Prosodic Cues in Infant-Directed Speech Facilitate Young Children's Conversational Turn Predictions

Marina Kalashnikova[1,2] and Heather Kember[2]

[1]Basque Center on Cognition Brain and Language, San Sebastian, Spain

[2]MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith, Australia

Corresponding author:

Marina Kalashnikova, BCBL. Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, San Sebastian, Guipuzcoa, 20009, Spain; t: +34 943 309 300 (ext: 234), m.kalashnikova@bcbl.eu

Acknowledgements:

This research was supported by an Australian Linguistics Society Research Grant "Come on kids, pay attention to your prosody!" to the second and first authors. The first author receives support from the European Union's Horizon 2020 Marie Sklodowska-Curie individual Fellowships European Programme under Grant Agreement No 798908 Optimising IDS. We thank Caitlin Hooper and Ruth Brookman for their assistance with the preparation or the stimuli and data collection. We are grateful to all the children and parents for their valuable time and interest in this research.

[This article has been accepted for publication in Journal of Experimental Child Psychology 24-05-2020]

Abstract

Experienced language users are able to predict when conversational turns approach completion, which allows them to attend to and comprehend their interlocutor's speech while planning and accurately timing their response. Adults primarily rely on lexico-syntactic cues to make such predictions, but it remains unknown what cues support these predictions in young children whose lexico-syntactic competence is still developing. This study assessed children's reliance on prosodic cues, specifically when predicting conversational turn transitions in infant-directed speech (IDS), the speech register that they encounter in day-to-day interactions, and that is characterized by exaggerated prosody compared to adult-directed speech (ADS). One- and three-year-olds completed an anticipatory looking paradigm in which their gaze patterns were recorded while they observed conversations that were produced in IDS or ADS and that contained prosodically complete (lexico-syntactic and prosodic cues) and prosodically incomplete (only lexico-syntactic cues) utterances. One-year-olds anticipated more turns that were signalled by prosodic cues (i.e., prosodically complete utterances) only in IDS, while three-year-olds did so in IDS and ADS. These findings indicate that children anticipate the completion of conversational turns by relying on prosodic information in speech, and the prosodic exaggeration of IDS supports this ability while children's linguistic and conversational skills are still developing.

Keywords: Conversation; turn-taking; infant-directed speech; prosody; anticipation; eye-tracking

Language is learned and processed in the context of face-to-face conversations. Speakers engage in conversations spontaneously and naturally, and they are unconscious of the highly complex linguistic processing skills that these exchanges require (Levinson, 2016). One of the main challenges in this process consists in timing conversational turns, which allows speakers to listen to their interlocutors and produce timely and coherent responses (Bögels & Levinson, 2017; Levinson & Torreira, 2015). Adults accurately time their responses in conversations, but this task represents a challenge for young children whose language processing abilities are still developing (Casillas, Bobb, & Clark, 2016), and who are in the process of acquiring their native language. This study investigated the role of infant-directed speech in the early development of turn-taking abilities. We propose that adults' use of this register provides infants with exaggerated prosodic cues that signal conversational turn transitions and fosters their early abilities to use and process language in the context of communicative interactions.

Conversational turns not only establish the structure for communicative exchanges (Schegloff, Sacks, & Jefferson, 1974), but they also reveal speakers' remarkable language-processing capacities (Levinson, 2016). That is, adults are typically able to take turns that rarely overlap, and that are separated by approximately 200 msec inter-turn silences (Stivers et al., 2009). This brief silence duration is astounding given that preparation for the production of individual words and utterances requires at least 600 msec (Bates et al., 2003; Indefrey & Levelt, 2004), so this implies that adults plan their own utterance before their interlocutor's turn is completed. That is, in order for conversations to follow a natural flow without unusually long silences or interruptions, speakers must predict the completion point of their interlocutor's turn and start planning their own turn before this point is reached (Bögels & Levinson, 2017; Levinson & Torreira, 2015). To successfully predict the completion point of a turn and time the onset of their response, experienced language users rely on their sensitivity to the rich visual (e.g., gaze, gestures), pragmatic (e.g., understanding that certain speech acts like questions require immediate responses), prosodic (e.g., final phrase lengthening, pitch modulations), and lexico-syntactic (e.g., content of the utterance, syntactic clause completion) cues in speech that sign post the structure of a conversation (see Clayman, 2013; Walker, 2013; Levinson & Torreira, 2015 for comprehensive reviews). However, conversational partners do not always have full access to a combination of these cues, which can result in reduced prediction accuracy. For instance, when listening to utterances that are prosodically but not syntactically complete, adults are more likely to perceive false completion points, and they are significantly slower at identifying existing completion points in syntactically complete but prosodically incomplete utterances (Bögels & Torreira, 2015). In these challenging cases, access to lexico-syntactic information is key as the ability to comprehend the content of the incoming turn allows adults to continue making predictions in the absence of other cues such as prosody (De Ruiter, Mitterer, & Enfield, 2006), and it strengthens their turn projection capacity allowing them to not only predict the number of remaining words in a turn but also their duration and content (Magyari & de Ruiter, 2012; Magyari, Bastiaansen, de Ruiter, & Levinson, 2014; Magyari, De Ruiter, & Levinson, 2017).

This poses the question of whether full access to lexico-syntactic information in speech is a pre-requisite for the development of mature turn-taking abilities. Infants start producing turn-taking behaviors already in their first months of life in proto-conversations with their caregivers (Bateson, 1975; Snow, 1977), but it has been suggested that the ability to time and predict turn transitions continues to develop well into childhood (Casillas et al., 2016). In fact, inter-turn silences increase when infants approach their first birthday (Hilbrink, Gattis, & Levinson, 2015), which is proposed to coincide with an increase in processing demands associated with the integration of language comprehension and production in their communicative interactions, and they remain longer than is expected for adults until after the age of five years (Casillas et al., 2016). However, despite producing longer inter-turn silences for several years, there is no evidence that children wait for a silence to occur before starting to plan their response (Lindsay, Gambi, & Rabagliati, 2019; Smith & McMurray, 2018), which suggests that they do engage in predicting turn completion points early on, and that these predictions may rely on cues that are available to children before they gain full access to the lexical and syntactic content of their native language. Specifically, prosodic skills develop in the first months of life (Nazzi, Bertoncini, & Mehler, 1998; Nazzi & Ramus, 2003), and they are manifested in infants' early sensitivity to the prosodic properties of the words in their native language (Höhle, Bijeljac-Babic, Herold, Weissenborn, & Nazzi, 2009; Jusczyk, Cutler, & Redanz, 1993; Jusczyk & Luce, 1994; Mattys, Jusczyk, Luce, & Morgan, 1999), but also the ability to identify prosodic cues to clause completion including final vowel lengthening, pitch modulations, and intonational patterns between six and eight months of age (Frota, Butler, & Vigário, 2014; Johnson & Seidl, 2008; Seidl, 2007; Soderstrom, Blossom, Foygel, & Morgan, 2008; Soderstrom, Seidl, Kemler Nelson, & Jusczyk, 2003).

