Stacking the evidence: Parents' use of acoustic packaging with preschoolers

Nathan R. George,
Adelphi University

Federica Bulgarelli,
Duke University and Pennsylvania State University

Mary Roe,
Pennsylvania State University

Daniel J. Weiss
Pennsylvania State University

Abstract

Segmenting continuous events into discrete actions is critical for understanding the world. As infants may lack top-down knowledge of event structure, caregivers provide audiovisual cues to guide the process, aligning action descriptions with event boundaries to increase their salience. This *acoustic packaging* may be specific to infant-directed speech, but little is known about when and why the use of this cue wanes. We explore whether acoustic packaging persists in parents' teaching of 2.5–5.5-year-old children about various toys. Parents produced a smaller percentage of action speech relative to studies with infants. However, action speech largely remained more aligned to action boundaries relative to non-action speech. Further, for the more challenging novel toys, parents modulated their use of acoustic packaging, providing it more for those children with lower vocabularies. Our findings suggest that acoustic packaging persists beyond interactions with infants, underscoring the utility of multimodal cues for learning, particularly for less knowledgeable learners in challenging learning environments.

Keywords

event segmentation; acoustic packaging; multi-modal synchrony; language

6.**Supplementary Material**
Supplementary data to this article can be found through the following repository: George, N., Bulgarelli, F., & Weiss, D. (2019, April 18). Acoustic Packing: Preschoolers.

1. Introduction

A child watches her father prepare a snack in the kitchen. He cuts strawberries, turns on a faucet, scrubs his hands, and turns off the faucet before drying his hands and serving the snack. Though there are no pauses between each action, we perceive this stream of activity as a series of discrete units. Segmenting events into individual actions may seem trivial, given that adults rely on top-down knowledge of goals and intentions (Zacks & Tversky, 2001). However, early in development event segmentation poses a greater challenge, as infants may lack top-down knowledge about events they regularly encounter (Baldwin, Baird, Saylor, & Clark, 2001; Meyer & Baldwin, 2011; Meyer, Hard, Brand, McGarvey, & Baldwin, 2011). In the eyes of an infant, a more suitable analogy might be a ballet novice identifying a *rond de jambe* during a dancer's routine, or a newcomer to basketball identifying the event boundaries of a pick and roll.

A bottom-up cue purported to assist in event segmentation is a form of audiovisual synchrony termed *acoustic packaging*, in which parents align the onset and offset of action descriptions with the boundaries of the action being referenced (Brand & Tapscott, 2007; Hirsh-Pasek & Golinkoff, 1996; Meyer et al., 2011; Rohlfing, Fritsch, Wrede, & Jungmann, 2006; Rolf, Hanheide, & Rohlfing, 2009; Wrede, Schillingmann, & Rohlfing, 2013). This alignment reinforces the beginnings and ends of actions, facilitating infants' segmentation of events (Brand & Tapscott, 2007). Little is known, however, about how parents adjust their use of acoustic packaging as children age. In the current paper, we extend our understanding of the role of bottom-up cues in the development of event perception by asking 1) whether acoustic packaging persists beyond infancy and 2) whether caregivers' use of this cue is influenced by characteristics of the learners (such as knowledge of events and language), and the learning context (such as task novelty).

1.1 Acoustic Packaging

Event segmentation plays an important role in domains such as language learning (Friend & Pace, 2011; Hirsh-Pasek & Golinkoff, 1996) and memory (Hespos, Grossman, & Saylor, 2010; Hespos, Saylor, & Grossman, 2009). It emerges early in development as infants show consistency in their perception of event boundaries (Baldwin et al., 2001; Saylor, Baldwin, Baird, & LaBounty, 2007). But how do infants gain a foothold into event structure prior to extensive knowledge of goals and intentions? Parental input to children may play a valuable role (see Levine, Buchsbaum, Hirsh-Pasek, & Golinkoff, 2018 for a review). For example, parents' actions are characterized by *motionese*, the use of exaggerated motion paths and heightened pauses at action boundaries that serve to focus learners' attention (Brand, Baldwin, & Ashburn, 2002; Brand, Shallcross, Sabatos, & Massie, 2007; Brand & Shallcross, 2008; Koterba & Iverson, 2009). Parents may also help learners discover event boundaries through language. Acoustic packaging is a property of infant-directed communication wherein caregivers align the onset and offset of action descriptions to the boundaries of the actions themselves (Hirsh-Pasek & Golinkoff, 1996). Acoustic packaging is thought to facilitate learning, consistent with the utility of multimodal synchrony for word learning (Bahrick & Lickliter, 2000; Bahrick & Lickliter, 2002; Gogate & Hollich, 2016; Gogate, Maganti, & Bahrick, 2015; Gogate, Prince, & Matatyaho, 2009; Nomikou, Koke, &

