HHS Public Access

Author manuscript

Dev Neuropsychol. Author manuscript; available in PMC 2020 June 01.

Published in final edited form as:

Dev Neuropsychol. 2016 ; 41(5-8): 324–341. doi:10.1080/87565641.2016.1256403.

The Multisensory Nature of Verbal Discourse in Parent–Toddler Interactions

Sumarga H. Suanda, **Linda B. Smith**, **Chen Yu**

Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana

Abstract

Toddlers learn object names in sensory rich contexts. Many argue that this multisensory experience facilitates learning. Here, we examine how toddlers' multisensory experience is linked to another aspect of their experience associated with better learning: the temporally extended nature of verbal discourse. We observed parent–toddler dyads as they played with, and as parents talked about, a set of objects. Analyses revealed links between the multisensory and extended nature of speech, highlighting inter-connections and redundancies in the environment. We discuss the implications of these results for our understanding of early discourse, multisensory communication, and how the learning environment shapes language development.

Toddlers learn object names in an environment rich in regularities and structure. Dependable patterns for learning can be observed at multiple time-scales. At the scale of individual utterances, parents' speech is often coupled to what their toddlers are seeing and doing, yielding utterances that are highly multisensory in nature (Frank, Tenenbaum, & Fernald, 2013; Gogate, Bahrick, & Watson, 2000; Harris, Jones, & Grant, 1983; Messer, 1978; Zukow-Goldring, 1990). At a slightly longer time-scale, as shown in (1) below, individual utterances are usually inter-connected, forming episodes of coherent discourse about one object (Frank et al., 2013; Messer, 1980). Both of these facets of the learning environment—its multisensory nature (e.g., Yu & Smith, 2012) and its temporally-extended nature (e.g., Schwab & Lew-Williams, 2016), have been shown independently to support lexical development. The key hypothesis of the current study is that these two facets are inter-related: extended episodes of verbal discourse are also highly multisensory. The importance of testing this hypothesis is in revealing the redundancies in toddlers' learning environment and in raising questions about the mechanisms by which the environment shapes learning.

(1) Mother: *oh there's a super car*?

Mother: *you like cars don't you?*

Mother: *what are you going to do with it*?

Mother: *are you going to make it go?*

(Messer, 1980).

CONTACT Sumarga H. Suanda, ssuanda@indiana.edu, Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN 47405-7007.

Statistical analyses of parents' speech to their children have convincingly demonstrated that multi-utterance episodes to the same referent, similar to (1), are pervasive in child-directed speech (Frank et al., 2013; Messer, 1980). A number of researchers have argued, and empirically demonstrated, that repetitive and interconnected utterances aid with speech perception (Bard & Anderson, 1983), word segmentation (Onnis, Waterfall, & Edelman, 2008), word-referent mapping (Schwab & Lew-Williams, 2016; Vlach & Johnson, 2013), semantic development (Clark, 2010), and syntax learning (Hoff-Ginsberg, 1986). By analyzing the linguistic features of discourse episodes, these data provide convincing evidence for the idea that the *verbal* properties of discourse facilitate learning. However, if toddlers' learning environment is as rich in its multisensory properties as many have suggested (see Estigarribia & Clark, 2007; Frank et al., 2013; Gogate et al., 2000; Yu & Smith, 2012), and the speech toddlers' hear is intricately tied to on-going nonverbal activity (Adamson & Bakeman, 2006), then extended episodes of verbal discourse likely co-occur with, and may even be driven by, extended episodes of sustained nonverbal activity. If so, it is possible that extended *verbal* discourse may also facilitate development through its underlying *nonverbal* components.

In the current study, we tested the hypothesis that episodes of extended discourse possess a number of nonverbal features relevant for learning. Based on previous research, we considered three types of nonverbal events. First, we considered toddlers' manual actions because research suggests that toddlers' object name learning is enhanced when parents name the objects with which their toddlers are manually engaged (Yu & Smith, 2012; see also Scofield, Hernandez-Reif, & Keith, 2009). Second, we considered parents' manual actions based on Gogate's and others' work showing how parents' sensorimotor behaviors as they named objects is associated with better learning (for reviews see Gogate & Hollich, 2010; Gogate, Walker-Andrews, & Bahrick, 2001). Finally, based on a series of recent studies showing that toddlers' *visual* experiences with objects also shapes learning, we considered toddlers' egocentric object views during play. In these recent studies, Smith, Yu, and their colleagues analyzed recordings from toddler-worn head cameras as toddlers played with novel objects with their parents (for review see Smith, Yu, Yoshida, & Fausey, 2015). They observed that when the objects parents named dominated toddlers' fields of view—by being both larger in image size and more centered in view than competitor objects (see Figure 1), those objects were more likely to be learned (Pereira, Smith, & Yu, 2014; Yu & Smith, 2012).

The specific goals of the current study were three-fold. First, we investigated the multisensory nature of episodes of extended verbal discourse, asking whether these episodes known to be rich in verbal properties were also episodes rich in nonverbal properties. Such a finding would have implications for our understanding of the mechanisms by which extended discourse influences learning. Second, we explored how analyzing the multisensory nature of extended discourse, as opposed to of individual utterances, could expand our understanding of the nature of early multisensory communication. Specifically, because episodes of discourse span multiple utterances and the silent gaps between them, a discourse-level analysis could speak to the multisensory coupling within discourse. If the nonverbal events in discourse occur primarily during the spoken utterances, and not the silent gaps between utterances, then such a finding would suggest highly synchronous

multisensory coupling (Gogate et al., 2001; Meyer, Hard, Brand, McGarvey, & Baldwin, 2011; Zukow-Goldring, 2001). If, however, the nonverbal events occur during both the speech and nonspeech segments, then such a finding would suggest a different conceptualization of how toddlers' verbal and nonverbal experiences are coupled. Our final goal was to examine whether the degree to which parents' speech is multisensory is correlated with the degree to which parents' speech is part of an extended discourse. Such a finding would suggest not only that extended discourses are multisensory but that they are more multisensory than brief discourses. Furthermore, such a finding would highlight the possibility that the multisensory aspect of speech influences its extended nature, and/or vice versa.

To address these goals, we observed parents and their toddlers engaged in object play, a context previously demonstrated to elicit both multisensory talk (e.g., Clark & Estigarribia, 2011; Gogate et al., 2000; Yu & Smith, 2012) and extended verbal discourse (e.g., Frank et al., 2013). During play, toddlers wore head cameras, allowing us to capture toddlers' visual experiences with objects. Additionally, we coded moment-by-moment parents' manual actions and toddlers' haptic exploration during play. Finally, we transcribed and coded parents' speech in detail, allowing us to re-construct the episodes of verbal discourse. The analyses we present focus primarily on the inter-relations between the multisensory and temporally-extended nature of parents' speech.

Methods

Participants

The present analyses were conducted on a corpus of audio-visual recordings of parents and their toddlers engaged in object play ($N = 100$; $M_{\text{toddler-age}} = 18.3$ mos; $SD = 4.3$; *Age Range*: 12.2–26.0; 42 girls, 58 boys; 85 mothers, 15 fathers). Analyses on a portion of these recordings have been reported previously (Pereira et al., 2014; Yu & Smith, 2013, in press), though all published reports differ from the current paper in their theoretical and empirical goals. Although this corpus combines data collected using slightly different recording equipment and stimuli (we describe these differences in the following sections), all observations were identical in three critical ways: (a) all recordings were of parent-toddler dyads engaged in brief trials of free play with a set of three novel objects; (b) all recordings were conducted in a laboratory setting; and (c) all recordings included toddler head-camera videos for capturing toddlers' egocentric views, third-person videos for coding toddler and parent manual actions, and complete audio recordings of parents' speech.

Study environment

Figure 2 depicts the experimental set-up. Toddlers sat in a chair at a table (61 cm × 91 cm × 51 cm) across from their parents who sat on floor cushions. The room's floor and floor-to-ceiling curtains were all white, and toddlers and their parents wore white smocks. This all-white set-up assisted the computer recognition of objects in the toddlers' head-camera images (see the Coding and data processing section).

Head camera and recording devices

During play, toddlers wore headgear (either a sports headband or an elastic cap) that was fixed with a small lightweight head camera (see Figure 2). For fifty-nine toddlers, the head camera was from Positive Science (see Franchak, Kretch, Soska, & Adolph, 2011), with a diagonal field of view of 100°. For the other 41 toddlers, the head camera was a KPC-VSN500 square camera with a diagonal field of view of 90°. Because estimates of visual dominance—our key vision measure (see the Coding of nonverbal properties section)—from the two head cameras were the same,[1] we merged the data collected from the two head cameras. Parents also wore headgear on which was mounted a hands-free professional-quality microphone, the ATM75 Cardioid Condenser Microphone from Audio-Technica. Play sessions were also captured through a bird's eye view camera and wall-mounted cameras behind toddlers' and parents' right shoulders (see Figure 2).

Stimuli

All dyads played with six novel objects that were organized into two sets of three. Objects were constructed in the lab to have simple shapes and a single main color to assist computer recognition (see the Coding and data processing section). The objects' sizes were comparable (approximately 270 cm^3; ranged from 9 × 6 × 4 cm to 10 × 6 × 5.5 cm) and allowed for toddler's grasping, picking up, and playing. Each object was paired with a novel word that was disyllabic and that adhered to the phonotactic constraints of English (e.g., "*habble*," "*wawa*," "*mapoo*").

Procedure

Prior to the session, we taught parents the labels for each of the objects. We instructed parents to play with their toddlers as they normally would and to use the names when talking about the objects. During the experiment, laminated cards listing the word-object pairings were taped to the parents' side of the table (out of toddlers' views). No further instructions about how parents should interact were provided as our goal was to encourage as natural as possible of a free-flowing play session in which parents and their toddlers interacted with the objects, and parents talked about those objects as they normally would during play.