Interestingly, existing experimental evidence demonstrates that while children can predict the timing of conversational turn transitions, they do not weigh prosodic cues to turn completion more heavily than adults or earlier than lexico-syntactic cues in this process (Casillas & Frank, 2017; Keitel & Daum, 2015; Keitel, Prinz, Friederici, Hofsten, & Daum, 2013; Lammertink, Casillas, Benders, Post, & Fikkert, 2015). This conclusion is based on studies that employed anticipatory looking paradigms in which participants observe video recordings of conversations and their gaze shifts from the speaker to the listener are recorded. This technique detects whether the gaze shift happens before or after the inter-turn silence, thus differentiating when participants predict vs. react to a turn transition. Using this method, Keitel and colleagues presented six-, 12-, 24-, and 36-month-old children and adults with conversations including natural adult-directed speech and flattened-intonation speech (Keitel & Daum, 2015; Keitel et al., 2013). Reliable turn prediction in natural speech was only detected in three-year-olds and adults, and only adults also predicted turns in the flattened pitch condition. Lammertink et al. (2015) presented 30-month-olds and adults with conversations recorded in infant-directed speech, which was manipulated to contain only prosodic or lexico-syntactic cues to turn completion. Children and adults predicted turn transitions in cases when lexico-syntax cued turn completion but prosody cued turn continuation, but only adults and not children made predictions in the reverse cases where only prosody cued turn completion, but lexico-syntax did not. These findings indicate that turn prediction abilities are still developing around the age of three years, and children rely on both the lexico-syntactic and prosodic cues in speech to predict when a conversational turn will reach completion.

A later study by Casillas and Frank (2017) demonstrated that turn prediction based solely on prosodic cues becomes evident only after the age of five years. In this study, children from one to six years and adults completed an anticipatory looking paradigm with conversations presented in four conditions: natural infant-directed speech, low-pass filtered speech (prosody only), flattened speech (lexico-syntactic cues only), and noise (no speech). Children and adults were able to predict turns in the natural speech but not the noise condition, but this was only the case for children after two years of age. Five- and six-year-olds as well as adults also anticipated turns in the prosody-only condition, but only adults did so when presented with lexico-syntactic cues alone without prosody information in speech. This study aligns with previous findings that after their second birthday, children are most successful in anticipating conversational turn transitions when presented with a full array of cues in the speech signal, but it also reveals that unlike adults, they are unable to extract the necessary information to predict a turn from lexico-syntactic content alone.

A noteworthy observation is that the four studies reviewed above used different speech registers to create their auditory stimuli. Keitel and colleagues (Keitel & Daum, 2015; Keitel et al., 2013) used stimuli recorded in adult-directed speech, while Lammertink et al. (Lammertink et al., 2015) and Casillas and Frank’s (Casillas & Frank, 2017) stimuli were recorded in infant-directed speech, and the latter were the only two studies to report that children could predict conversational turns by 30 months of age. Therefore, it remains plausible that children are able to rely on prosodic cues to support their early predictions about conversational turn structure, but only when those cues are representative of infant-directed speech, the type of speech that they encounter in their day-to-day communicative interactions.

In comparison with adult-directed speech (ADS), infant-directed speech (IDS) is prosodically and acoustically exaggerated (see Cristia, 2013; Golinkoff, Can, Soderstrom, & Hirsh-Pasek, 2015; Soderstrom, 2007 for reviews), and is manifested in slow rate (Panneton, Kitamura, Mattock, & Burnham, 2006), exaggerated pitch height and range (Fernald et al., 1989; Fernald & Kuhl, 1987; Fernald & Simon, 1984), positive vocal affect (Fernald, 1993; Kitamura & Burnham, 2003; Singh, Morgan, & Best, 2002), and acoustically exaggerated vowels (Burnham, Kitamura, & Vollmer-Conna, 2002; Kalashnikova, Carignan, & Burnham, 2017; Kuhl et al., 1997). From their first days of life, infants prefer listening to IDS over ADS (Dunst, Gorman, & Hamby, 2012). Based on this early preference, which is attributed to the prosodic and acoustic features of this register that likely emerge from parents' intention to transmit positive emotions in their voice (Fernald, 1992; Fernald et al., 1989; Fernald & Kuhl, 1987), IDS has been proposed to serve an attentional function in early communicative development (Spinelli, Fasolo, & Mesman, 2017). According to this attentional account, infants attend more to communicative interactions that employ the type of speech that they prefer, and which is specifically catered to their linguistic and socio-cognitive needs (Papoušek, 2007). Consequently, the heightened attention to the task incurred by this register leads to more successful performance in language-processing tasks that use IDS stimuli (Singh et al., 2002; Singh, Morgan, & White, 2004).

On the other hand, a linguistic account proposes that in addition to its attention-grabbing properties, the prosodic and linguistic exaggeration in IDS facilitate infants' processing and encoding of language-specific information conveyed in this register. For instance, acoustic exaggeration of vowel sounds in IDS relative to ADS has been proposed to result in clearer speech, which assists infants' task of discriminating sound categories in their speech input (Kuhl et al., 1997). Indeed, the degree of acoustic exaggeration of vowels in individual mothers' IDS significantly correlates to their infants' concurrent and future speech perception ability and vocabulary size (Hartman, Ratner, & Newman, 2017; Kalashnikova & Burnham, 2018; Liu, Kuhl, & Tsao, 2003). This function of vowel exaggeration has been the focus of recent debates in the literature (Martin et al., 2015; McMurray, Kovack-Lesh, Goodwin, & McEchron, 2013; Miyazawa, Shinya, Martin, Kikuchi, & Mazuka, 2017), but it is not the only component of IDS that may serve a linguistic function. Other IDS features including exaggerated pitch range and slow speech rate have been shown to assist infants' performance in language-processing tasks such as phonetic discrimination, speech segmentation, word recognition, and word learning (Graf Estes & Hurley, 2013; Ma, Golinkoff, Houston, & Hirsh-Pasek, 2011; Schreiner & Mani, 2017; Song, Demuth, & Morgan, 2010; Thiessen, Hill, & Saffran, 2005; Trainor & Desjardins, 2002).

As mentioned above, studies that have employed IDS stimuli have recorded successful turn-taking in 30-month-old children, which may have been due to the attentional benefits of using this register for stimuli presentation as proposed by the attention account of IDS. However, according to the linguistic account, it is possible that specific IDS features further assist infants' early capacity to anticipate conversational turn completion. In line with this view, the prosodic exaggeration in IDS is proposed to amplify the cues to which infants must attend in conversations, specifically final-utterance lengthening and pitch modulations (Fisher & Takura, 1996; Koponen & Lacerda, 2003). This possibility received support from a recent computational modeling study by Ludusan and colleagues (Ludusan, Cristia, Martin, Mazuka, & Dupoux, 2016) who demonstrated that models trained with IDS input were significantly more successful in identifying clause boundaries compared to models trained with ADS. Acoustic analyses of the speech signal showed that pitch cues to clausal boundaries were less reliable in IDS, but IDS did contain exaggerated pause and syllabic nuclei durations compared to ADS. Given that these cues may also reliably signal turn completion (Local & Walker, 2012), it is possible that exposure to IDS can similarly boost infants' ability to make predictions about conversational turn transitions.

This study aimed to assess the extent to which young children rely on prosodic and lexico-syntactic cues to conversational turn transitions in interactions that involve infant- and adult-directed speech. We also assessed the effects of children's lexico-syntactic competence on turn prediction by comparing performance by one- and three-year-old children. At one year of age, children understand many words in their language (Dale & Fenson, 1996), and they are sensitive to the prosodic and intonation cues to clause completion (Seidl, 2007; Soderstrom et al., 2008, 2003), but their expressive language skills and syntactic competence are still developing. In comparison, three-year-olds have more advanced receptive and expressive language abilities (Clark, 1995), and importantly this is the age at which children have previously been demonstrated to successfully anticipate conversational turns in the presence and absence of prosodic cues (Keitel & Daum, 2015; Keitel et al., 2013; Lammertink et al., 2015). Following previous studies, we employed an anticipatory looking paradigm in which children observed conversational interactions between two puppets. Puppet interactions were chosen to make them more attractive to young children, but importantly, also to enable us to manipulate the auditory stimuli without causing discontinuities in the video signal (Keitel & Daum, 2015).

