aspects of early language acquisition*. Volume **1** (pp. 97–123). John Benjamins.

Fernald, A., **Taeschner, T.**, **Dunn, J.**, **Papousek, M.**, **de Boysson-Bardies, B.**, & **Fukui, I.** (1989). A cross-language study of prosodic modifications in mothers' and fathers' speech to preverbal infants. *Journal of Child Language*, **16**(3), 477–501. https://doi.org/10.1017/S0305000900010679

Fisher, C., & **Tokura, H.** (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development*, **67**(6), 3192–3218. https://doi.org/10.1111/j.1467-8624.1996.tb01909.x

Gee, J., & **Grosjean, F.** (1983). Performance structures: A psycholinguistic and linguistic ap- praisal. *Cognitive Psychology*, **15**, 411–458. https://doi.org/10.1016/0010-0285(83)90014-2

Geffen, S., **Burkinshaw, K.**, **Athanasopoulou, A.**, & **Curtin, S.** (under review). Utterance-initial prosodic differences between statements and questions in infant-directed speech.

Gerken, L. A. (1996). Phonological and distributional information in syntax acquisition. In J. Morgan & C. Demuth (Eds.), *Signal to syntax* (pp. 411–425). Erlbaum.

Gerken, L. A., **Jusczyk, P. W.**, & **Mandel, D.** (1994). When prosody fails to cue syntactic structure: 9-month-olds' sensitivity to phono- logical vs. syntactic phrases. *Cognition*, **51**, 237–265. https://doi.org/10.1016/0010-0277(94)90055-8

Gilkerson, J., & **Richards, J. A.** (2009). *Impact of adult talk, conversational turns and TV during the critical 0-4 years of child development* (LENA Technical Report LTR-01-2). LENA Foundation.

Gleitman, L. R., **Newport, E. L.**, & **Gleitman, H.** (1984). The current status of the motherese hypothesis. *Journal of Child Language*, **11**(1), 43–79. https://doi.org/10.1017/S0305000900005584

Gleitman, L. R., & **Wanner, E.** (1982). The state of the state of the art. In E. Wanner & L.R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp.3–48). Cambridge University Press.

Goldman-Eisler, F. (1972). Pauses, clauses, sentences. *Language And Speech*, **15**, 103–113. https://doi.org/10.1177/002383097201500201

Graf Estes, K., & **Hurley, K.** (2013). Infant-directed prosody helps infants map sounds to meanings. *Infancy*, **18**(5), 797–824. https://doi.org/10.1111/infa.12006

Hilton, C. B., **Moser, C. J.**, **Bertolo, M.**, **Lee-Rubin, H.**, **Amir, D.**, **Bainbridge, C. M.**, **Simson, J.**, **Knox, D.**, **Glowacki, L.**, **Alemu, E.**, **Galbarczyk, A.**, **Jasienska, G.**, **Ross, C. T.**, **Neff, M. B.**, **Martin, A.**, **Cirelli, L. K.**, **Trehub, S. E.**, **Song, J.**, **Kim, M.**, **Schachner, A.**, … **Mehr, S. A.** (2022). Acoustic regularities in infant-directed speech and song across cultures. *Nature Human Behavior*, **6**(11), 1545–1556. https://doi.org/10.1038/s41562-022-01410-x

Hirsh Pasek, K., **Kemler Nelson, D. G.**, **Jusczyk, P. W.**, **Cassidy, K. W.**, **Druss, B.**, & **Kennedy, L.** (1987). Clauses are perceptual units for young infants. *Cognition*, **26**(3), 269–286. https://doi.org/10.1016/S0010-0277(87)80002-1

Honda, K. (2004). Physiological factors causing tonal characteristics of speech: from global to local prosody. In *Speech Prosody 2004, International Conference.*

Johnson, E. K., & **Seidl, A.** (2008). Clause segmentation by 6-month-old infants: A crosslinguistic perspective. *Infancy*, **13**(5), 440–455. https://doi.org/10.1080/15250000802329321

Jusczyk, P., **Hirsh-Pasek, K.**, **Kemler-Nelson, D.**, **Kennedy, L.**, **Woodward, A.**, & **Piwoz, J.** (1992). Perception of acoustic correlates of major phrasal units by young infants. *Cognitive Psychology* **24**, 252–293. https://doi.org/10.1016/0010-0285(92)90009-Q

Kaye, K. (1980). Why we don't talk 'baby talk' to babies. *Journal of Child Language*, **7**, 489–507. https://doi.org/10.1017/S0305000900002804