Once parents and toddlers put on their smocks and were fitted with the recording equipment, an experimenter put one set of three objects on the table and the play session began. Across the studies aggregated in this corpus, the number of trials completed and the precise trial durations differed. For one of the studies (n = 23), dyads completed two trials lasting approximately 2 minutes long. For the remaining dyads (n = 77), dyads completed up to four trials lasting approximately 1.5 minutes long. For each trial across all dyads, parents and toddlers played with one of two object sets. We swapped object sets after each trial to keep toddlers engaged. If toddlers became fussy before the trial ended, we ended the trial early. We included only trials that lasted approximately 1 minute long (the shortest trial was 50 seconds). Five toddlers did not complete the maximum number of trials (2 or 4). On average

[1]Mean proportion of time objects were visually dominant for Cameras 1 (M = .18, SD = .06) and 2 (M = .19, SD = .09) were not significantly different from each other, p = .63.

across all dyads, the play sessions lasted a total of 5.3 minutes (*SD* = 1.3) across an average of 3.5 trials (*SD* = 0.9).

Coding and data processing

Speech transcription—Parents' speech during each trial was fully transcribed and divided into utterances, defined as strings of speech between two periods of silence lasting at least 400 ms (Pereira et al., 2014; Yu & Smith, 2012). Utterances that contained reference to one of the objects were marked as "referential utterances". These included utterances when parents named an object (e.g., "that's a *habble*"), employed a pronoun referring to an object (e.g., "can you push *it*?"), or used an alternate concrete noun referring to an object (e.g., "don't throw *the toy*"). For each referential utterance, trained coders annotated the intended referent object by watching the video. On average, parents produced 19.5 utterances per minute (*SD* = 4.1), 11.1 of which were referential (*SD* = 3.7).

Head camera image processing—We sampled toddlers' head cameras at a rate of 10 frames per second. Using an in-house automated computer vision algorithm, we derived estimates for the image size of each object in toddlers' fields of view. Briefly, the program accomplishes this by first separating out nonwhite object pixels from the white background. These object pixels are then merged into object blobs based on color similarity. Finally, each object blob is given an object label based on its color. Early tests comparing the vision algorithm's ability to detect objects to that of human coders yielded high agreement (91–95% agreement, Yu, Smith, Shen, Pereira, & Smith, 2009; Smith, Yu, & Pereira, 2011; see Yu et al., 2009; for further technical details on the program). The number of pixels each object occupies for each frame is logged (see Figure 3).

Coding of nonverbal properties—For each session, we coded frame-by-frame three nonverbal properties of interest: (a) how visually dominant objects were in toddlers' fields of view, (b) which objects parents touched, and (c) which objects toddlers touched.

We determined visual dominance through the automated vision algorithm. We first determined the percentage of pixels of toddlers' fields of view that was taken up by each object. We then considered an object to be visually dominant if that object occupied more than 5% of the toddlers' field of view (see Figure 3; see also Yu & Smith, 2012).[2] Trained coders watched the entire session from multiple angles (head camera, third-person view cameras) and annotated frame-by-frame when parents and their toddlers touched each of the objects. Figure 4 depicts a representative time series of these nonverbal properties, along with referential utterances, over the course of a trial.

Reliability coding—We assessed reliability by having a second coder independently code a random selection of 25% of the participants. Reliability of each manually coded variable was determined by the Cohen's kappa (κ) statistic (*Reference Coding: .85; Toddler Touch: .93; Parent Touch: .94*). Reliability was high based on conventional guidelines (see Bakeman & Gottman, 1997).

[2]We treated visual dominance as a binary variable to match the other nonverbal variables (parent touch and toddler touch). All current results hold when different threshold of visual dominance (3% or 7% of Field of View) were employed.

Results

In what follows, we address three issues. First, we examined the multisensory nature of extended episodes of verbal discourse by analyzing the nonverbal signals to reference during these episodes. Second, we investigated in more depth the nature of the multisensory signal by analyzing the synchrony between verbal and nonverbal events within extended discourse episodes. Finally, we asked whether the multisensory nature of verbal discourse is associated with discourse length by comparing the nonverbal properties of *extended* discourse episodes to those of *brief* discourse episodes.

Multisensory nature of extended discourse

We identified episodes of extended verbal discourse by first clustering adjacent referential utterances to the same object into episodes of discourse, and then classifying those episodes as either extended or brief. Specifically, from the stream of referential utterances parents produced (see Figure 5), we marked consecutive utterances to the same object as an episode of discourse about that object (Frank et al., 2013; Messer, 1980). On average, adjacent utterances to the same object were separated by 3.3 s (SD = 2.4). A discourse episode's onset was defined as the onset of the first referential utterance to that object; its offset was defined as the offset of the last referential utterance to that object. Utterances that referred to multiple objects (M = 1.4 utterances per minute, SD = 1.3) were not counted as part of a discourse episode and would be considered to have terminated the discourse sequence. We allowed for nonreferential utterances to occur within a discourse episode as to maximize the difference between the current discourse-level analysis and previous utterance-level analyses of child-directed speech.

As Figure 5 illustrates, discourse episodes varied in both their duration and run length (i.e., the number of utterances within an episode). For each dyad, we computed the median episode duration (M = 5.0 s; SD = 3.0 s) and run length (M = 2.0 utterances; SD = .75), and classified episodes as "extended discourse" if that episode was above the median in both duration and run length. All other episodes were considered to be episodes of "brief discourse."[3] Table 1 provides descriptive statistics of the verbal properties of discourse episodes.

Nonverbal properties of extended discourse utterances—Of particular interest were the *nonverbal* correlates to extended verbal discourse and how the co-occurrence between these properties produces highly multisensory utterances. Table 2 provides descriptive statistics of the three nonverbal properties of interest during extended discourse. These descriptive data paint a clear picture: episodes of extended discourse are dense and rich not only in their *verbal* information, but also in their *nonverbal* signals. Figure 6 illustrates the convergence of verbal and nonverbal properties. The figure depicts the proportion of utterances within episodes of extended discourse that overlapped with nonverbal events on the talked about object. We considered utterances to overlap with nonverbal events so long as there was any overlap between the utterance and the nonverbal

[3]Other reasonable methods for distinguishing extended and brief discourse revealed similar trends (see Suanda, Smith, & Yu, 2016).

events. To examine whether these proportions were significantly different from what would be expected by chance, we constructed randomly sampled segments from the interaction (restricted to times within trials) that matched each extended discourse utterance in duration and target object. We then computed the proportion of those segments that overlapped with nonverbal events directed at the target. Finally, we compared the occurrence rate of nonverbal events in the observed episodes of extended discourse to the occurrence rate in the randomly sampled segments. To ensure the robustness of this comparison, we repeated the sampling process 1000 times and computed the mean occurrence of nonverbal events across all iterations. As Figure 6 illustrates, parents' speech in extended discourse co-occurred with each nonverbal property at a rate much greater than would be expected by chance (as indexed by the pseudo-sample simulations)[4]: visual dominance: $t(99) = 6.34$, $p < .001$, *Cohen's d* = 1.49, *Mean Odds Ratio* (*OR*) = 1.50 (1.43–1.58)[5]; visible parent action: $t(99) = 15.27$, $p < .001$, $d = 1.77$, $OR = 1.99$ (1.87–2.12); haptic exploration: $t(99) = 10.89$, $p < .001$, $d = 1.40$, $OR = 1.58$ (1.48–1.67).

Of course, the co-occurrences between speech and the different nonverbal properties were not independent of one another. For example, in some cases as parents talked, toddlers and parents might have been jointly manipulating an object, providing both a visual (i.e., visible parent action) and a haptic signal (i.e., toddler manual object exploration) of the referent. In other cases, toddlers might be manually exploring an object close to their body, also creating an experience with that object that spans multiple modalities. Thus, we examined the extent to which parents' speech in extended discourse co-occurs with at least one nonverbal modality (i.e., *bimodal* references) and with multiple nonverbal modalities (i.e., *multimodal* references).

As highlighted in Figure 6D and 6E, the occurrence of bimodal and multimodal utterances in extended discourse were far greater than what we would expect given base-rate levels, $t_{\text{bimodal}}(99) = 24.55$, $p < .001$, $d = 2.06$, $OR = 1.40$ (1.36–1.45), $t_{multimodal}(99) = 11.51$, $p < .001$, $d = 1.69$, $OR = 1.96$ (1.84–2.08). These results highlight two additional points about utterances in extended discourse. First, nearly all utterances (94.0%, $SD = 7.9\%$) co-occurred with at least one nonverbal event. Second, many utterances in extended discourse (43.3%, $SD = 17.0\%$) co-occurred not only with one but with multiple nonverbal properties, providing toddlers with redundant cues to their parents' intended referents.

How verbal and nonverbal events couple

Many previous analyses of parents' naming behavior have revealed that verbal and nonverbal properties of parent–toddler interactions are inter-connected (Brand, Baldwin, & Ashburn, 2002; Frank et al., 2013; Gogate et al., 2000; Meyer et al., 2011; Rohlfing, Fritsch, Wrede, & Jungmann, 2006; Yu & Smith, 2012). Most previous studies however have analyzed parents' speech at the level of individual utterances (see Clark & Estigarribia, 2011; Frank et al., 2013; Meyer et al., 2011; Rohde & Frank, 2014; for notable exceptions).

4. Simulations were performed in MATLAB.
5. These Odds Ratios are to be interpreted as the increase in likelihood that a nonverbal event occurred during an utterance within extended discourse over what would be expected by chance. We present the mean Odds Ratios across subjects along with the 95% confidence intervals around those means.