In this study, children observed interactions produced in IDS and ADS, and in each register, the completion of half of the turns was marked by complete lexico-syntactic and prosodic cues, but in the other half, lexico-syntactic cues indicated completeness but prosodic cues did not (Bögels & Torreira, 2015; Grosjean & Hirt, 1996). This manipulation enabled us to assess whether children employed prosodic cues when making predictions about turn transitions. If this were the case, children should produce more anticipatory gaze shifts when predicting the completion of prosodically complete than prosodically incomplete utterances (Bogels & Torreira, 2015). However, if children only employ lexico-syntactic cues, then they should predict turn completion similarly often in both cases given that lexico-syntactic cues always predict turn completion in our task (De Ruiter et al., 2006). Additionally, we constructed two hypotheses regarding the effects of speech register and children's age on performance. First, if it is the case that children's reliance on prosodic cues is supported by the prosodic exaggeration characteristic of IDS in line with the linguistic function account of this register, then we predicted that one- and three-year-olds would anticipate more turns in the complete than incomplete prosody condition only in the IDS and not the ADS register (prosody condition

× register interaction). Alternatively, if it is the case that children do not develop an early ability to rely on prosodic cues, but only accurately predict turn completion when it is signaled by both prosody and lexico-syntax (Casillas & Frank, 2017; Keitel & Daum, 2015; Lammertink et al., 2015), then we expected that only three-year-olds and not one-year-olds would predict more turn transitions in the complete than incomplete prosody condition (prosody condition × age interaction) given that at this age children are expected to have full access to the prosodic and lexico-syntactic information of their native language regardless of the speech register in which it occurs.

Method

Thirty-eight children participated in this study: 20 were one year-old (M = 12.47 months, SD = 2.33, 10 female), and 18 were three years-old (M = 36.96 months, SD = .69, 10 female). Children were recruited via a database of families that have expressed interest in taking part in research in an infancy lab at a local university. All children were acquiring English in a monolingual context with no exposure to additional languages. Parental reports indicated that all children were typically developing, were not at risk for any developmental language or cognitive disorders and had normal hearing and vision. An additional 15 children participated but were excluded from the final sample because they failed to complete the experiment due to fussiness (10 one-year-olds) or loss of interest in the task (2 three-year-olds), equipment failure (2 one-year-olds), and experimenter error (1 one-year-old). This study received approval from the Western Sydney University Human Ethics Committee (approval number: H9142), and all written informed consent was collected from the caregivers of all children prior to their participation in the study.

Materials and Apparatus

Four dialogues on topics that are likely to be familiar and interesting to young children (going to the beach, riding bicycles, playing with toys, and going to a birthday party) were used to record the stimuli for this task. Each dialogue included 25 utterances with each utterance ranging from 5 to 12 words in length. Of these, twenty-one utterances were declarative sentences, and four were in the form of questions. Our objective was to focus on declarative statements given that questions are characterized by additional distinct lexico-semantic and intonational cues (Keitel et al., 2013; Casillas & Frank, 2017). However, some questions were included in the dialogues to resemble natural conversations as much as possible (i.e., it is unusual to have a conversation in which both interlocutors only produce declarative statements). Two female native speakers of Australian English were recorded acting out the dialogues. They were instructed to produce the dialogues as if they were engaged in a natural conversation. That is, they were not provided with specific instructions about pauses between their turns or intonational exaggeration. Two dialogues were initiated by one of the speakers and two dialogues by the other. The speakers were recorded producing each dialogue twice: once using ADS, and once using IDS. It must be noted that even though no infant addressees were present during the recording, the scripts used for IDS displayed an image of a smiling baby to help the speakers imagine their target audience. Moreover, both speakers had extensive training in identifying and producing IDS utterances in our lab.

The dialogue scripts are presented in the Appendix. The speakers produced all utterances as they appeared in these scripts, but after the recording, the 12 longest utterances from each dialogue were manipulated to create the complete and incomplete prosody conditions. Praat software (Boersma & Weenink, 2010) was used to remove the final part of each of these utterances, so that the resulting utterance was similar in length to the non-manipulated utterances and was grammatically complete. For instance, the utterance originally recorded as “sometimes you can see fish under the water” was truncated to “sometimes you can see fish”. Thus, its lexico-syntactic structure would signal the end of a conversational turn (i.e., a complete statement), but its prosody would signal that the speaker would produce more words before finishing their turn (i.e., an incomplete statement). The remaining 13 utterances were not manipulated. Therefore, an utterance recorded as “the blue water is so pretty” was kept in its original form, so that its lexico-syntactic structure and prosody signaled the completion of a conversational turn (i.e., a complete statement).

As a result, the final dialogues each comprised 25 grammatically complete utterances, 13 of which were prosodically complete and 12 prosodically incomplete. The prosodically incomplete utterances were placed in randomly assigned positions within the dialogue with the only constraint that the first and last utterances of each dialogue were prosodically complete. Note that one of the prosodically complete utterances marked the end of each dialogue and did not trigger a conversational turn, so the analyses were based on 12 utterances in each condition per dialogue.

Videos of two female hand puppets "speaking to each other" were recorded in sync with each dialogue and were later edited to ensure that puppets' mouth opening and closing coincided with the beginning and end of the utterances. The position of the two puppets on the screen (left vs. right) and the speaking order (first speaker vs. second speaker) were counterbalanced across the dialogues and the IDS and ADS conditions. An example of the visual and auditory stimuli presented as part of one conversational turn is illustrated in Figure 1.

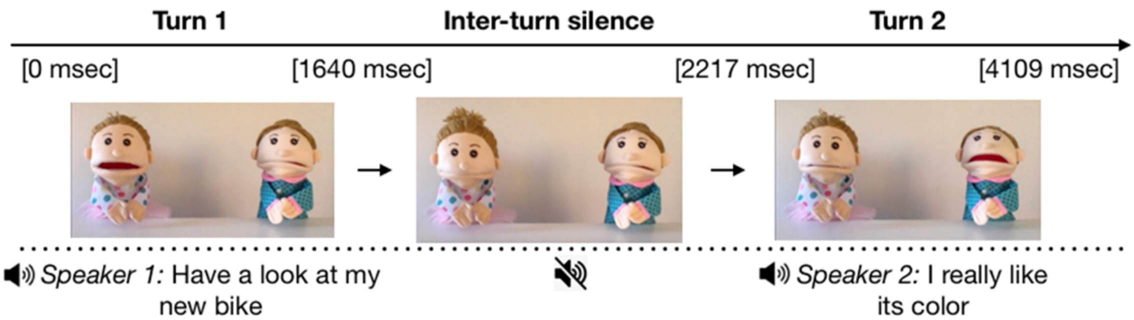

Figure 1. *Structure of a sample extract from the audio-visual stimuli.*

The duration of the audio-visual clips ranged from 45.5 to 52.3 seconds for ADS (*M* = 48.9 seconds, *SD* = 2.9) and from 57.6 to 66.2 for IDS (*M* = 60.9 seconds, *SD* = 3.8). Table 1 presents the pitch and duration measures for individual utterances in the complete and incomplete prosody conditions in the two registers. As expected, utterances in IDS had higher pitch and pitch range, were longer, and were separated by longer inter-turn silences than utterances in ADS. Given that the utterances ended in different words that varied in their length, number of syllables, and phonological form, we were unable to extract a single measure of final phrase lengthening that would allow for a direct comparison of the prosodically complete and incomplete utterances in the IDS and ADS registers. However, it was possible to calculate the IDS/ADS ratio given that each utterance was used in the IDS and ADS conditions. In this case, a ratio > 1 would denote that the final word was lengthened in IDS compared to ADS, a ratio of 1 would denote that the length of the final word was similar in the two registers, and a ratio < 1 would denote that the final word was shortened in IDS compared to ADS. One-sample t-tests showed that the ratios were significantly greater than 1 for utterances in both the complete prosody, $t(47) = 6.873$, $p < .001$, and incomplete prosody conditions, $t(47) = 4.355$, $p < .001$, so the final word was always lengthened in IDS compared to ADS. Furthermore, the ratios did not differ between the two prosody conditions, $t(94) = .868$, $p = .388$, suggesting that the degree to which the final word was lengthened in IDS compared to its ADS counterpart was similar when prosody signaled final phrase completion and when it did not. Therefore, in all cases, words in IDS were longer than in ADS, which included words in the final utterance position even if the utterance was truncated to create the incomplete prosody manipulation.