Rohlfing, 2017). A distinctive feature of acoustic packaging, however, is that parents' alignment of action descriptions to actions not only highlights word mappings, but also allows for the edges of speech segments to act as a supporting cue for event segmentation, one that does not require the child to have any prior knowledge of the event or the labeling phrase.

Parents use acoustic packaging when teaching their infants about actions. Meyer and colleagues (2011) asked parents to demonstrate the functionality of simple toys (i.e., stacking rings, nesting cups) to their 6- to 13-month-old infants. In their sample, 57% of parent utterances described ongoing actions, and the boundaries of these action descriptions were more synchronous with the boundaries of actions than were the boundaries of non-action utterances. Further, when familiarized with a stream of goal-directed actions in a sequence, 9.5-month-olds reliably preferred an action pair that had aligned with the boundaries of a linguistic label relative to an observed action pair that did not align with the label (Brand & Tapscott, 2007). In sum, acoustic packaging is both available in infant-directed communication and valuable to naive learners.

1.2 Developmental Trajectory

As acoustic packaging is posited to help infants learn event boundaries (Brand & Tapscott, 2007; Meyer et al., 2011), and labels (Hirsh-Pasek & Golinkoff, 1996), parents' reliance on this cue may decline as children develop. This appears to be the case for other forms of synchrony that have been loosely termed "multimodal motherese" (Gogate, Bolzani & Betancourt, 2006; Gogate, Bahrick, & Watson, 2000), a parallel to the simplified, exaggerated, and melodic speech used to facilitate attention and processing during early stages of language acquisition (e.g., Grieser & Kuhl, 1988; Newport, 1977). For example, label-action synchrony (i.e., temporal proximity of linguistic label to referent) is known to wane between 5–30 months of age, with pre-lexical children receiving more synchronous input (Gogate et al., 2000; Gogate, et al., 2015). During the second year of life, parents also shift from describing actions as they are being performed to pre-labeling actions to highlight word meanings (Tomasello & Kruger, 1992), changing the nature of the synchrony. This change may reflect a shift toward social and grammatical cues and away from low-level perceptual cues (Hollich, et al., 2000; see also Gogate & Maganti, 2017). Further, older children may have greater capacity for maintaining relationships over longer durations, afforded by increases in memory capacity (Gogate, et al., 2015). Importantly, while these studies examine the proximity of a label to an action in language instruction, they do not specifically address whether boundaries in speech align with action boundaries to highlight event structure. Indeed, the aligning of both onsets and offsets of speech (specifically action speech) to coincide with the beginnings and ends of actions sets acoustic packaging apart from other forms of synchrony in its utility for cueing event structure.

Understanding the role of audiovisual synchrony in event segmentation across the lifespan has attracted interest beyond human development, as evidenced by research efforts in the fields of artificial intelligence and robotics. The characteristics of infant-directed speech that promote bottom-up learning are thought to be critical for developing intelligent systems that mimic infant social learning (Schillingmann, Wrede, & Rohlfing, 2009; Wrede et al., 2013).

To that end, researchers in these fields have compared boundary alignment in infant- and adult-directed interactions more broadly, finding that the boundaries of speech (of any kind) and action are less aligned overall in adult-directed speech relative to infant-directed speech. Further, infant-directed speech contains simpler packages with fewer actions per package (Schillingmann, Wrede, & Rohlfing, 2009; Wrede et al., 2013). While these studies shed light on some changes in alignment across the lifespan, it is important to note that these comparisons are agnostic with respect to speech category. It is possible that, despite an overall decline in speech-action synchrony, parents continue to provide tighter alignment of action speech to actions, even when teaching older learners. A deeper understanding of how acoustic packaging changes across development may therefore yield insights relevant to several disciplines.