An analysis of parents' speech at the level of discourse may provide unique insight. In the context of understanding multisensory communication, a discourse-level analysis sheds light on the nature of the coupling between verbal and nonverbal events. That is, are the multisensory patterns we observed primarily driven by synchronous verbal and nonverbal events (see Figure 7A), suggestive of highly attentive parents who time-locks each utterance to nonverbal events (Gogate et al., 2000; Meyer et al., 2011; Zukow-Goldring, 1996)? Or do the multisensory patterns reflect a broader sense of coupling between verbal and nonverbal events (see Figure 7B), with some utterances more synchronous with nonverbal events and other utterances less synchronous with nonverbal events (see Meyer et al., 2011)? To understand the coupling between parents' speech and the on-going nonverbal events, we compared, at the frame level, the amount of overlap between nonverbal properties and *speech segments* (portions of discourse in which parents were talking about the object; see Figure 8) and *nonspeech segments* (portions of discourse in which parents were silent).

Figure 8 shows the mean proportion of time in speech and nonspeech segments that overlapped with nonverbal events. As the figure highlights, both speech and nonspeech segments were largely similar in their overlap with the three nonverbal properties. The occurrence of each of the nonverbal properties were well above what would be expected by chance[6] for both speech-segments $t_{\text{vis-dom}}$ (99) = 4.15, $p < .001$, $d = .96$, $OR = 1.47$ (1.33–1.60), $t_{\text{vis-action}}$ (99) = 13.90 $p < .001$, $d = 1.81$, $OR = 2.17$ (1.91–2.44), $t_{\text{haptic}}(99) = 8.12$, $p < .001$, $d = 1.11$, $OR = 1.61$ (1.45–1.78), and nonspeech segments, $t_{\text{vis-dom}}(99) = 5.56$, $p < .001$, $d = 1.19$, $OR = 1.62$ (1.46–1.78), $t_{\text{vis-action}}(99) = 9.42$, $p < .001$, $d = 1.30$, $OR = 1.84$ (1.65–2.03), $t_{\text{haptic}}(99) = 8.95$, $p < .001$, $d = 1.26$, $OR = 1.70$ (1.51–1.89). We interpret these results as consistent with the broad notion of coupling between the verbal and nonverbal properties of interactions. It is likely that some of parents' utterances, as many have convincingly demonstrated (Gogate et al., 2000; Gogate, Maganti, & Bahrick, 2015; Meyer et al., 2011; Zukow-Goldring, 1996), were tightly coupled to nonverbal behaviors. In fact, as illustrated in the example in Figure 9(B), we did find a reliable difference in how prevalent parents' manual actions were during speech and nonspeech segments, $t(99) = 6.50$, $p < .001$, $d = .65$, $OR = 1.22$ (1.11–1.32). It is also likely that many other utterances however were not particularly well-timed with nonverbal behaviors. Interestingly, for example, we observed that object visual dominance (Figure 9A) and toddlers' haptic exploration (Figure 9C) appeared to more reliably occur when during nonspeech segments, $t_{\text{vis-dom}}(99) = 3.68$, $p < .001$, $d = .37$, $OR = 1.21$ (1.00–1.42), $t_{\text{haptic}}(99) = 3.25$, $p < .001$, $d = .33$, $OR = 1.06$ (1.00–1.11).

Linking the multisensory and discursive properties of parent speech

Thus far, we have shown how jointly considering the multisensory and extended nature of parents' discourse expands our understanding of these two properties. Herein, we ask whether these two aspects of parents' speech are actually correlated with each other. That is, is speech that is part of an extended discourse actually more multisensory than speech that is

6. In deriving chance-level overlap, we constructed randomly sampled segments of the entire interaction that matched each extended discourse in duration and target object, computed the proportion of time within those samples that overlapped with the nonverbal events, and then repeated that process across the 1000 iterations. Chance level overlap was the mean proportion of overlap across all iterations.

part of a brief discourse? If it is, then this would potentially suggest that one may actually influence the other. For example, perhaps parents' speech that is more multisensory, which could both be the result of parents following toddlers' visual and manual attention (Tomasello & Farrar, 1986) and the result of speech further attracting toddlers' attention (Baldwin & Markman, 1989; Gogate, Bolzani, & Betancourt, 2006; Rader & Zukow-Goldring, 2012, 2015), is more conducive to establishing sustained discourse about an object. If instead speech in extended discourse is no more multisensory than speech in brief discourse, then this would suggest that the multisensory patterns we observed simply reflect the fact that all of parents' speech is highly multisensory, regardless of discourse length. Thus, we compared the nonverbal properties of referential utterances that were part of an extended discourse with the nonverbal properties of referential utterances that were part of a *brief* discourse (see Figure 10). Of interest was whether there was anything unique about the multisensory nature of extended discourse.

Figure 10B–D illustrates the differences between the nonverbal properties during extended discourse to those during brief discourse. Parents' speech in extended discourses was more likely to co-occur with toddlers' haptic exploration of the referent, $t(99) = 2.71$, $p < .01$, $d = .27$, $OR = 1.27$ (1.11–1.44), and with visual dominance of that referent, $t(99) = 2.25$, $p < .05$,[7] $d = .22$, $OR = 1.20$ (1.05–1.36); there were no differences between parent actions across extended and brief discourse, $p = .37$. When we looked beyond the occurrence of individual modalities and examined the co-occurrence of multiple modalities with parents' speech, we found that parents' speech in extended discourse ($M = 43.2\%$, $SD = 17.0$) was more likely to co-occur with multiple modalities than parents' speech in brief discourse ($M = 37.5\%$, $SD = 16.1$), $t(99) = 2.95$, $p < .01$, $d = .30$, $OR = 1.40$ (1.19–1.61); there were no differences in the percentage of utterances co-occurring with at least one nonverbal modality ($M_{ext.} = 94.1\%$, $SD_{ext.} = 7.8$; $M_{brief} = 94.3\%$, $SD_{brief} = 6.1$), $p = .87$. In sum, these findings demonstrate that there is indeed a correlation between the duration of a discourse episode and the nonverbal properties of the utterances within that episode.

General discussion

Toddlers learn object names from sensory rich contexts: they hear words, they touch and look at objects, and their caregivers produce visible gestures and social signals. Fortunately for toddlers, many of these nonverbal events are tightly aligned with the topics of parents' speech (Frank et al., 2013; Gogate & Hollich, 2010; Messer, 1983; Zukow-Goldring, 1990). Thus, this sensory rich nature of toddlers' experience is a virtue, not a nuisance, for learning (see Gogate et al., 2001; Yu & Smith, 2012). In the current study, we investigated how the multisensory facet of toddlers' learning environment relates to a different, but no less pervasive, facet of the environment: the extended structure of verbal discourse. Our results demonstrate inter-relations between these dimensions that have largely been studied independently, providing novel insights into the toddlers' learning environment and how the environment could support learning. Herein, we discuss the implications of these data,

[7]The fact that the results for visual dominance mirror the results for toddlers' manual actions is likely due to the fact that visual dominance is more tightly coupled to toddlers' manual actions (see Yu & Smith, 2012).

highlighting the value of jointly considering multiple aspects of toddlers' learning environment.

Role of extended discourse in lexical development

The organization of parents' speech into discourse episodes is a noticeable feature of child-directed speech (Frank et al., 2013; Messer, 1980). Much empirical and theoretical work has highlighted how this prominent feature of toddlers' learning experience could support language development. For example, existing work has shown that the repetition of utterance properties (e.g., words, sentence structures) common within an episode, could aid speech perception (Bard & Anderson, 1983), word segmentation (Onnis et al., 2008), and syntax acquisition (Hoff-Ginsberg, 1986). Additionally, the rich verbal descriptions that often follow the introduction of new object names within discourse episodes (see Clark, 2010) could help toddlers learn the deeper meanings of new words (see also Sullivan & Barner, 2016). All of these studies point to the idea that a candidate mechanism by which extended discourse could facilitate learning is through its *verbal* features.

By going beyond an analysis of the verbal properties of extended discourse and analyzing their nonverbal correlates, we suggest a different pathway through which these extended episodes would benefit learners. We observed that these episodes possessed several rich nonverbal features previously demonstrated by observational and experimental studies to facilitate toddlers' object name learning. Specifically, we found that inside these episodes, parents' object references co-occurred with moments when those objects were visually salient (Axelsson, Churchley, & Horst, 2012; Pereira et al., 2014; Yu & Smith, 2012), moments when toddlers were manually engaged with those objects (Pereira, Smith, & Yu, 2008; Scofield et al., 2009; Yu & Smith, 2012), and moments when parents produced visible gestures and actions with those objects (Booth, McGregor, & Rohlfing, 2008; Gogate et al., 2000; Rader & Zukow-Goldring, 2012). If extended discourse episodes are packed with multiple converging nonverbal events known to be associated with better learning, then extended episodes of *verbal* discourse may also facilitate language learning in part through their *nonverbal* correlates. Future research that investigates how the constellation of verbal and nonverbal aspects of discourse is related to vocabulary growth, how the relevant factors may shift with age, and whether there are individual differences in which factors matter will go a long way in bettering our understanding of how the language learning environment shapes language learning trajectories.

Nature of multisensory communication

The analysis of multisensory events over the course of a discourse episode, as opposed to over individual utterances, also provides new insights into the real-time dynamics of multisensory communication, and its potential role for learning. Although we found that nonverbal events did tend to occur when parents spoke, creating multisensory events (see also Clark & Estigarribia, 2011; Gogate et al., 2000; Meyer et al., 2011; Zukow-Goldring, 1996), we also found that within an extended discourse episode, nonverbal events occurred when parents did not speak. We suggest that this is in part due to the different time scales at which spoken utterances and nonverbal events occurred. For example, whereas on average parents' utterances lasted 1.5 seconds long, toddlers' holding events lasted 5.5 seconds long.

Thus, when parents' speech overlapped with toddlers' holding, the holding event likely began well before parents spoke and continued well after. Although our analyses do not speak to the implication of these dynamics for learning, they do raise a question about how multisensory events shape learning: is it via the moments when speech and nonverbal events overlap, via the sustained nonverbal activity enveloping parents' speech, or both? The current results also highlight the value of analyzing a phenomenon (in this case the coupling of verbal and nonverbal processes) at multiple time-scales of analysis (see also Meyer et al., 2011; Rohde & Frank, 2014). Had we simply considered utterance-level data, our findings could be interpreted as consistent with the notion of synchronous coupling between verbal and nonverbal events. Only by going beyond the window of individual utterances were we able to conclude that the data were more consistent with a broader notion of coupling that goes beyond synchrony between verbal and nonverbal events.