Table 1. Pitch and duration measures of utterances comprising the IDS and ADS dialogues (*IDS > ADS, $p < .01$).

	ADS		IDS	
	Complete prosody	Incomplete prosody	Complete prosody	Incomplete prosody
Mean F0 (Hz)	223.53 (35.38)	233.02 (33.29)	253.17 (37.72)*	259.89 (46.29)*
F0 Range (max Hz-min Hz)	235.56 (113.69)	203.09 (90.71)	285.72 (113.17)*	268.06 (108.96)*
Duration (msec)	1565.19 (328.71)	1321.59 (327.90)	2082.46 (449.82)*	1720.09 (479.59)*
Inter-turn silence (msec)	0.52 (0.03)	0.51 (0.02)	0.55 (0.08)*	0.55 (0.11)*

Stimuli presentation and data collection were controlled using Tobii Studio software, and a Tobii-X120 eye-tracker was used to record participants' gaze data at a 120Hz rate. Stimuli were presented on a 22-in display mounted on top of the eye-tracker, and audio was presented in stereo mode over loudspeakers hidden behind a curtain under the eye-tracker.

Procedure

During the task, children sat approximately 60cm away from the eye-tracker. One-year-olds sat on their parent's lap, and three-year-olds sat independently on a chair. Three-year-olds' caregivers also were present in the room, but they sat behind the child and were instructed to remain silent. Children were not given any specific instructions about the task, but they were told that they would watch some fun movies. Prior to the start of the task, all children completed a 5-point infant gaze calibration routine. Next, the clips were presented. In total, each child watched 8 clips (4 IDS and 4 ADS). Two presentation orders were constructed with IDS and ADS clips appearing in alternating order and ensuring that the IDS and ADS versions of the same dialogue were not presented consecutively. Given that it was more challenging for the three-year-olds to passively watch the clips without any specific instructions, they were shown additional brief cartoon videos after every two dialogues to maintain them engaged in the task.

Processing of Eye-tracking Data

Two areas of interest (AoI) were defined encompassing each puppet's face. The size and location of the AoIs were identical for each video. Gaze data were then extracted from Tobii Studio and information about the register condition and utterance type was added. The eye-tracking R package (Dink & Ferguson, 2015) was used in R (R Core Team, 2013) to process the raw data. In this step, the start time of each clip was re-set to zero, and the anticipatory and non-anticipatory time windows were defined. The number of gaze shifts produced for each conversational turn was computed. First, during the *non-anticipatory* time window, a score of 1 was assigned if a gaze shift from the speaking puppet to the listening puppet was produced during the speaking puppet's turn. Next, in the *anticipatory* time window, a score of 1 was assigned if a gaze shift satisfied the criteria defined by Casillas and Frank (2017). That is, a gaze shift was considered anticipatory if (1) the participant fixated the non-speaking puppet, (2) this fixation occurred within the anticipation window that lasted from 200msec prior to the end of the speaking puppet's turn until 200msec after the beginning of the responder's turn, and (3) the participant had fixated on the speaking puppet for at least 100msec before the beginning of the anticipation window (i.e., the participant had to switch their gaze from the speaker to the listener and not simply fixate the listener).

Results

Preliminary Analyses

Prior to analyzing gaze shift data, children's overall gaze patterns during the IDS and ADS clips were compared to assess that they attended to the stimuli and complied with the task. For this purpose, we calculated the proportions of looking time directed to the speaking puppet out of the total looking time to the two puppets during each utterance. These scores are presented in Table 2, and as can be seen, one-sample *t*-test analyses confirmed that children in the two age groups looked at the speaking puppet above chance levels (chance = .5). Therefore, children were engaged in the task and looked at each puppet during its speaking turn. In addition, paired *t*-test analyses indicated that the proportion of looking at the speaking puppet was similar in IDS and ADS conditions for one-year-olds, but three-year-olds looked at the speaking puppet significantly more when they heard IDS than ADS.

Table 2. Proportion of looking time to the speaking puppet during the ADS and IDS clips (**$p < .01$).

	ADS		IDS		IDS vs. ADS
Age	Mean (SD)	*t*	Mean (SD)	*t*	*t*
1-year-olds	.65 (.072)	9.55**	.65 (.071)	9.42**	-0.417

3-years-olds	.69 (.069)	10.57**	.74 (.061)	16.25**	3.638**

Gaze Shift Analyses

Children's gaze data during the key time windows (non-anticipatory and anticipatory) were analyzed using binomial generalized linear mixed effects (GLME) models conducted using the lme4 package (Bates, 2005) in R (R Core Team, 2013). When appropriate, the lsmeans package (Lenth, 2016) was used for post-hoc least square means pairwise comparisons, which controlled for multiple comparisons using the Tukey correction method.

Non-anticipatory gaze shifts. First, children's gaze switching behaviors during the non-anticipatory time window of each conversational turn were compared across ages and registers. These analyses capture children's random tendency to switch their gaze from one speaker to another. Given that during this window, the prosodically manipulated and non-manipulated utterances were identical, these analyses did not include the prosody condition as a factor. A binomial GLME model (Model 1) was constructed with switch score as the dependent variable and Age (1 year, 3 years) and Register (IDS, ADS) as the predictor variables and random intercepts for participants. The output of this model is shown in Table 3. As can be seen it yielded no main effects of age and register, and no significant interaction. This result enables us to conclude that any effects of age, register, or prosody condition identified on children's anticipatory looking rates in our next set of analyses reflect differences in their tendency to switch their gaze in anticipation of an upcoming conversational turn and not differences in their overall tendency to switch their gaze from one speaker to another in the course of a conversation. Furthermore, this indicates that it was not the case that children fixated the speaking puppet more consistently in the IDS compared to the ADS condition, but instead they were equally likely to switch their gaze between the puppets while observing the conversations produced in both registers.

Table 3. Output of GLME Model 1 assessing random gaze switches in the non-anticipatory time window.

	Estimate	SE	*Z* value	*p*
Intercept	-0.549	0.123	-4.473	<*.001*
Age [3 years]	0.116	0.163	0.711	.477
Register [IDS]	0.153	0.116	1.315	.189
Age [3 yers] × Register [IDS]	-0.017	0.162	-0.107	.914

Anticipatory gaze shifts. The proportions of anticipatory looks produced during the anticipatory time window in each register and condition by one- and three-year-olds are displayed in Figure 2. The GLME Model 2 included children's anticipatory gaze scores as the dependent variable, Age (1 year, 3 years), Register (IDS, ADS), and Prosody Condition (Complete, Incomplete) as the independent variables, and random intercepts for participants (detailed output for Model 1 is shown in Table 4). A main effect of register indicated that children produced more anticipatory looks during the IDS than the ADS clips, $\beta = .565$, $SE = .277$, $Z = 2.038$, $p = .042$. The main effects of age and prosody condition were not significant, but there were significant interactions of age × prosody condition, $\beta = -1.365$, $SE = .413$, $Z = -3.309$, $p = .001$, and register × prosody condition, $\beta = -.899$, $SE = .393$, $Z = -2.291$, $p = .022$.