Kemler Nelson, D. G., **Hirsh-Pasek, K.**, **Jusczyk, P. W.**, & **Cassidy, K. W.** (1989). How the prosodic cues in motherese might assist language learning. *Journal of Child Language*, **16**(1), 55–68. https://doi.org/10.1017/S030500090001343X

Kuhl, P. K., **Andruski, J. E.**, **Chistovich, I. A.**, **Chistovich, L. A.**, **Kozhevnikova, E. V.**, **Ryskina, V. L.**, **Stolyarova, E. I.**, **Sundberg, U.**, & **Lacerda, F.** (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science*, **277**, 684–6. https://doi.org/10.1126/science.277.5326.684

Lany, J., & **Shoaib, A.** (2020). Individual differences in non-adjacent statistical dependency learning in infants. *Journal of Child Language*, **47**(2), 483–507. https://doi.org/10.1017/S0305000919000230

LENA Foundation. (2018). *LENA Foundation.* https://www.lena.org/

Liu, H. M., **Kuhl, P. K.**, & **Tsao, F. M.** (2003). An association between mother's speech clarity and infants' speech discrimination skills. *Developmental Science*, **6**(3), F1–F10. https://doi.org/10.1111/1467-7687.00275

Ludusan, B., **Cristià, A.**, **Martin, A.**, **Mazuka, R.**, & **Dupoux, E.** (2016). Learnability of prosodic boundaries: Is infant-directed speech easier? *The Journal of the Acoustical Society of America*, **140**(2), 1239–1250. https://doi.org/10.1121/1.4960576

Ma, W., **Golinkoff, R. M.**, **Houston, D. M.**, & **Hirsh-Pasek, K.** (2011). Word learning in infant-and adult-directed speech. *Language Learning and Development*, **7**(3), 185–201. https://doi.org/10.1080/15475441.2011.579839

Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual production. *The Journal of the Acoustical Society of America*, **71**, 173. https://doi.org/10.1121/1.387344

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk: Volume 1: Transcription format and programs.* (3rd). Lawrence Erlbaum.

Many Babies Consortium. (2020). Quantifying sources of variability in infancy research using the infant-directed speech preference. *Advances in Methods and Practices in Psychological Science*, **3**, 24–52. https://doi.org/10.1177/2515245919900809

Martin, A., **Schatz, T.**, **Versteegh, M.**, **Miyazawa, K.**, **Mazuka, R.**, **Dupoux, E.**, & **Cristià, A.** (2015). Mothers speak less clearly to infants than to adults. *Psychological Science*, **26**(3), 341–347. https://doi.org/10.1177/0956797614562453

Nazzi, T., **Nelson, D. G. K.**, **Jusczyk, P. W.**, & **Jusczyk, A. M.** (2000). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic well formedness. *Infancy*, **1**(1), 123–147. https://doi.org/10.1207/S15327078IN0101_11.

Nencheva, M. L., **Piazza, E. A.**, & **Lew-Williams, C.** (2021). The moment-to-moment pitch dynamics of child-directed speech shape toddlers' attention and learning. *Developmental Science*, **24**(1), e12997. https://doi.org/10.1111/desc.12997

Nespor, M., & **Vogel, I.** (1986). *Prosodic phonology*. Foris.

Paschen, L., **Fuchs, S.**, & **Seifart, F.** (2022). Final lengthening and vowel length in 25 languages. *Journal of Phonetics*, **94**(9), 101179. https://doi.org/10.1016/j.wocn.2022.101179

Pegg, J. E., **Werker, J. F.**, & **McLeod, P. J.** (1992). Preference for infant-directed over adult-directed speech: Evidence from 7-week-old infants. *Infant Behavior and Development*, **15**(3), 325–345. https://doi.org/10.1016/0163-6383(92)80003-D

Pena, M., **Maki, A.**, **Kovacčić, D.**, **Dehaene-Lambertz, G.**, **Koizumi, H.**, **Bouquet, F.**, & **Mehler, J.** (2003). Sounds and silence: An optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences of the United States of America*, **100**(20), 11702–11705. https://doi.org/10.1073/pnas.1934290100

Pouw, W., **Paxon, A.**, **Harrison, S. J.**, & **Dixon, J. A.** (2020). Acoustic information about upper limb movement in voicing. *Psychological and Cognitive Sciences*, **117**(21), 11364–11367. https://doi.org/10.1073/pnas.2004163117