Linking extended discourse and multisensory communication

The current findings go beyond demonstrating that parents' speech is highly multisensory. The current data demonstrates a link between the degree to which parents' speech is multisensory with the degree to which parents' speech is part of an extended episode of verbal discourse. If these multi-utterance conversations are rich in their quantity and quality of linguistic input, as previous analyses of parent–toddler discourse have suggested (e.g., Messer, 1980), then our results highlight just how inter-connected the different dimensions of the learning environment actually are: multisensory utterances, which may have their own virtues for learning (see Gogate et al., 2001; Yu & Smith, 2012), also may help establish conversations with other properties that facilitate learning, including repetitions of object names (Messer, 1980), variation sets (Onnis et al., 2008), aligned syntactic structures (Hoff-Ginsberg, 1986), and rich semantic networks (Clark, 2010). In other words, multisensory utterances may not only have direct effects on learning via their inter-sensory redundancy, they may also have cascading indirect effects on learning via the co-occurring linguistic variables.

Structure in the language learning environment: Implications for developmental neuropsychology

Recent theoretical perspectives on brain development posit that a better understanding of the nature and structure of the learning environment is highly relevant to understanding the processes of brain network development (Byrge, Sporns, & Smith, 2014; Chiel & Beer, 1997; Engel, Maye, Kurthen, & Konig, 2013). This is because under these accounts, the brain, body, and environment act as a coupled system whose components continuously influence and are influenced by each other. Thus, investigations into the structure of the learning environment provide insight into the input statistics that mold brain networks (Byrge et al., 2014), and represent part of the puzzle of understanding how the brain-body-behavior system shapes learning (see also Johnson, 2011). Data on the developing brain of children who experience impoverished (e.g., Hackman & Farah, 2009) or deprived learning environments (e.g., Chugani et al., 2001), as well as research on how environmental factors shape learning in populations with atypical neurodevelopment (e.g., Rowe, Levine, Fisher, & Goldin-Meadow, 2009) further point to the value of investigations of the structure of the learning environment for developmental neuropsychology.

With this broad perspective in mind, the current depiction of the toddlers' language learning environment makes contact with recent advancements in the study of brain development in three ways. First, the current data underscores how toddlers' learning experiences traverse multiple modalities (auditory, visual, and tactile) and are tightly bound to the behavior of their social partners. Relatedly, the last 20 years of research into the neural underpinnings of cognitive competencies have highlighted that in contrast to viewing cognition as resulting from the operations of single brain areas, from unisensory neural building blocks, and from processes within independent individuals, cognition is better understood as emerging from the dynamic connections between many areas within large-scale networks (Bressler & Menon, 2010; Sporns, 2011; Sporns, Chialvo, Kaiser, & Hilgetag, 2004), from neural mechanisms that are multisensory through and through (Ghazanfar & Schroeder, 2006; Hyde, Jones, Flom, & Porter, 2011; Reynolds, Bahrick, Lickliter, & Guy, 2014), and from neural processes that are coupled to the neural processes of social partners (Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2011). Consistent with the brain-body-environment framework, multisensory interactive discourse, like that observed in the current study, may thus help shape and be shaped by large-scale multisensory coupled brain networks.

Second, the current data also depict a learning environment that is filled with redundancies and inter-connections. This finding is consistent with a perspective of language development as a process deeply tied to other facets of development, including cognitive development (Smith, 2013), perceptuo-motor development (e.g., Iverson, 2010), and social development (Tomasello, 2003). Mounting evidence from developmental cognitive neuroscience (e.g., Borgstrom, Von Koss Torkildsen, & Lindgren, 2015; Junge, Cutler, & Hagoort, 2012; for review, see Kuhl, 2010), developmental neuropsychology (e.g., D'Souza & Karmiloff-Smith, 2011) and neurocomputational modeling (Mayor & Plunkett, 2010) also supports linkages between the development of language and other domains.

Finally, the current study highlights that language learners are exposed to multisensory regularities that span multiple time scales. Although the existence of regularities at different time-scales is widely recognized, precisely how multisensory information along those time-scales is encoded, processed and used in the service of language learning is not well understood (see also Rohde & Frank, 2014). Recent theoretical and empirical developments in the cognitive neuroscience of human memory suggest that in processing everyday stimuli, like language, the brain relies on a hierarchy of networks, with each network encoding and integrating information at increasingly longer time-scales (for review, see Hasson, Chen, & Honey, 2015). Importantly, activation of the functional networks specializing in longer time-scales is associated with greater comprehension and retention of information (Hasson, Nusbaum, & Small, 2007; Yarkoni, Speer, & Zacks, 2008). Although much of this data come from neuroimaging studies with adults, they nonetheless point to candidate neural mechanisms through which information at longer time-scales (e.g., the discourse time-scale) could influence learning (see also Christiansen & Chater, 2016).

Future work

In the current study, we investigated the multisensory nature of verbal discourse in parent-toddler interactions. One unique contribution of the current investigation is its analysis of the visual properties available to toddlers during discourse as measured via small toddler-worn headcameras. The advantage of this approach is that it offers a window into the language learning environment from the toddler learners' perspective, as opposed to the more commonly employed third-person perspective of the learning environment (for an analysis and discussion of the importance of understanding the learners' perspective, see Yoshida & Smith, 2008; Yurovsky, Smith, & Yu, 2013). A limitation of this approach is that observations obtained from headcameras only reflect the information that is available to toddlers as opposed to the information to which toddlers are actually attending (for discussion, see Smith et al., 2015). Future research that measures toddlers' eye-gaze patterns (e.g., via head-mounted eye-tracking, see Franchak et al., 2011) may speak more directly to how toddlers' visual attention waxes and wanes during discourse, as well as how it shapes parents' verbal discourse. A second limitation of the current study is that it focuses on the nature of toddlers' learning environment and not on toddlers' learning itself. Thus, although our results suggests a powerful learning environment that is sensory-rich with many redundant cues for learning, the extent to which these cues support toddlers' learning and development remains to be seen. Finally, like many studies in the developmental sciences, the current study relied on a sample of convenience (for discussion on the implications of different sampling methods in developmental psychology, see Bornstein, Jager, & Putnick, 2013). Future research that tests whether the current findings generalize to the broader population (including, but not limited to, diverse socio-demographic subgroups and different language learning experiences) is important.

Conclusion

The notion that parents' speech to their language learning toddlers is highly multisensory is neither novel nor controversial. Additionally, the notion that the multisensory facet of parents' speech facilitates toddlers' development is gaining traction. The multisensory nature of parents' speech however is only one of many dimensions of toddlers' experience that supports acquisition. We suggest, and believe to have demonstrated herein, that one important way forward is to investigate how different dimensions of the environment intertwine. The promise of such a multi-dimensional approach is a more complete understanding of how the toddlers' environment and experiences shape learning and of how best to intervene when learning goes awry.

Acknowledgments

We thank many members of the Computational Cognition and Learning Laboratory, especially Maddie Bruce, Danielle Rosenstein, and Jessica Steinhiser, for their assistance in this research.

Funding

This research was supported in part by the National Science Foundation (SBE-BCS0924248, SBE-SMA1203420) and the National Institutes of Health (R01-HD074601, R21-EY017843, K99-HD082358).

References

Adamson LB, & Bakeman R (2006). Development of displaced speech in early mother-child conversations. Child Development, 77, 186–200. doi:10.1111/cdev.2006.77.issue-1 [PubMed: 16460533]

Axelsson EL, Churchley K, & Horst JS (2012). The right thing at the right time: Why ostensive naming facilitates word learning. Frontiers in Psychology, 3, 88. doi:10.3389/fpsyg.2012.00088 [PubMed: 22470363]

Bakeman R, & Gottman JM (1997). Observing interaction: An introduction to sequential analysis. Cambridge, UK: Cambridge University Press.

Baldwin DA, & Markman EM (1989). Establishing word-object relations: A first step. Child Development, 60, 381–398. [PubMed: 2924658]

Bard EG, & Anderson AH (1983). The unintelligibility of speech to children. Journal of Child Language, 10, 265–292. doi:10.1017/S0305000900007777 [PubMed: 6874768]

Booth AE, McGregor KK, & Rohlfing KJ (2008). Socio-pragmatics and attention: Contributions to gesturally guided word learning in toddlers. Language, Learning, and Development, 4, 179–202. doi:10.1080/15475440802143091

Borgstrom K, Von Koss Torkildsen J, & Lindgren M (2015). Event-related potentials during word mapping to object shape predict toddlers' vocabulary. Frontiers in Psychology, 6, 143. [PubMed: 25762957]

Bornstein MH, Jager J, & Putnick DL (2013). Sampling in developmental science: Situations, shortcomings, solutions, and standards. Developmental Review, 33, 357–370. doi:10.1016/j.dr.2013.08.003 [PubMed: 25580049]

Brand R, Baldwin DA, & Ashburn LA (2002). Evidence for 'motioinese': Modifications in mothers' infant directed action. Developmental Science, 5, 72–83. doi:10.1111/1467-7687.00211

Bressler SL, & Menon V (2010). Large-scale brain networks in cognition: Emerging methods and principles. Trends in Cognitive Sciences, 14, 277–290. doi:10.1016/j.tics.2010.04.004 [PubMed: 20493761]

Byrge L, Sporns O, & Smith LB (2014). Developmental process emerges from extended brain-body-behavior networks. Trends in Cognitive Sciences, 18, 395–403. doi:10.1016/j.tics.2014.04.010 [PubMed: 24862251]