A further consideration is that alignment may be tighter in environments in which learning is challenging. For instance, when audio-visual synchrony is provided, infants are better able to process streams of speech in noise (Hollich, Newman, & Jusczyk, 2005). Relatedly, research on infant-directed speech suggests that speech accommodation is more pronouncedin noisier environments (Newman, 2003), and even persists to some degree in adult-directed speech when facilitating comprehension with non-native interlocutors for whom the speech patterns are novel (Smith, 2007). Yet, to the best of our knowledge, speech-action boundary alignment has exclusively been studied in the context of demonstrating simple, familiar toys (e.g., stacking rings) in order to achieve controlled comparisons across ages (Schillingmann, et al., 2009; Wrede et al., 2013). Thus, it is possible that acoustic packaging could still be provisioned to older learners in contexts involving more challenging tasks.

This study investigated the extent to which acoustic packaging persists beyond speech directed to infants, providing the first glimpse at whether acoustic packaging might still be provisioned for preschool-aged children. We were further interested in better understanding the factors that influence when parents provide acoustic packaging. This includes both the perceived novelty of the task for the child as well as the child's overall vocabulary. We hypothesized that if acoustic packaging declines as a function of children's top-down knowledge of event structure, it may persist for more novel events regardless of age. Moreover, if parents use acoustic packaging to highlight word-to-world mappings, they may do so less as children's vocabularies increase. Adopting the procedures and analyses of Meyer and colleagues (2011), parents demonstrated a series of toys to preschool-aged children. In contrast to previous work, we investigated how children's top-down knowledge impacts parents' use of acoustic packaging by recording interactions with toys that we surmised would be familiar (stacking rings and nesting cups) and comparing them to interactions with a toy that we thought would be novel (a jungle-themed gear toy). We also assessed the role of language comprehension by correlating the degree of acoustic packaging with children's receptive vocabularies (Peabody Picture Vocabulary Test, Dunn & Dunn, 2007).

2. Methods

2.1 Participants

Fifty-six parent-child dyads participated in this study, recruited from a large-university community. Children ranged from 2.5 to 5.5 years of age (23 males; M_{age} = 49 months, SD = 9.39 months), and each participated along with one parent (6 males). Fourteen additional participants were tested, but not included in the analyses. For seven siblings, only the first child was included in analyses for two reasons. First, we wanted to ensure that each parent contributed the same amount of data. Second, we wanted to avoid introducing differences in teaching strategies as a result of having previously taught the toy to another child. Other participants were excluded from analyses due to missing recordings (n=1) or not completing the tasks (n=6). Written parental consent and verbal child assent were obtained for all participants.

2.2 Materials

Three toys were used in parent-child interactions. Two toys were chosen to be familiar and relatively simple to assemble: a Fisher-Price Brilliant Basics™ Rock-a-Stack, consisting of five stackable colored rings, and PlayGo™ Rainbow Stacking Cups, 10 colored cups that can be stacked or nested by size. The third toy was selected to be novel and more complicated to assemble, a Learning Resources Movin' Monkeys™ building set. This forest-themed gears toy consists of structural components (e.g., bases, trees) as well as gears, cranks, and decorations that form a simple machine (see Figures 1 and 2). The use of two simple toys and one complex toy was designed to help balance the number of interactions; we anticipated that the simpler toys would elicit relatively shorter interactions.

Photos of each assembled toy were provided to parents. A single photo was used for the stacking rings, and two photos were used for the stacking cups: one of the cups nested and one of the cups stacked (see Figure 1). Given the complexity of the novel toy, four photos were provided: one showing the completed toy, and three from the instruction booklet demonstrating how to assemble the various components (see Figure 2).

The Peabody Picture Vocabulary Test (PPVT-4) assessed receptive vocabulary (Dunn & Dunn, 2007). A survey assessed parent and child familiarity with the toys using a 10-point Likert scale (10 being most familiar) and inquired about the frequency with which the child interacted with each toy, at the age they most frequently played with it.