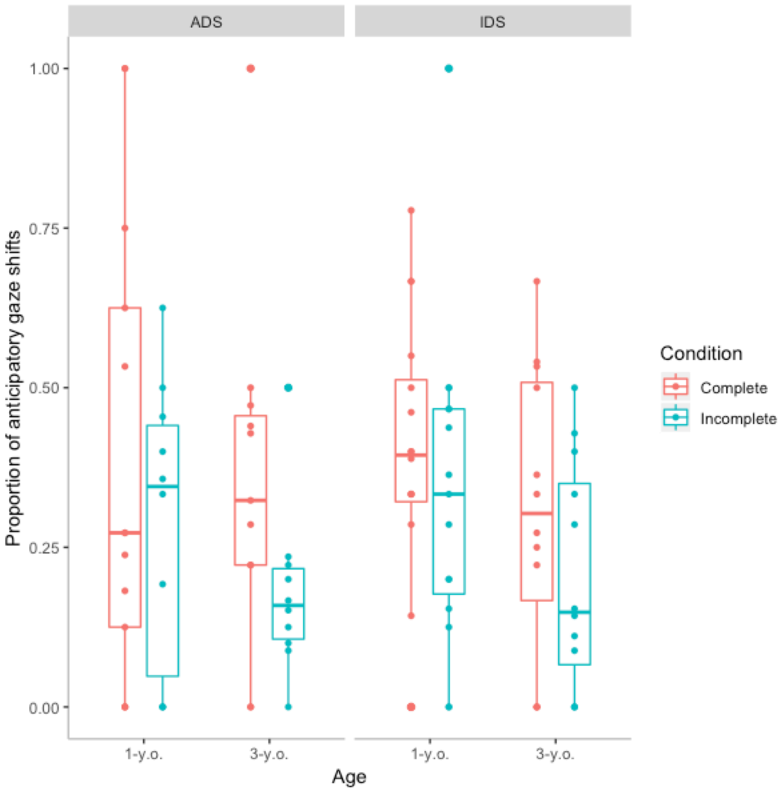

Figure 2. *Proportion of anticipatory gaze shifts produced during the anticipatory window by one- and three-year-olds in the complete and incomplete prosody conditions of the IDS and ADS registers.*

Table 4. Output of GLME Model 2 assessing children's gaze switches during the anticipatory looking time window.

	Estimate	SE	*Z* value	*p*
Intercept	-0.879	0.244	-3.6	*<.001*
Age [3 years]	0.209	0.324	0.644	.519
Register [IDS]	0.565	0.277	2.038	*.042*
Prosody condition [Incomplete]	0.331	0.301	1.1	.271
Age [3 yers] × Register [IDS]	-0.371	0.375	-0.991	.322
Age [3 years] × Prosody condition [Incomplete]	-1.365	0.413	-3.309	*.001*
Register [IDS] × Prosody condition [Incomplete]	-0.899	0.393	-2.291	*.022*
Age [3 years] × Register [IDS] × Prosody [Incomplete]	0.765	0.561	1.363	.173

In order to identify the source of the significant two-way interactions, data for the IDS and ADS registers were analyzed separately in two identical GLME models (Models 3 and 4) with the exception that Register was no longer included as a factor (DV = anticipatory looking score, IV = age, prosody condition, Random Intercept = participant). The output of Models 3 and 4 is presented in Table 5. Model 3, which assessed children's performance when the stimuli were in IDS, yielded no main effect of Age, but a significant effect of Prosody Condition, β = -.547, *SE* = .251, *Z* = -2.177, *p* = .029. That is, one- and three-year-old children produced more anticipatory looks in the complete prosody than in the incomplete prosody condition when they listened to conversations produced in IDS.

On the contrary, Model 4 for ADS, showed no main effects of Age and Prosody Condition but a significant Age × Prosody Condition interaction, β = -1.370, *SE* = .411, *Z* = -3.331, *p* = .001. Follow up least square means comparisons showed that the difference in performance between the complete and incomplete prosody conditions was significant for the three-year-olds, Odds Ratio = 2.822, *SE* = .794, *Z* = 3.686, *p* = .001, but not for the one-year-olds, Odds Ratio = .717, *SE* = .215, *Z* = -1.111, *p* = .683. Therefore, when presented with ADS stimuli, only three-year-olds showed differential performance across the two prosody conditions.

Table 5. Output of GLME Models 3 (IDS) and 4 (ADS).

Model 3: IDS				
	Estimate	*SE*	*Z* value	*p*
(Intercept)	-0.328	0.196	-1.678	.093
Age [3 years]	-0.140	0.296	-0.474	.636
Prosody conditon [Incomplete]	-0.547	0.251	-2.177	*.029*
Age [3 years] × Prosody condition [Incomplete]	-0.627	0.379	-1.654	.098
Model 4: ADS				
	Estimate	*SE*	*Z* value	*p*
(Intercept)	-0.840	0.236	-3.556	*.001*
Age [3 years]	0.204	0.314	0.650	.516
Prosody conditon [Incomplete]	-0.333	0.299	1.111	.267
Age [3 years] × Prosody condition [Incomplete]	-1.370	0.411	-3.331	*<.001*

Discussion

This study assessed one- and three-year-old children's ability to employ prosodic and lexico-syntactic cues to predict conversational turn completion when these occurred in the context of infant-directed and adult-directed speech. Our results showed that children produced more anticipatory gaze shifts, which indicates that they were predicting that a turn transition was about to take place, when stimuli were presented in IDS than in ADS. Most importantly, one-year-olds predicted turn-completion more successfully in the complete than in the incomplete prosody condition only in IDS, whereas three-year-olds did so regardless of the speech register used in the conversation. Therefore, children's reliance on the speech register and prosodic information therein corresponded to their age and linguistic ability. One-year-olds were more likely to make predictions about the turn structure of a conversation when speech was infant-directed and when both the lexico-syntactic and prosodic cues in the stimuli unambiguously signaled turn-completion. On the other hand, three-year-olds also anticipated most turns when they were signaled by the combination of lexico-syntactic and prosodic cues, but they were able to do so independently of the properties of each specific speech register.

Even though young infants start engaging in conversational turn taking already in their first months of life (Gratier et al., 2015; Hilbrink et al., 2015), adult-like turn-taking competence is slow to develop (Casillas et al., 2016). This is in part due to the demands of integrating linguistic processing into early proto-conversations between infants and their caregivers (Hilbrink et al., 2015), and infants' developing ability to detect and encode the various linguistic and paralinguistic cues that enable conversational partners to make predictions about the structure of a conversation. Previous research has proposed that children are unable to make reliable predictions about the timing of conversational turn transitions before the age of two years (Casillas & Frank, 2017), with some suggesting that this ability develops even later in childhood (Keitel & Daum, 2015), and that from early on, children's predictions are based on a combination of lexico-syntactic and prosodic cues despite having much earlier access to prosody than lexico-syntax in their native language (Casillas & Frank, 2017; Lammertink et al., 2015). The present findings suggest that these abilities are already manifested around children's first birthday, and they can be elicited in experimental settings by providing children with infant-directed stimuli.

Specifically, our findings demonstrate that when one-year-olds listened to stimuli presented in IDS, they produced more anticipatory gaze switches at the end of utterances that were prosodically complete than at the end of utterances that were prosodically incomplete. We acknowledge that our study had several limitations that prevent us from clearly determining to what extent each factor (register and prosodic cues) influenced one-year-olds' performance in our task. First, our conservative sample size and our study design did not allow for additional tests of whether one-year-olds did or did not show some turn anticipation in the prosodically incomplete IDS utterances or any of the ADS utterances. Our task also was unable to discern whether one-year-olds also relied on lexico-syntactic cues to turn completion in addition to prosody. It is possible that they utilized some of this information given that by 12 months, children already have impressive receptive vocabulary abilities (e.g., see Bergelson & Swingley, 2012), and we specifically constructed utterances with lexical and grammatical properties of IDS that were age-appropriate for the youngest children in our sample. Importantly, it is unlikely that 12-month-olds relied solely on lexico-syntactic cues and not on prosody given that they did not show turn prediction in the ADS condition, which contained identical lexico-syntactic content, and their predictions were specifically sensitive to prosodic cues that signaled utterance completion.