Chiel HJ, & Beer RD (1997). The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment. Trends in Neuroscience, 20, 553–557. doi:10.1016/S0166-2236(97)01149-1

Christiansen MH, & Chater N (2016). The now-or-never bottleneck: A fundamental constraint on language. Behavioral and Brain Sciences, 39, e62. [PubMed: 25869618]

Chugani HT, Behen ME, Muzik O, Juhasz C, Nagy F, & Chugani DC (2001). Local brain functional activity following early deprivation: A Study of postinstitutionalized Romanian orphans. NeuroImage, 14, 1290–1301. doi:10.1006/nimg.2001.0917 [PubMed: 11707085]

Clark EV (2010). Adult offer, word-class, and child uptake in early lexical acquisition. First Language, 30, 250–269. doi:10.1177/0142723710370537

Clark EV, & Estigarribia B (2011). Using speech and gesture to introduce new objects to young children. Gesture, 11, 1–23. doi:10.1075/gest.11.1

D'Souza D, & Karmiloff-Smith A (2011). When modularization fails to occur: A developmental perspective. Cognitive Neuropsychology, 28, 276–287. doi:10.1080/02643294.2011.614939 [PubMed: 22185238]

Engel AK, Maye A, Kurthen M, & Konig P (2013). Where's the action? The pragmatic turn in cognitive science. Trends in Cognitive Sciences, 17, 202–209. doi:10.1016/j.tics.2013.03.006 [PubMed: 23608361]

Estigarribia B, & Clark EV (2007). Getting and maintaining attention in talk to young children. Journal of Child Language, 34, 799–814. doi:10.1017/S0305000907008161 [PubMed: 18062359]

Franchak JM, Kretch KS, Soska KC, & Adolph KE (2011). Head-mounted eye tracking: A new method to describe infant looking. Child Development, 82, 1738–1750. doi:10.1111/cdev.2011.82.issue-6 [PubMed: 22023310]

Frank MC, Tenenbaum JB, & Fernald A (2013). Social and discourse contributions to the determination of reference in cross-situational word learning. Language Learning & Development, 9, 1–24. doi:10.1080/15475441.2012.707101

Ghazanfar AA, & Schroeder CE (2006). Is neocortex essentially multisensory? Trends in Cognitive Sciences, 10, 278–285. doi:10.1016/j.tics.2006.04.008 [PubMed: 16713325]

Gogate LJ, Bahrick LE, & Watson JD (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. Child Development, 71, 878–894. doi:10.1111/cdev.2000.71.issue-4 [PubMed: 11016554]

Gogate LJ, Bolzani LH, & Betancourt E (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. Infancy, 9, 259–288. doi:10.1207/s15327078in0903_1

Gogate LJ, & Hollich G (2010). Invariance detection within an interactive system: A perceptual gateway to language development. Psychological Review, 117, 496–516. doi:10.1037/a0019049 [PubMed: 20438235]

Gogate LJ, Maganti M, & Bahrick L (2015). Cross-cultural evidence for multimodal motherese: Asian-Indian mothers' adaptive use of synchronous words and gestures. Journal of Experimental Child Psychology, 129, 110–126. doi:10.1016/j.jecp.2014.09.002 [PubMed: 25285369]

Gogate LJ, Walker-Andrews A, & Bahrick LE (2001). The intersensory origins of word comprehension: An ecological-dynamic systems view. Developmental Science, 4, 1–18. doi:10.1111/desc.2001.4.issue-1

Hackman DA, & Farah MJ (2009). Socioeconomic status and the developing brain. Trends in Cognitive Sciences, 13, 65–73. doi:10.1016/j.tics.2008.11.003 [PubMed: 19135405]

Harris M, Jones D, & Grant J (1983). The nonverbal context of mothers' speech to infants. First Language, 4, 21–30. doi:10.1177/014272378300401003

Hasson U, Chen J, & Honey CJ (2015). Hierarchical process memory: Memory as an integral component of information processing. Trends in Cognitive Sciences, 19, 304–313. doi:10.1016/j.tics.2015.04.006 [PubMed: 25980649]

Hasson U, Ghazanfar AA, Galantucci B, Garrod S, & Keysers C (2011). Brain-to-brain coupling: A mechanism for creating and sharing a social world. Trends in Cognitive Sciences, 16, 114–121. doi:10.1016/j.tics.2011.12.007

Hasson U, Nusbaum HC, & Small SL (2007). Brain networks subserving the extraction of sentence information and its encoding to memory. Cerebral Cortex, 17, 2899–2913. doi:10.1093/cercor/bhm016 [PubMed: 17372276]

Hoff-Ginsberg E (1986). Function and structure in maternal speech: Their relation to the child's development of syntax. Developmental Psychology, 22, 155–163. doi:10.1037/0012-1649.22.2.155

Hyde DC, Jones BL, Flom R, & Porter CL (2011). Neural signatures of face-voice synchrony in 5-month-old human infants. Developmental Psychobiology, 53, 359–370. doi:10.1002/dev.v53.4 [PubMed: 21271561]

Iverson JM (2010). Developing language in a developing body: The relationship between motor development and language development. Journal of Child Language, 37, 229–261. doi:10.1017/S0305000909990432 [PubMed: 20096145]

Johnson MH (2011). Interactive specialization: A domain-general framework for human functional brain development? Developmental Cognitive Neuroscience, 1, 7–21. doi:10.1016/j.dcn.2010.07.003 [PubMed: 22436416]

Junge C, Cutler A, & Hagoort P (2012). Electrophysiological evidence of early word learning. Neuropsychologia, 50, 3702–3712. doi:10.1016/j.neuropsychologia.2012.10.012 [PubMed: 23108241]

Kuhl PK (2010). Brain mechanisms in early language acquisition. Neuron, 67, 713–727. doi:10.1016/j.neuron.2010.08.038 [PubMed: 20826304]

Mayor J, & Plunkett K (2010). A neurocomputational account of taxonomic responding and fast mapping in early word learning. Psychological Review, 117, 1–31. doi:10.1037/a0018130 [PubMed: 20063962]

Messer DJ (1978). The integration of mothers' referential speech with joint play. Child Development, 49, 781–787. doi:10.2307/1128248

Messer DJ (1980). The episodic structure of maternal speech to young children. Journal of Child Language, 7, 29–40. doi:10.1017/S0305000900007017 [PubMed: 7372737]

Messer DJ (1983). The redundancy between adult speech and nonverbal interaction: A contribution to acquisition? In Golinkoff RM (Ed.), The transition from prelinguistic to linguistic communication (pp. 147–165). Hillsdale, NJ: Erlbaum.

Meyer M, Hard B, Brand RJ, McGarvey M, & Baldwin DA (2011). Acoustic packaging: Maternal speech and action synchrony. IEEE Transactions on Autonomous Mental Development, 3, 154–162. doi:10.1109/TAMD.2010.2103941

Onnis L, Waterfall HR, & Edelman S (2008). Learn locally, act globally: Learning language from variation set cues. Cognition, 109, 423–430. doi:10.1016/j.cognition.2008.10.004 [PubMed: 19019350]

Pereira AF, Smith LB, & Yu C (2008). Social coordination in toddler's word learning: Interacting systems of perception and action. Connection Science, 20, 73–89. doi:10.1080/09540090802091891 [PubMed: 20953274]

Pereira AF, Smith LB, & Yu C (2014). A bottom-up view of toddler word learning. Psychonomic Bulletin & Review, 21, 178–185. doi:10.3758/s13423-013-0466-4 [PubMed: 23813190]

Rader ND, & Zukow-Goldring P (2012). Caregivers' gestures direct infant attention during early word learning: The importance of dynamic synchrony. Language Sciences, 34, 559–568. doi:10.1016/j.langsci.2012.03.011

Rader ND, & Zukow-Goldring P (2015). The role of speech-gesture synchrony in clipping words from the speech stream: Evidence from infant pupil responses. Ecological Psychology, 27, 290–299. doi:10.1080/10407413.2015.1086226

Reynolds GD, Bahrick LE, Lickliter R, & Guy MW (2014). Neural correlates of intersensory processing in five-month-old infants. Developmental Psychobiology, 56, 355–372. doi:10.1002/dev.21104 [PubMed: 23423948]

Rohde H, & Frank MC (2014). Markers of topical discourse in child-directed speech. Cognitive Science, 38, 1634–1661. doi:10.1111/cogs.2014.38.issue-8 [PubMed: 24731080]

Rohlfing KJ, Fritsch J, Wrede B, & Jungmann T (2006). How can multimodal cues from child-directed interaction reduce learning complexity in robots? Advanced Robotics, 20, 1183–1199. doi:10.1163/156855306778522532

Rowe ML, Levine SC, Fisher JA, & Goldin-Meadow S (2009). Does linguistic input play the same role in language learning for children with and without early brain injury? Developmental Psychology, 45, 90–102. doi:10.1037/a0012848 [PubMed: 19209993]

Schwab JF, & Lew-Williams C (2016). Repetition across successive sentences facilitates young children's word learning. Developmental Psychology, 52, 879–886. doi:10.1037/dev0000125 [PubMed: 27148781]

Scofield J, Hernandez-Reif M, & Keith AB (2009). Preschool children's multimodal word learning. Journal of Cognition and Development, 10, 306–333. doi:10.1080/15248370903417662

Smith LB (2013). It's all connected: Pathways in visual object recognition and early noun learning. American Psychologist, 68, 618–629. doi:10.1037/a0034185 [PubMed: 24320634]

Smith LB, Yu C, & Pereira AF (2011). Not your mother's view: The dynamics of toddler visual experience. Developmental Science, 14, 9–17. doi:10.1111/j.1467-7687.2009.00947.x [PubMed: 21159083]

Smith LB, Yu C, Yoshida H, & Fausey CM (2015). Contributions of head-mounted cameras to studying the visual environments of infants and young children. Journal of Cognition and Development, 16, 407–419. doi:10.1080/15248372.2014.933430 [PubMed: 26257584]

Sporns O (2011). The human connectome: A complex network. Annals of the New York Academy of Sciences, 1224, 109–125. doi:10.1111/j.1749-6632.2010.05888.x [PubMed: 21251014]

Sporns O, Chialvo DR, Kaiser M, & Hilgetag CC (2004). Organization, development and function of complex brain networks. Trends in Cognitive Sciences, 8, 418–425. doi:10.1016/j.tics.2004.07.008 [PubMed: 15350243]

Suanda SH, Smith LB, & Yu C (2016). More than words: The many ways extended discourse facilitates word learning In Papafragou A, Grodner D, Mirman D, & Trueswell JC (Eds.),

Proceedings of the 38th Annual Conference of the Cognitive Science Society (pp. 1835–1840). Austin, TX: Cognitive Science Society.