2.3 Procedure

2.3.1 Warm-up—During the warm-up phase, the child was escorted to a testing room and tested on the PPVT-4. Another experimenter accompanied the parent to a separate testing room to familiarize them with the toys and assembly pictures. The parent was informed that the goal of the study was to demonstrate to the child how to use each toy, and that they could interact with the toys for as long as they wanted. Parents were encouraged to look at toy pieces and assembly instructions. They were not required to do anything further (e.g., practice building) though they could ask for clarification regarding assembly.

2.3.2 Interactions—Following the warm-up phase, the parent and child were brought into a soundproof testing room and provided one toy at a time (order counterbalanced across subjects). The parent was instructed to demonstrate the construction of the toy to the child prior to the child interacting with it, mirroring previous research with infants (Meyer et al., 2011). Sessions were untimed, with interactions lasting until the child lost interest, as determined by the parent. Session lengths ranged from 7–53 minutes. All sessions were video recorded.

2.3.3 Coding—Audio and video components of recordings were separated for analysis by blind coders. The audio components of the recordings were imported into Praat (Boersma, 2001) and silences of 200ms or greater were used to define the boundaries of speech segments (see Meyer et al., 2011). Each speech segment was subsequently transcribed for content. Blind coders classified them as either referring to ongoing action descriptions or other speech (hereafter referred to as non-action speech). Ongoing action descriptions were utterances that contained an action verb in the present tense, whereas all other verbal segments were classified into one of several categories adapted from Meyer and colleagues (2011). These additional categories included attention getting (utterances to refocus the child's attention), goal setting (utterances that referred to future actions), action completion (utterances that referred to an already completed action), or other (all other utterances, see Meyer et al., 2011; see also Table 1 for examples). Intercoder reliability was calculated for 20% of speech segments, revealing moderate agreement K=0.70 (Landis & Koch, 1977).

The video was analyzed by coders to determine the onset and offset of actions related to assembling the toy. As in Meyer and colleagues (2011), onsets of actions were defined as the point at which the parent began moving a component of a toy down towards its intended place from an apex. The offset was defined as the moment at which a component of the toy was fully attached (i.e., settled, not moving; see Figure 3). Intercoder reliability was calculated for 20% of action boundaries. When raters had differing numbers of actions identified for a given task, missing actions were imputed by identifying the boundaries closest to those identified by the other rater. Results revealed excellent agreement, ICC=.99 (Cicchetti, 1994).

To assess alignment, difference scores were created. For the onset of each speech segment, the closest action onset was identified and the absolute value of the difference between the two onsets was calculated. The same procedure was used for offsets.

2.3.4 Data Cleaning—We removed all speech segments that occurred before the start of first action and after the end of the last action. This isolated speech used during the task, ensuring that comparisons were not biased by large difference scores resulting from unrelated conversations (e.g., parent helping child get comfortable). After computing the difference scores, we removed data points where the difference between speech and action was greater than 30 seconds. This conservative cut-off was above 2.5 standard deviations from the mean for the difference scores of both onsets and offsets and led to the exclusion of 240 data points (1.9% of data).

3. Results

Parents rated their children's familiarity with the novel (M = 2.08, SD= 1.94) and familiar (M = 8.78, SD = 1.81) tasks as significantly different, $t(52)$ = 19.45, p<.001; however, there were also differences between the familiar tasks. Parents rated their children as being significantly more familiar with rings (M = 9.19, SD = 1.75) compared to cups (M = 8.38, SD = 2.44), $t(52)$ = 2.65, p<.05. To account for these differences, we categorized each task for each child based on individual parent ratings (familiar: above 5, novel: 5 or less). Thus, each level of our novelty variable contained a mix of tasks, though they predominantly patterned according to our original intentions (i.e., the jungle toy largely appraised as novel; cups and rings tended to appraise as more familiar)

3.1 Prevalence of Action Speech

We first looked at the percentage of action utterances produced by parents broken down by task. Fewer than 10% of the utterances produced by parents were action utterances (compared to 57% in speech to infants in Meyer et al., 2011), and this percentage was fairly consistent across tasks (see Table 1).