The finding that one-year-old children produced more anticipatory gaze switches when utterances were prosodically complete and produced in IDS, however, leads us to conclude that young children can rely on their early prosodic competence to make predictions about the structure of a conversation. Before their first birthday, infants are sensitive to prosodic cues that mark grammatical clause boundaries such as pitch modulations and final phrase lengthening (Johnson & Seidl, 2008; Seidl & Cristià, 2008; Soderstrom et al., 2003). This enables young infants to parse continuous speech into complete clausal units and extract meaningful linguistic information from these units (Nazzi, Kemler Nelson, Jusczyk, & Jusczyk, 2000). Pitch and durational components are both exaggerated in IDS compared to ADS (Fernald & Simon, 1984; Fernald & Kuhl, 1987), but previous research has casted doubt on the linguistic functionality of this exaggeration. That is, pitch and durational modulations in natural IDS aimed at transmitting positive affect and attract infants' attention to speech may increase the variability in these components and decrease their informativeness as prosodic cues to clausal completion (Kempe et al., 2010). Our findings suggest that this is not the case. As demonstrated by the acoustic analyses of our stimuli, IDS utterances contained significantly more variable pitch contours compared to ADS utterances, but this additional prosodic variability did not interfere with infants' sensitivity to the prosodic cues that signaled turn completion. These results converge with the findings from computational modeling experiments demonstrating that models are more successful at categorizing durational cues to utterance completion from IDS than ADS (Ludusan et al., 2016). In fact, it has been proposed that the high prosodic variability in IDS, which results in reduced predictability, may instead increase the informational value of this register, which results in greater attention to the IDS signal in young infants and boosts their ability to extract and encode linguistic and paralinguistic information from their input (Räsänen, Kakouros, & Soderstrom, 2018).

In addition to the prosodic cues that specifically signal utterance completion (Local & Walker, 2012), IDS has several features that may contribute to enhancing children's ability to anticipate conversational turn transitions. First, IDS is characterized by longer inter-turn silences (Cooper & Aslin, 1990; Fernald & Kuhl, 1987; Martin, Igarashi, Jincho, & Mazuka, 2016), which was also the case in our stimuli. When the upcoming turn follows a longer silence, children have more time to react to the completion of the previous turn and direct their attention to the next speaker. Casillas and Frank (2017) directly demonstrated that this is the case not only for children but also for adults whose prediction accuracy was positively related to the duration of the silence gaps between two turns. However, our findings suggest that it is unlikely that longer inter-turn-silences were the only cue that enabled one-year-olds to predict conversational transitions in our task. The duration of inter-turn silences was comparable for the complete and incomplete prosody utterances in IDS (see Table 1), so if infants were only anticipating turn transitions during the long silence gaps in this register, no effect of prosody condition would have been observed. Nevertheless, it is likely that long inter-turn silences are a powerful feature of IDS that caregivers use to scaffold early lexical and morphosyntactic development (Soderstrom, 2007), and which also yields benefits for the early conversational turn taking abilities in young infants. The use of short utterances that are separated by extended inter-turn silences may decrease the load associated with speech processing during conversational interactions and enable infants to encode the prosodic information that reliably signals when turn transitions are about to occur.

As mentioned in the Introduction, the prosodic exaggeration in IDS also serves an attentional function, which may have additionally fostered children's performance in this task. Indeed, regardless of age, children produced more anticipatory gaze shifts in the IDS compared to the ADS condition. However, we argue that it is unlikely that an overall preference for IDS stimuli could account for all the findings observed in this study. First, three-year-olds looked more at the speaking puppet in the IDS than the ADS condition, but this was not the case for one-year-olds. Second, children's tendency to produce non-anticipatory gaze switches was not affected by the speech register, which suggests that it was not the case that children fixated the speaking puppet more consistently or were more likely to shift their gaze between the puppets when they heard IDS or ADS. Finally, the main effect of register observed for the anticipatory time window was qualified by a prosody condition by register interaction. Therefore, we conclude that children were sensitive to prosodic cues to turn transitions over above the general attention-getting properties of this register. This is the first study to demonstrate this effect in relation to early turn-taking competence, but our findings dovetail with evidence showing that children are more successful in various language processing tasks when presented with IDS compared to ADS stimuli including phonetic discrimination (Trainor & Desjardins, 2002), speech segmentation (Thiessen et al., 2005), word recognition (Singh et al., 2004; Song et al., 2010), and word learning (Graf Estes & Hurley, 2013; Ma et al., 2011).

The three-year-olds in this study showed more accurate anticipation of conversational turn transitions for prosodically complete utterances not only in IDS but also in ADS. This results pattern was expected given that previous studies have reported successful turn predictions for children after 30 months of age in response to both infant- and adult-directed stimuli (Casillas & Frank, 2017; Keitel & Daum, 2015; Keitel et al., 2013; Lammertink et al., 2015). These findings indicate that children's sensitivity to the prosodic markers to conversational turn transitions generalize to ADS during their third year of life. We consider two possible explanations for the time frame at which this transition takes place. First, child-directed speech, the speech register directed to pre-school children of two years of age and older, continues to be distinct from ADS along several prosodic and linguistic dimensions, but its pitch height and range become significantly less exaggerated compared to IDS during children's second year of life (Ko, Seidl, Cristia, Reimchen, & Soderstrom, 2016; Narayan & McDermott, 2016; Vosoughi & Roy, 2012). This effect is also likely to be bi-directional since parents tend to produce speech with prosodic, affective, and linguistic properties that correspond to their infants' linguistic and communicative needs (Kitamura & Burnham, 2003), and infants prefer the type of speech register that is suited to their age (Kitamura & Lam, 2009; Spence & Moore, 2003). Therefore, children's more mature language processing skills and their more extensive exposure to less prosodically exaggerated speech result into a reduced reliance on the prosodically exaggerated components of IDS for successful linguistic processing. The second and related possibility is that children become capable to rely on their more advanced linguistic competence to compensate for exposure to less exaggerated and less engaging ADS stimuli. That is, in this task, three-year-olds did have access to the lexico-syntactic content of the utterances, and this additional information could have supported their conversational turn predictions even when the prosodic cues were less salient in ADS compared to IDS. Similar transitions to successful linguistic processing in ADS occur at different time points depending on the demands of specific experimental tasks. For instance, infants become able to segment words from continuous speech in ADS at 16 months of age (Mani & Patzold, 2016), but they only become successful at learning novel words in ADS at 27 months (Ma et al., 2011). Therefore, it appears that as children accrue more advanced linguistic competence, their reliance on the specific features of IDS gradually decreases allowing them to succeed in language processing tasks regardless of the speech register in use.

Three-year-old children in this study demonstrated the ability to predict the completion of ongoing conversational turns when it was signaled by a combination of prosodic and lexico-syntactic cues. This result replicates the findings from previous literature that indicate that despite having access to lexico-syntactic information that can reliably indicate that a turn has been completed, children and adults make more accurate predictions when this information is complemented with prosodic cues to turn completion (Bogels & Torreira, 2015; Casillas & Frank, 2017; Keitel & Daum, 2015; Lammertink et al., 2015). The current design did not include a condition in which turn transitions were signaled by lexico-syntactic cues in the absence of contradictory prosodic information (e.g., pitch-flattened speech), but previous studies suggest that while adults can make turn predictions in such conditions (De Ruiter et al., 2006), this is not the case for pre-school children (Casillas & Frank, 2017). Therefore, at three years of age, children exhibit sophisticated abilities to make predictions about the turn structure of an ongoing conversation, but these abilities are still developing; while for adults prosodic information is useful but not essential for turn prediction accuracy (De Ruiter et al., 2006), children's predictions are restricted to cases signaled by both prosodic and lexico-syntactic cues.