Sullivan J, & Barner D (2016). Discourse bootstrapping: Preschoolers use linguistic discourse to learn new words. Developmental Science, 19, 63–75. doi:10.1111/desc.12289 [PubMed: 25702754]

Tomasello M (2003). Constructing a language: A usage-based theory of language acquisition. Cambridge, MA: Harvard University Press.

Tomasello M, & Farrar MJ (1986). Joint attention and early language. Child Development, 57, 1454–1463. [PubMed: 3802971]

Vlach HA, & Johnson SP (2013). Memory constraints on infants' cross-situational statistical learning. Cognition, 127, 375–382. doi:10.1016/j.cognition.2013.02.015 [PubMed: 23545387]

Yarkoni T, Speer NK, & Zacks JM (2008). Neural substrates of narrative comprehension and memory. NeuroImage, 41, 1408–1425. doi:10.1016/j.neuroimage.2008.03.062 [PubMed: 18499478]

Yoshida H, & Smith LB (2008). What's in view for toddlers? Using a head camera to study visual experience. Infancy, 13, 229–248. doi:10.1080/15250000802004437 [PubMed: 20585411]

Yu C, & Smith LB (2012). Embodied attention and word learning by toddlers. Cognition, 125, 244–262. doi:10.1016/j.cognition.2012.06.016 [PubMed: 22878116]

Yu C, & Smith LB (2013). Joint attention without gaze following: Human infants and their parents coordinate visual attention to objects through eye-hand coordination. PLoS ONE, 8, e79659. doi:10.1371/journal.pone.0079659 [PubMed: 24236151]

Yu C, & Smith LB (in press). Multiple sensory-motor pathways lead to coordinated visual attention. Cognitive Science. doi:10.1111/cogs.12366

Yu C, Smith LB, Shen H, Pereira AF, & Smith TG (2009). Active information selection: Visual attention through the hands. IEEE Transactions on Autonomous Mental Development, 2, 141–151.

Yurovsky D, Smith LB, & Yu C (2013). Statistical word learning at scale: The baby's view is better. Developmental Science, 16, 959–966. [PubMed: 24118720]

Zukow-Goldring P (1990). Socio-perceptual bases for the emergence of language: An Alternative to innatist approaches. Developmental Psychobiology, 23, 705–726. doi:10.1002/(ISSN)1098-2302 [PubMed: 2286299]

Zukow-Goldring P (1996). Sensitive caregivers foster the comprehension of speech: When gestures speak louder than words. Early Development and Parenting, 5, 195–211. doi:10.1002/(SICI)1099-0917(199612)5:4<195::AID-EDP133>3.0.CO;2-H

Zukow-Goldring P (2001). Perceiving referring actions. Developmental Science, 4, 28–30.

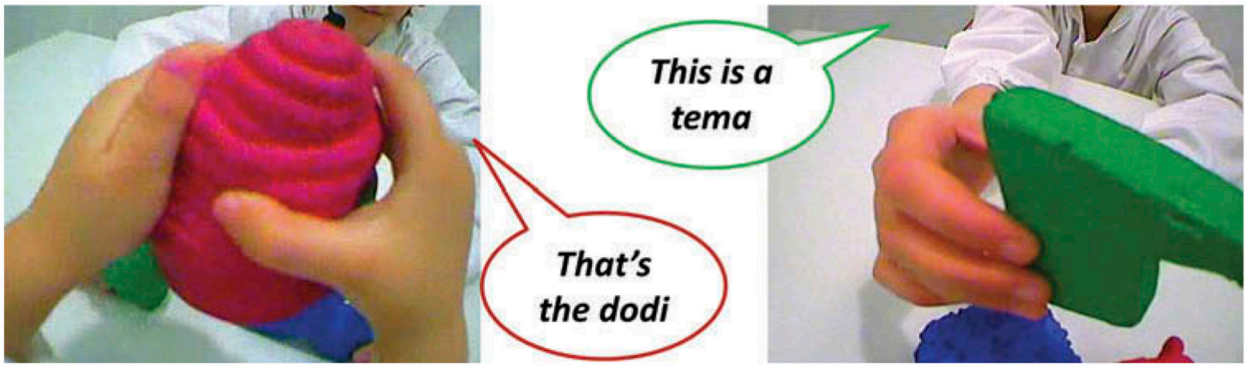

Figure 1.
Example image frames from toddler egocentric cameras found to be associated with object name learning (see Pereira et al., 2014).

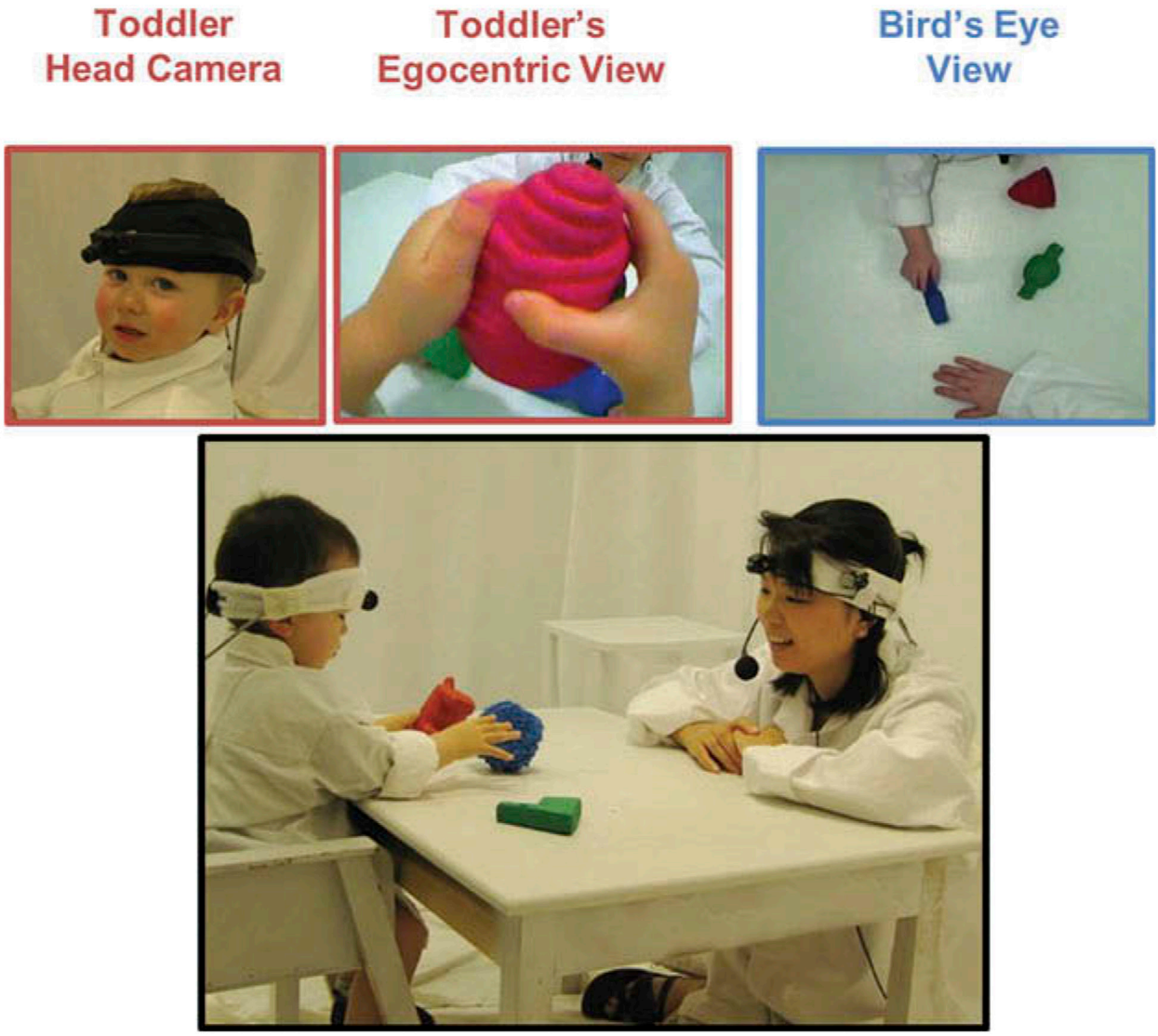

Figure 2.
The experimental set-up: Toddlers equipped with head cameras and their parents played with mono-colored objects in a laboratory room; play sessions were recorded from multiple angles.