3.2 Predictors of Acoustic Packaging

We conducted two separate generalized linear mixed models in SPSS (version 25), one examining predictors of onset alignment (i.e., difference score between utterance onset and nearest action onset), and the second examining predictors of offset alignment (i.e., difference score between utterance offset and nearest action offset). In both models, data were clustered by parent-child dyad. Along with fixed and random intercepts, the following fixed factors were entered as predictors: Utterance Type (action vs. non-action speech; dummy coded), Novelty (familiar vs. novel based on parent ratings; dummy coded), and Receptive Vocabulary (grand mean centered PPVT). In addition, all two- and three-way interactions between Utterance type and the variables of Novelty, and Receptive Vocabulary were added as fixed factors. Finally, Age and Session Length (both grand mean centered) were entered as covariates (see Appendix A for full models).

For the onset model, there was a significant covariate effect of Session Length, $F(1, 8,554)$ = 457.19, p<.001. Importantly, we also found a main effect of Utterance Type, $F(1, 8,554)$ = 5.44, p<.05, suggesting that acoustic packaging persists in speech to preschool-aged children, with action speech onsets more aligned to action onsets than other types of speech. With regards to our hypotheses about the roles of learner and task characteristics, we also observed main effects of Receptive Vocabulary, $F(1, 8,554)$ = 5.81, p<.05, and Novelty, $F(1, 8,554)$ = 3.87, p<.05, which can be best understood in the context of a significant three-way interaction between Utterance Type, Novelty, and Receptive Vocabulary, $F(2, 8,554)$ = 13.01, p<.001. Within the novel task, non-action speech onsets became less aligned to action onsets as Receptive Vocabulary increased, β = .041, p<.001. The decrease in alignment was greater for action speech, though it was marginally significant, β = .057, p = .055. A different pattern is observed in the familiar task, as neither non-action nor action speech onset alignment related to Receptive Vocabulary, $\beta_{\text{non-action}}$ = −.002, p = .71; β_{action} = −.011, p = .48.

The offset model revealed a similar pattern of results. There remained a significant covariate effect of Session Length, $F(1, 8,554) = 479.17$, $p < .001$; however, only a marginal effect of Utterance Type, $F(1, 8,554) = 3.54$, $p = .060$. In terms of learner and task characteristics, we only observed a marginal effect of Novelty, $F(1, 8,554) = 3.54$, $p = .060$, but again observed a significant main effect of PPVT, $F(1, 8.554) = 5.17$, $p < .05$, and a significant three-way interaction between Utterance Type, Novelty, and Receptive Vocabulary, $F(2, 8,554) = 11.13$, $p < .001$. For novel tasks, both action and non-action speech became less aligned with action boundaries as Receptive Vocabulary increased, though to varying degrees, $\beta_{\text{non-action}} = .036$, $p < .001$; $\beta_{\text{action}} = .060$, $p < .05$. Similar to onsets, neither action nor non-action offset alignment were related to Receptive Vocabulary in the familiar task, $\beta_{\text{non-action}} = -.003$, $p = .62$; $\beta_{\text{action}} = -.013$, $p = .43$ (see Figure 4).

4.   Discussion

Our study provided the first data, to our knowledge, regarding the use of acoustic packaging in speech to 2.5–5.5-year-olds. Parents described to their children how to assemble both familiar and novel toys and we measured the relative alignment of action speech and action boundaries. We were interested in determining whether acoustic packaging might decrease in response to both changes in the learner's knowledge as well as the demands of a specific task. We found that boundaries for action speech are more aligned to action boundaries than are boundaries for non-action speech, revealing a continued use of acoustic packaging at this age. Consistent with our predictions regarding the influence of task and child-specific factors, we observed that the tighter alignment of action speech and action boundaries (relative to non-action speech and action boundaries) dissipates as vocabularies increase, but only for novel tasks. When the task is challenging, parents appear to be more sensitive to the child's need for packaging, and thus modulate their speech to rely more on this cue for children with smaller vocabularies. Our results demonstrate the importance of measuring acoustic packaging in relation to both learner characteristics and the specific learning context. More broadly, our results suggest that acoustic packaging, which serves as a scaffold for event segmentation (see Levine, et al., 2018), is available to learners even after language production is well underway.