Conversational turns mark the structure of communicative interactions and reveal the complex linguistic processing involved during spontaneous conversations. It is, therefore, not surprising that young language users find it difficult to combine their emerging language-specific knowledge of prosodic markers to turn transitions and the lexico-syntactic content required for successful speech comprehension and production. This study demonstrates that children before the age of three years benefit from exposure to infant-directed speech in this process, and even after the age of three, they continue to employ all the prosodic and lexico-syntactic information available in their input to accurately predict when a conversational turn will end. Caregivers employ infant-directed speech spontaneously when interacting with their young infants. Importantly, this speech register may not only introduce infants to conversational structure in their native language, but it also directs their attention to the subtle prosodic cues to alleviate the processing challenges incurred during conversational turn-taking.

References

Bates, D. (2005). Fitting linear mixed models in R. *R News, 5*(1), 27–30.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., ... & Wicha, N. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*(2), 344-380.

Bateson, M. C. (1975). Mother-infant exchanges: The epigenesis of conversational interaction. *Annals of the New York Academy of Sciences, 263*, 101–113. doi: 10.1111/j.1749-6632.1975.tb41575.x

Bergelson, E., & Swingley, D. (2012). At 6-9 months, human infants know the meanings of many common nouns. *Proceedings of the National Academy of Sciences, 109*(9), 3253–3258. https://doi.org/10.1073/pnas.1113380109

Boersma, P., & Weenink, D. (2010). Praat: Doing phonetics by computer (version 5.3.16).

Bogels, S., & Torreira, F. (2015). Listerners use intonational phrase boundaries to project turn ends in spoken interaction. *Journal of Phonetics, 52*, 46–57.

Bögels, S., & Levinson, S. C. (2017). The brain behind the response: Insights into turn-taking in conversation from neuroimaging. *Research on Language and Social Interaction*, *50*(1), 71-89.

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

Casillas, M., Bobb, S., & Clark, E. (2016). Turn-taking, timing, and planning in early language acquisition. *Journal of Child Language, 43*(6), 1310–1337. https://doi.org/10.1017/s0305000915000689

Casillas, M., & Frank, M. C. (2017). The development of children's ability to track and predict turn structure in conversation. *Journal of Memory and Language, 92*, 234–253. https://doi.org/10.1016/j.jml.2016.06.013

Clark, E. V. (1995). Language acquisition: The lexicon and syntax. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition (2nd ed.), Vol. 11. Speech, language, and communication* (p. 303–337). Academic Press. https://doi.org/10.1016/B978-012497770-9/50011-X

Clayman, S. E. (2013). Turn-constructional units and the transition-relevance place. *The Handbook of Conversation Analysis*, 150-166.

Cooper, R. B., & Aslin, R. . (1990). Preference for infant-directed speech in the first month after birth. *Child Development, 61*(5), 1584–1595.

Cristia, A. (2013). Input to Language: The Phonetics and Perception of Infant-Directed Speech. *Linguistics and Language Compass, 7*(3), 157–170. https://doi.org/10.1111/lnc3.12015

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments, and Computers, 28*(1), 125–127. https://doi.org/10.3758/BF03203646

De Ruiter, J. P., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language, 82*(3), 515–535.

Dink, J. W., & Ferguson, B. (2015). eyetrackingR: An R Library for Eye-tracking Data Analysis.

Dunst, C. J., Gorman, E., & Hamby, D. W. (2012). Preference for infant-directed speech in preverbal young children. *Center for Early Literacy Learning Reviews, 5*(1), 1–13.

Fernald, A. (1992). Human maternal vocalizations to infants as biologically relevant signals: An evolutionary perspective. *The Adapted Mind*, 391–428. https://doi.org/10.1007/BF00852474

Fernald, A. (1993). Approval and disapproval : Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development, 64*(3), 657–674.

Fernald, A., & Simon, T. (1984). Expanded intonation contours in mothers' speech to newborns. *Developmental Psychology, 20*(1), 104–113. https://doi.org/10.1037/0012-1649.20.1.104

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

Fernald, H., & Kuhl, P. (1987). Acoustic determinants of infant preference for motherse speech. *Infant Behaviour and Development, 10*, 279–293.

Fisher, C., & Tokura, H. (1996). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. *Child Development*, *67*(6), 3192-3218.

Frota, S., Butler, J., & Vigário, M. (2014). Infants' perception of intonation: Is it a statement or a question? *Infancy, 19*(2), 194–213. https://doi.org/10.1111/infa.12037

Golinkoff, R. M., Can, D. D., Soderstrom, M., & Hirsh-Pasek, K. (2015). (Baby)Talk to Me: The Social Context of Infant-Directed Speech and Its Effects on Early Language Acquisition. *Current Directions in Psychological Science, 24*(5), 339–344. https://doi.org/10.1177/0963721415595345

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

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interactions between mothers and infants. *Frontiers in Psychology, 6*(1167), 236–245. https://doi.org/doi: 10.3389/fpsyg.2015.01167

Grosjean, F., & Hirt, C. (1996). Using prosody to predict the end of sentences in English and French: Normal and brain-damaged subjects. *Language and Cognitive Processes*, *11*(1-2), 107-134.

Hartman, K. M., Ratner, N. B., & Newman, R. S. (2017). Infant-directed speech (IDS) vowel clarity and child language outcomes. *Journal of Child Language, 44*(5), 1140–1162. https://doi.org/10.1017/S0305000916000520

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: a longitudinal study of mother–infant interaction. *Frontiers in Psychology, 6*(September), 1–12. https://doi.org/10.3389/fpsyg.2015.01492

Höhle, B., Bijeljac-Babic, R., Herold, B., Weissenborn, J., & Nazzi, T. (2009). Language specific prosodic preferences during the first half year of life: Evidence from German and French infants. *Infant Behavior and Development, 32*(3), 262–274. https://doi.org/10.1016/j.infbeh.2009.03.004

Indefrey, P., & Levelt, W. J. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101-144.

Johnson, E., & 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. W., Cutler, A., & Redanz, N. (1993). Infants' Preference for the Predominant Stress Patterns of English Words. *Child Development, 64*(3), 675–687.

Jusczyk, P. W., & Luce, P. A. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*(5), 630–645. https://doi.org/10.1006/jmla.1994.1030

Kalashnikova, M., Carignan, C., & Burnham, D. (2017). The origins of babytalk: Smiling, teaching or social convergence? *Royal Society Open Science, 4*(8). https://doi.org/10.1098/rsos.170306

Kalashnikova, M., & Burnham, D. (2018). Infant-directed speech from seven to nineteen months has similar acoustic properties but different functions. *Journal of Child Language, 45*(05), 1035–1053. https://doi.org/10.1017/S0305000917000629

Keitel, A., & Daum, M. M. (2015). The use of intonation for turn anticipation in observed conversations without visual signals as source of information. *Frontiers in Psychology, 6*(108), 265–273. https://doi.org/doi: 10.3389/fpsyg.2015.00108

Keitel, A., Prinz, W., Friederici, A. D., Hofsten, C. von, & Daum, M. M. (2013). Perception of conversations: The importance of semantics and intonation in children's development. *Journal of Experimental Child Psychology, 116*(2), 264–277. https://doi.org/10.1016/j.jecp.2013.06.005

Kempe, V., Schaeffler, S., & Thoresen, J. C. (2010). Prosodic disambiguation in child-directed speech. *Journal of Memory and Language, 62*(2), 204–225. https://doi.org/10.1016/j.jml.2009.11.006