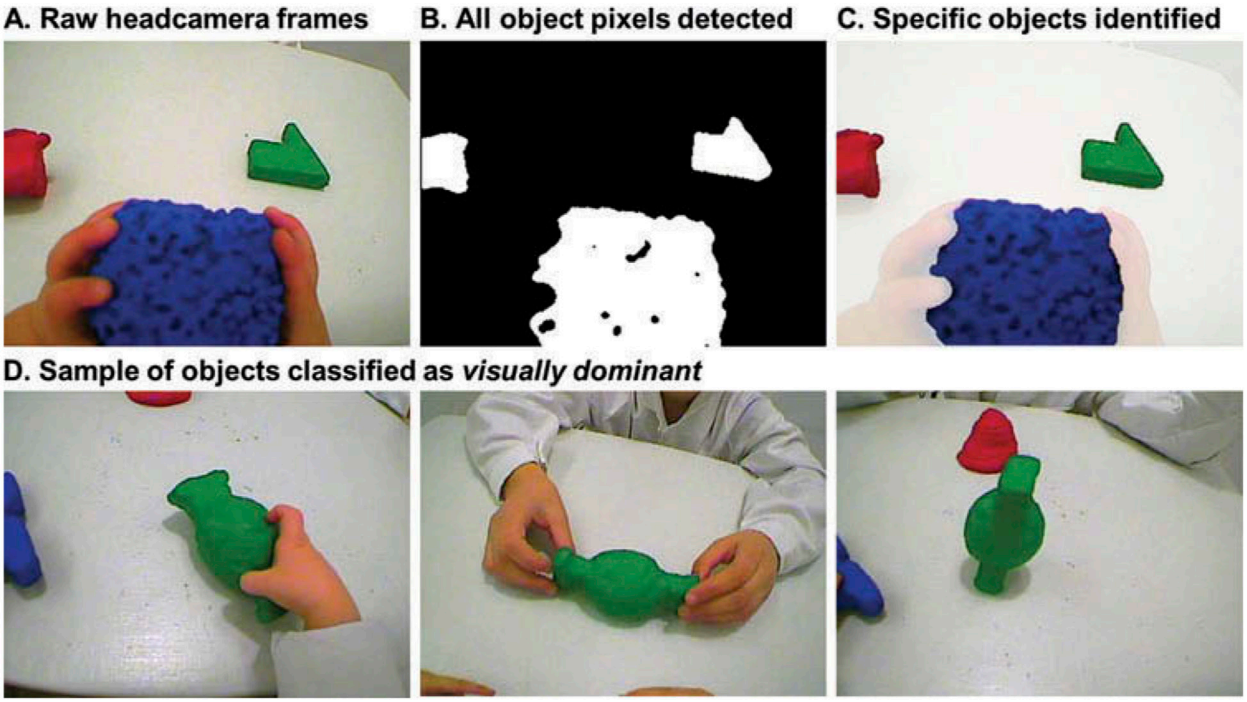

Figure 3.

Illustration of the automated processing of a single head camera image frame (A–C). Sampling of objects that were classified as visually dominant using the 5% field of view threshold (D).

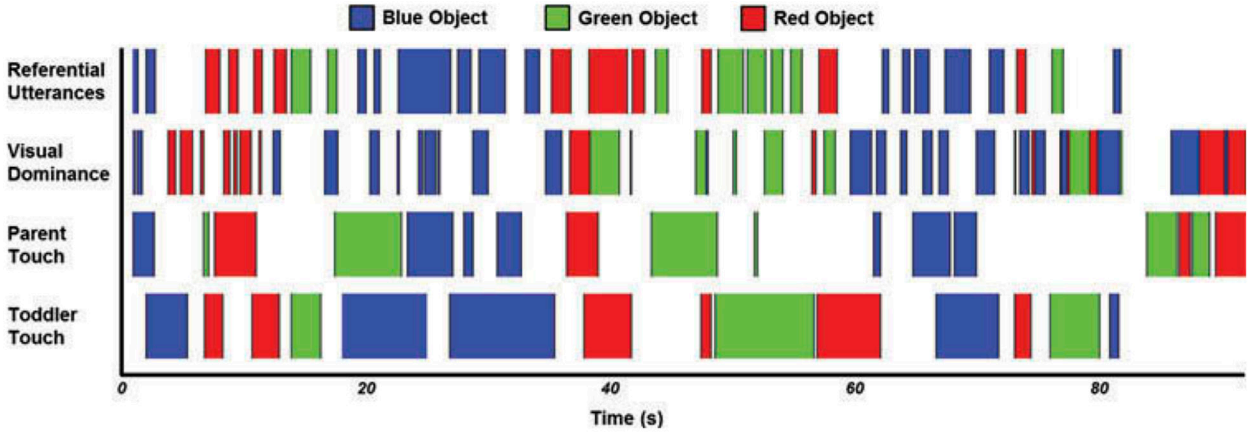

Figure 4.

Representative time series of parents' referential utterances (top row) and the nonverbal events (bottom three rows); colors reflect the target object of speech and nonverbal event.

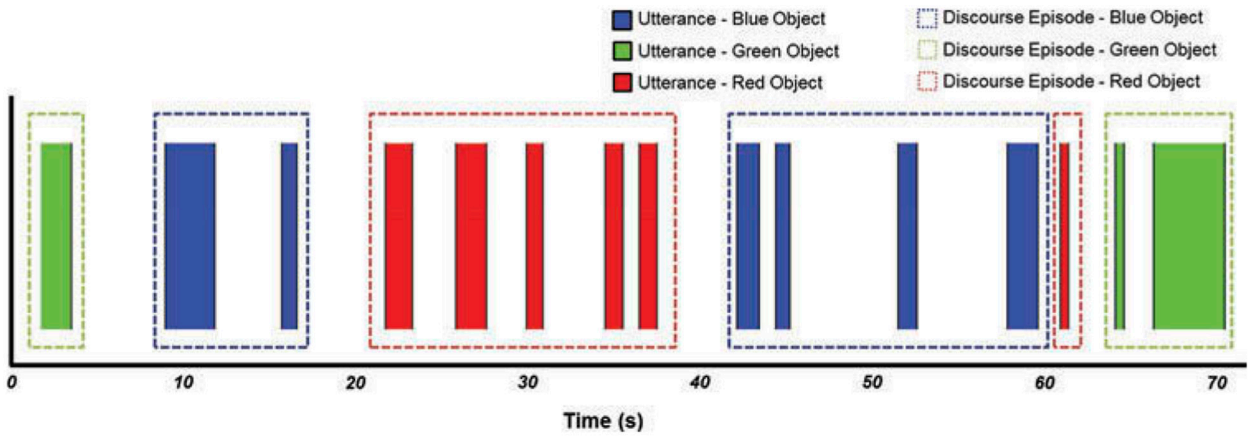

Figure 5.

Example time series of parents' utterances to different objects over the course of a trial. Adjacent utterances to the same object were merged into episodes of discourse.

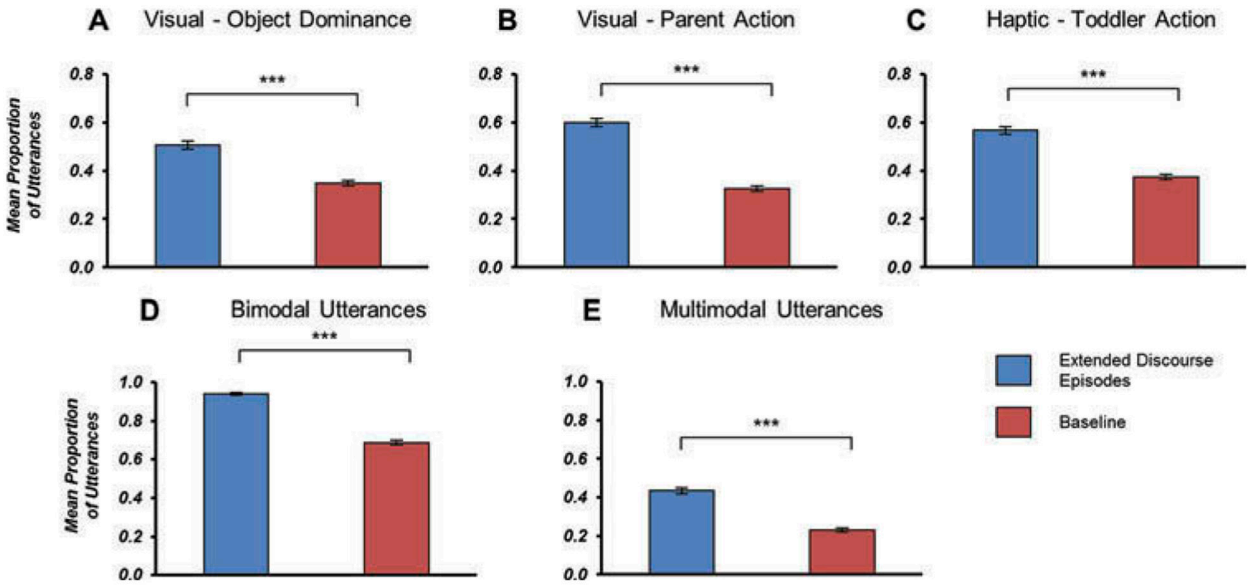

Figure 6.

Mean proportion of utterances within extended discourse episodes that overlapped with each nonverbal property (A–C), and that were classified as at least bimodal (D) or multimodal (E).