Central to the aims of our study, we found that acoustic packaging in child-directed speech is dependent upon both learner characteristics and task demands. While acoustic packaging was provided in descriptions of the familiar tasks, a more complex pattern emerged within the novel tasks. Specifically, parents of children with lower receptive vocabularies provided tighter synchrony across all types of speech as compared to parents of children with higher receptive vocabularies. Further, the alignment of action speech and actions declined more drastically, ultimately converging to be comparable to the alignment of non-action speech in children with the largest vocabularies (see Figure 4b, 4d). This pattern of results demonstrates the importance of perceived novelty in the developmental trajectory of acoustic packaging. The demands of a novel task appear to heighten parents' attention to the degree to which their children require extra support for learning. Similar to studies of synchrony in word learning contexts (Gogate, et al. 2000; Gogate, et al., 2015), parents subsequently modulate their use of acoustic packaging in response to learner characteristics, in this case child vocabulary. As children demonstrate greater proficiency in language, parents not only

reduce the overall alignment of speech to actions, but focus less on acoustic packaging, allowing the alignment of action speech and action boundaries to vary as freely as they do the alignment of non-action speech and action boundaries.

There are at least two possible reasons underlying the observed relationship between the child's vocabulary and parents' use of acoustic packaging. First, consistent with acoustic packaging being an important source of information for event segmentation (e.g., Brand & Tapscott, 2007; Levine et al., 2018; Wrede et al., 2013), parents may come to rely less on this cue as a function of their child demonstrating knowledge of events through language. Alternatively, consistent with the role of acoustic packaging in language learning (Hirsh-Pasek & Golinkoff, 1996), parents' decreased use of acoustic packaging may be elicited in response to the child possessing a larger vocabulary, and thus needing less support for word-to-world mappings (see also Gogate & Maganti, 2017). Testing situations in which event and language knowledge can be differentiated (e.g., familiar events being taught to second-language learners) may help delineate the extent to which each factor contributes to the observed decline of acoustic packaging beyond infancy.

Another factor that may contribute to the overall looser alignment of speech and action boundaries for the novel task is the broader shift in parent-child communication in these age groups. As children age and their vocabularies develop, they contribute more to conversations. We found that the percentage of parental speech devoted to describing action drops drastically in interactions with preschoolers (~9%) relative to infant-directed speech (~57%; Meyer et al., 2011). In contrast, there was a sharp rise in utterances categorized as "other" (~75%, up from roughly 12% for infant-directed speech; Meyer et al., 2011). Increases in this category may reflect a shift to more dynamic exchanges in which parents responded to children's utterances rather than just describing the task. For example:

Parent: "Let's see if we keep those on the very top."

Child: "He's walking."

Parent: "The monkey's walking?"

Child: "No. Ride the monkey. Ride the monkey."

Parent: "You're riding the monkey?"

Child: "Yeah"

Parent: "I don't think most monkeys give people rides."

This new conversational dynamic creates a wider range of topics for discussion, including child-driven shifts toward tangential content, which in turn may create larger lags between the smaller percentage of action speech and the task-relevant actions described. Future research might examine the extent to which declines in acoustic packaging can be predicted by this increased conversational role for the child.

We note that the trajectory of acoustic packaging in development is reminiscent of the use of other forms of speech accommodation (Smith, 2007), including the synchrony observed in language instruction contexts (Gogate, et al., 2000; Gogate, et al., 2015), as each appears less pronounced when learners exhibit greater understanding of language. Further, just as infant-directed speech becomes more pronounced in noisier environments (Newman, 2003), the quality of acoustic packaging increases in learning contexts that are perceived to be novel, particularly when interacting with children who are thought to require greater learning support. Acoustic packaging therefore represents another strategy flexibly employed by parents (and possibly other interlocutors) to highlight aspects of the input, providing naïve learners with cues that assist in both language learning and event segmentation.