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. https://doi.org/10.1207/S15327078IN0401_5

Kitamura, C., & Lam, C. (2009). Age-specific preferences for infant-directed affective intent. *Infancy, 14*(1), 77–100. https://doi.org/10.1080/15250000802569777

Ko, E.-S., Seidl, A., Cristià, A., Reimchen, M., & Soderstrom, M. (2016). Entrainment of prosody in the interaction of mothers with their young children. *Journal of Child Language, 43*(2), 284–309. https://doi.org/10.1017/s0305000915000203

Koponen, E., & Lacerda, F. (2003). Final lengthening in infant directed speech may function as a cue to phrase constituents. *Phonum, 9*, 9-12.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V, Ryskina, V. L., … Lacerda, F. (1997). Cross-Language Analysis of Phonetic Units in Language Addressed to Infants. *Science, 277*(5326), 684–686. https://doi.org/10.1126/science.277.5326.684

Lammertink, I., Casillas, M., Benders, T., Post, B., & Fikkert, P. (2015). Dutch and English toddlers' use of linguistic cues in predicting upcoming turn transitions. *Frontiers in Psychology, 6*(495), 274–291. https://doi.org/doi: 10.3389/fpsyg.2015.00495

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33.

Levinson, S. C. (2016). Turn-taking in human communication - Origins and implications for language processing. *Trends in Cognitive Sciences, 20*(1), 6–14. https://doi.org/10.1016/j.tics.2015.10.010

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in psychology*, *6*, 731.

Lindsay, L., Gambi, C., & Rabagliati, H. (2019). Preschoolers optimize the timing of their conversational turns through flexible coordination of language comprehension and production. *Psychological Science, 30*(4), 504–515. https://doi.org/10.1177/0956797618822802

Liu, H. M., Kuhl, P. K., & Tsao, F. M. (2003). An association between mothers' speech clarity and infants' speech discrimination skills. *Developmental Science, 6*(3), 1–10. https://doi.org/10.1111/1467-7687.00275

Local, J., and Walker, G. (2012). How phonetic features project more talk. *Journal of the International Phonetic Association, 42*, 255–280. doi: 10.1017/S0025100312000187

Ludusan, B., Cristia, 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

Magyari, L., & de Ruiter, J. P. (2012). Prediction of turn-ends based on anticipation of upcoming words. *Frontiers in Psychology, 3*(OCT), 1–9. https://doi.org/10.3389/fpsyg.2012.00376

Magyari, L., Bastiaansen, M. C., de Ruiter, J. P., & Levinson, S. C. (2014). Early anticipation lies behind the speed of response in conversation. *Journal of Cognitive Neuroscience*, *26*(11), 2530-2539.

Magyari, L., De Ruiter, J. P., & Levinson, S. C. (2017). Temporal preparation for speaking in question-answer sequences. *Frontiers in Psychology*, *8*, 211.

Mani, N., & Patzold, W. (2016). Sixteen-month-old infants' segment words from infant- and adult-directed speech. *Language Learning and Development, 12*(4), 499–508. https://doi.org/10.1080/15475441.2016.1171717

Martin, A., Igarashi, Y., Jincho, N., & Mazuka, R. (2016). Utterances in infant-directed speech are shorter, not slower. *Cognition, 156*, 52–59. https://doi.org/10.1016/j.cognition.2016.07.015

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

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 494*, 465–494.

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. https://doi.org/10.1016/j.cognition.2013.07.015

Miyazawa, K., Shinya, T., Martin, A., Kikuchi, H., & Mazuka, R. (2017). Vowels in infant-directed speech: More breathy and more variable, but not clearer. *Cognition, 166*, 84–93. https://doi.org/10.1016/j.cognition.2017.05.003

Narayan, C. R., & McDermott, L. C (2016). Speech rate and pitch characteristics of infant-directed speech: Longitudinal and cross-linguistic observations. *The Journal of the Acoustical Society of America, 139*, 1272, doi: 10.1121/1.4944634

Nazzi, T, Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 756–766. https://doi.org/10.1037/0096-1523.24.3.756

Nazzi, T., Kemler Nelson, D. G., 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

Nazzi, T., & Ramus, F. (2003). Perception and acquisition of linguistic rhythm by infants. *Speech Communication, 41*, 233–243. https://doi.org/10.1016/S0167-6393(02)00106-1

Panneton, R., Kitamura, C., Mattock, K., & Burnham, D. (2006). Slow speech enhances younger but not older infants' perception of vocal emotion. *Research in Human Development, 3*(1), 7–19. https://doi.org/10.1207/s15427617rhd0301_2

Papoušek, M. (2007). Communication in early infancy: An arena of intersubjective learning. *Infant Behavior and Development, 30*(2), 258–266. https://doi.org/10.1016/j.infbeh.2007.02.003

Räsänen, O., Kakouros, S., & Soderstrom, M. (2018). Is infant-directed speech interesting because it is surprising? Linking properties of IDS to statistical learning and attention at the prosodic level. *Cognition, 178*, 193–206. https://doi.org/10.1016/j.cognition.2018.05.015

Schegloff, E., Jefferson, G., & Sacks, H. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696-735.

Schreiner, M. S., & Mani, N. (2017). Listen up! Developmental differences in the impact of IDS on speech segmentation. *Cognition, 160*, 98–102. https://doi.org/10.1016/j.cognition.2016.12.003

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

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

Singh, L., Morgan, J. L., & White, K. S. (2004). Preference and processing: The role of speech affect in early spoken word recognition. *Journal of Memory and Language, 51*(2), 173–189. https://doi.org/10.1016/j.jml.2004.04.004

Smith, N. A., & McMurray, B. (2018). Temporal responsiveness in mother-child dialogue: A longitudinal analysis of children with normal hearing and hearing loss. *Infancy, 23*(3), 410–431. https://doi.org/10.1111/infa.12227

Snow, C. E. (1977). Mothers' speech research: From input to interaction. *Talking to children: Language input and acquisition*, *3149*.

Soderstrom, M. (2007). Beyond babytalk: Re-evaluating the nature and content of speech input to preverbal infants. *Developmental Review, 27*(4), 501–532. https://doi.org/10.1016/j.dr.2007.06.002

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., Kemler Nelson, D. G., & 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

Song, J. Y., Demuth, K., & Morgan, J. (2010). Effects of the acoustic properties of infant-directed speech on infant word recognition. *The Journal of the Acoustical Society of America, 128*(1), 389–400. https://doi.org/10.1121/1.3419786

Spence, M. J., & Moore, D. S. (2003). Categorization of infant-directed speech: Development from 4 to 6 months. *Developmental Psychobiology, 42*(1), 97–109. https://doi.org/10.1002/dev.10093

Spinelli, M., Fasolo, M., & Mesman, J. (2017). Does prosody make the difference? A meta-analysis on relations between prosodic aspects of infant-directed speech and infant outcomes. *Developmental Review, 44*(April), 1–18. https://doi.org/10.1016/j.dr.2016.12.001

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587-10592.

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

Trainor, L. J., & Desjardins, R. N. (2002). Pitch characteristics of infant-directed speech affect infants' ability to discriminate vowels. *Psychonomic Bulletin & Review, 9*(2), 335–340. https://doi.org/10.3758/BF03196290

Vosoughi, S., & Roy, D. (2012). A longitudinal study of prosodic exaggeration in child-directed speech. *Proceedings of the 6th International Conference on Speech Prosody, SP 2012, 1*, 194–197.

Walker, G. (2013). "Phonetics and prosody in conversation," in Handbook of Conversation Analysis, Eds T. Stivers and J. Sidnell (Chichester: Wiley- Blackwell), 455–474.