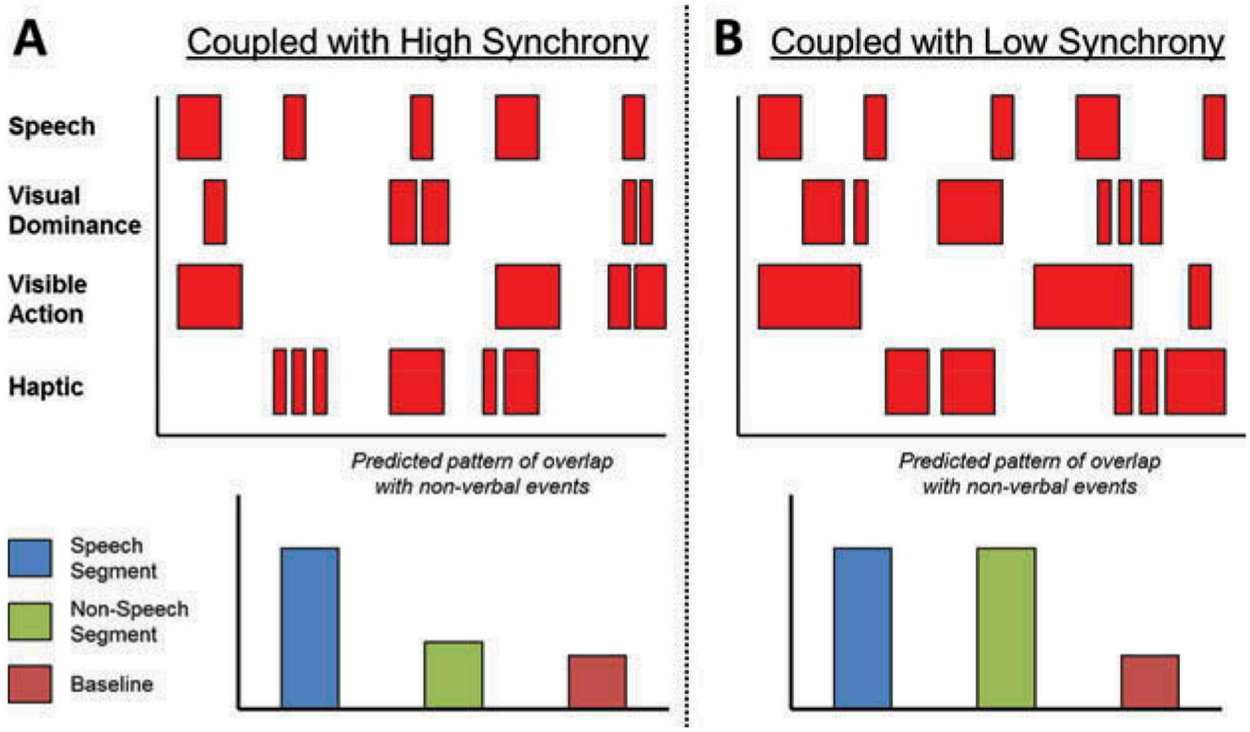

Figure 7.

Hypothetical data (top row) and predictions (bottom row) for two hypotheses on the synchronous nature of the multisensory coupling in parents' speech. Highly synchronous coupling would lead to greater overlap with nonverbal events in speech versus nonspeech segments; less synchronous coupling would lead to greater parity in overlap with nonverbal events between speech and nonspeech segments.

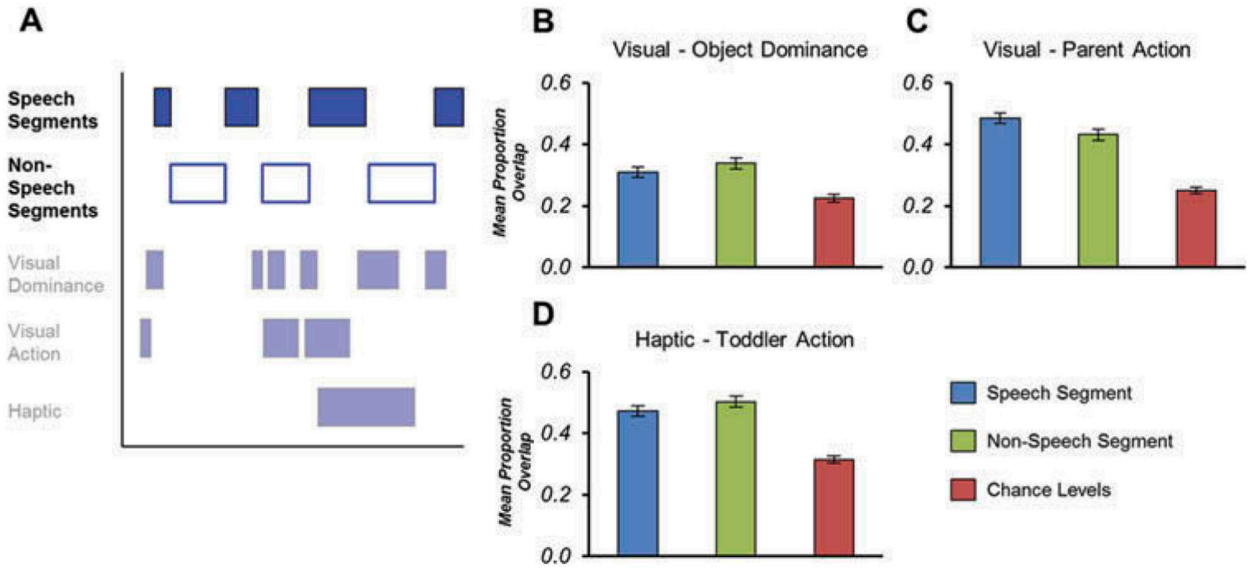

Figure 8.

Illustration of the relevant windows of analysis for speech and nonspeech segments (A). Mean proportion of speech and nonspeech segments that overlapped with the three nonverbal properties (B–D).

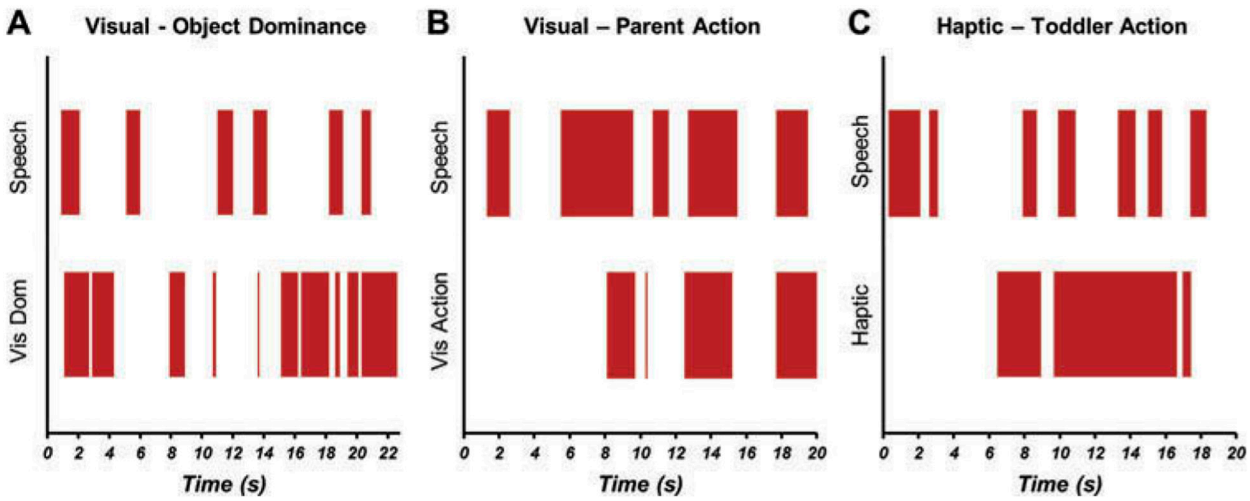

Figure 9.

Example time series of discourse episodes depicting the coupling between verbal and each nonverbal property.

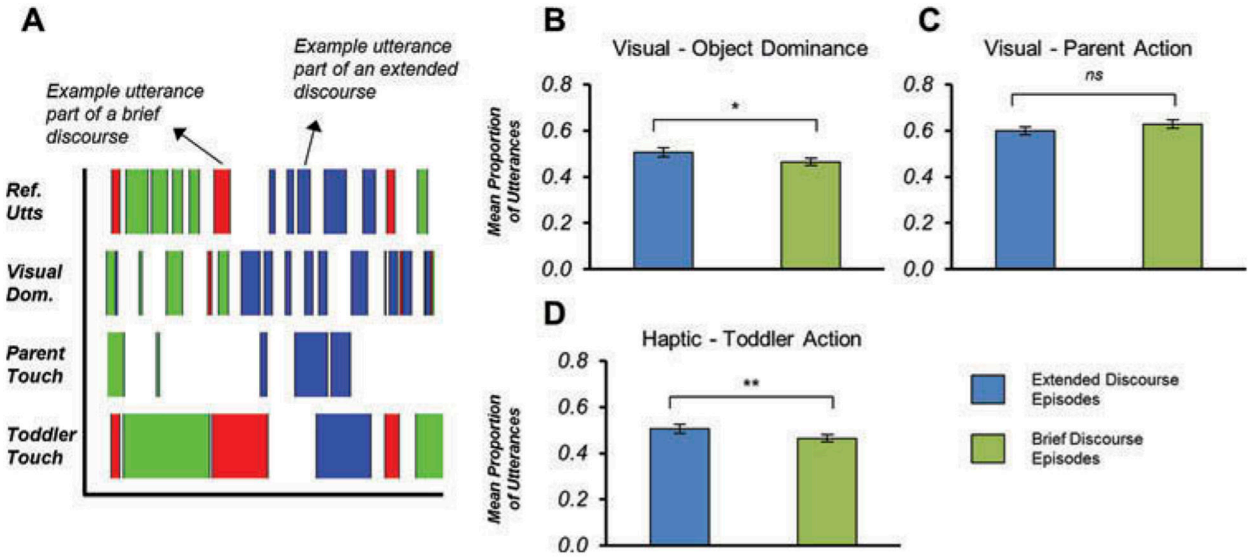

Figure 10.

(A) Examples of relevant moments for analysis for utterances in extended and brief episodes; (B–D) Mean proportion of utterances that contained some overlap with the three nonverbal properties.

Table 1.

Descriptive verbal statistics of extended and brief episodes of discourse

	Extended Episodes	Brief Episodes
Mean Number of Episodes	7.8 (3.8)	14.0 (7.0)
Mean Duration of Episode	16.2 (9.2)	3.3 (2.1)
Mean Run Length of Episode	4.3 (1.6)	1.4 (0.4)

Note. Standard deviations in parentheses.

Table 2.

Mean proportion of overlap with nonverbal properties

	Extended Episodes	Entire Interaction
Visual Object Dominance	.32 (.17)	.19 (.09)
Visible Parent Action	.45 (.17)	.23 (.08)
Toddler Haptic	.48 (.17)	.26 (.08)

Note. Standard deviations in parentheses.