Ramírez-Esparza, N., **García-Sierra, A.**, & **Kuhl, P. K.** (2014). Look who's talking: Speech style and social context in language input to infants are linked to concurrent and future speech development. *Developmental Science*, **17**(6), 880–891. https://doi.org/10.1111/desc.12172

Ramírez-Esparza, N., **García-Sierra, A.**, & **Kuhl, P. K.** (2017). Look who's talking NOW! Parentese speech, social context, and language development across time. *Frontier in Psychology*, **8**:1008. https://doi.org/10.3389/fpsyg.2017.01008

Read, C., & **Schreiber, P. A.** (1982). Why short subjects are harder to find than long ones. In E. Wanner & L. R. Gleitman (Eds.), *Language acquisition: The state of the art* (pp.78–101). Cambridge University Press.

Rondal, J. A. (1980). Fathers' and mothers' speech in early language development. *Journal of Child Language*, **7**(2), 353–369. https://doi.org/10.1017/S0305000900002671

Rstudio Team. (2019). *Rstudio: Integrated Development Environment for R.* Boston, MA. http://www.rstudio.com/

Seidl, A., & **Johnson, E. K.** (2006). Infant word segmentation revisited: Edge alignment facilitates target extraction. *Developmental science*, **9**(6), 565–573. https://doi.org/10.1111/j.1467-7687.2006.00534.x

Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, **57**(1), 24–48. https://doi.org/10.1016/j.jml.2006.10.004

Seidl, A., & **Cristià, A.** (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, **11**(4), 596–606. https://doi.org/10.1111/j.1467-7687.2008.00704.x

Selkirk, E. O. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry*, **11**(3), 563–605. http://www.jstor.org/stable/4178179

Singh, L., **Morgan, J. L.**, & **Best, C. T.** (2002). Infants' listening preferences: Baby talk or happy talk? *Infancy*, **3**(3), 365–394. https://doi.org/10.1207/S15327078IN0303_5

Soderstrom, M., **Blossom, M.**, **Foygel, R.**, & **Morgan, J. L.** (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language*, **35**(4), 869–902. https://doi.org/10.1017/S0305000908008763

Soderstrom, M., **Seidl, A.**, **Nelson, D. G. K.**, & **Jusczyk, P. W.** (2003). The prosodic bootstrapping of phrases: Evidence from prelinguistic infants. *Journal of Memory and Language*, **49**(2), 249–267. https://doi.org/10.1016/S0749-596X(03)00024-X

Stern, D., **Spieker, S.**, **Barnett, R.**, & **MacKain, K.** (1983). The prosody of maternal speech: Infant age and context related changes. *Journal of Child Language*, **10**(1), 1–15. https://doi.org/10.1017/S0305000900005092

Tamis-LeMonda, C. S., **Kuchirko, Y.**, **Luo, R.**, **Escobar, K.**, & **Bornstein, M. H.** (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental Science*, **20**(6), e12456. https://doi.org/10.1111/desc.12456

Thiessen, E. D., **Hill, E. A.**, & **Saffran, J. R.** (2005). Infant-directed speech facilitates word segmentation. *Infancy*, **7**(1), 53–71. https://doi.org/10.1207/s15327078in0701_5

Thompson, A. (2019). *Who's Talking to Whom and Does It Matter? The Impact of Multiple Speakers, Overheard Speech, and Child-directed Speech on Infants' Language Development* (Publication No. 27930178) [Doctoral dissertation, University of Notre Dame]. ProQuest Dissertations Publishing.

Weisleder, A., & **Fernald, A.** (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, **24**(11), 2143–2152. https://doi.org/10.1177/0956797613488145

Wellmann, C., **Holzgrefe, J.**, **Truckenbrodt, H.**, **Wartenburger, I.**, & **Hohle, B.** (2012). How each prosodic boundary cue matters: Evidence from German infants. *Frontiers in Psychology*, **3**, 580. https://doi.org/10.3389/fpsyg.2012.00580

Wightman, C. W., **Shattuck-Hufnagel, S.**, **Ostendorf, M.**, & **Price, P. J.** (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *The Journal of the Acoustical Society of America*, **91**(3), 1707–1717. https://doi.org/10.1121/1.402450

Cite this article: Wang T., Yu E.C., Huang R., & Lany J. (2024). Acoustic cues to phrase and clause boundaries in infant-directed speech: Evidence from LENA recordings. *Journal of Child Language* **51,** 1193–1212, https://doi.org/10.1017/S030500092300034X