The finding that acoustic packaging persists in child-directed speech provides a useful bridge between the studies of acoustic packaging in infants (Meyer, et al., 2011), and subsequent comparisons to adult-directed speech (Schillingmann et al., 2009; Wrede et al., 2013). Much like Meyer and colleagues (2011), we find that, in some contexts, parents continue to structure their speech such that the boundaries of actions are preferentially reinforced by the boundaries of labelling phrases. At the same time, our results within the novel task also reflect trends observed in artificial intelligence work (Schillingmann et al., 2009; Wrede et al., 2013), indicating that the overall alignment of speech and action boundaries appears looser for older learners (differences of several seconds) relative to infant-directed speech (differences under a second; Meyer, et al., 2011). These dual perspectives highlight an important question: does overall speech-action alignment (Schillingmann et al., 2009; Wrede et al., 2013) or the tighter alignment of action speech to actions relative to non-action speech (Meyer, et al., 2011) best explain the utility of audiovisual alignment for event segmentation? While we did not explicitly test how children benefit from the use of acoustic packaging, both aspects of alignment may have important roles. For example, there is evidence to suggest that alignment of verbs and actions provided to 6-month-olds in everyday routines (e.g., diapering) predicts later vocabulary at 24 months (Nomikou, Koke, & Rohlfing, 2017). However, even though alignment was looser in the input provided to older children in our study, such alignment may still be valuable for learners. Increases in cognitive abilities such as working memory, a known predictor of vocabulary development (Verhagen & Leseman, 2016), may permit older children to make use of relative differences in alignment between action and non-action speech, even as overall speech-action alignment widens (see also Gogate, et al., 2015). Further, we note the observation that young learners do encounter multisensory regularities at longer timescales, such as in extended discourse (Suanda, et al., 2016). It is possible that acoustic packaging may similarly operate across multiple hierarchical levels, potentially serving additional roles in binding actions into broader routines (Hirsh-Pasek & Golinkoff, 1996). Addressing which types of alignment are most valuable for event segmentation as children age will require additional research investigating how event knowledge coalesces as a function of alignment across age groups and timescales.

While our results suggest that the novelty of the learning situation impacts acoustic packaging, we concede that the tasks we chose also differed along other dimensions. For instance, the jungle themed gears toy required a multitude of actions (e.g., snapping,

hanging, twisting, etc.) whereas the cups and rings toys were more restricted (e.g., stacking by size). This difference in complexity could contribute to the observed differences in acoustic packaging, as the variety of action terms used may have encouraged a greater focus on the use of packaging compared to the simple, repeated actions of the familiar task. We further note that the familiar toys were restricted in terms of the order of assembly (i.e., larger pieces first, then progressively smaller). Given that event boundaries correspond with uncertainty about what comes next (Hard, Recchia, & Tversky, 2011; Kurby & Zacks, 2008), the increased variability with respect to the assembly of the novel toy may have yielded greater variability in the structure of the input. That said, it is unlikely that these features alone can account for our findings, as we note that some parents rated the cups as less familiar and the jungle toy as more familiar and consequently, our analyses did not strictly pit the cups and rings against the jungle toy. In sum, given the variability in our tasks, our study provides a useful platform to begin teasing apart the most relevant factors that undergird the extended use of acoustic packaging beyond speech to infants.

In sum, our experiment provides a new perspective on the role of acoustic packaging in building event and linguistic knowledge. While it has been acknowledged that acoustic packaging provides a valuable scaffold for infant language learners' understanding of events (Nomikou, et al., 2017; Levine, et al., 2018), here we demonstrate that acoustic packaging does not disappear after infancy. Rather, reliance on this bottom-up cue appears to stem from sensitivity to learners' knowledge of events and language as well as the demands of a task. Consequently, our study suggests that acoustic packaging may be more prevalent than previously thought. Further, our results suggest that studies assessing acoustic packaging should manipulate both task difficulty and the perceived knowledge of learners with respect to both events and language, the latter of which may be quite low in future populations of interest, such as low-proficiency adult second-language learners. Our results bear straightforward implications for extending this work to other domains, such as computational approaches to language learning (e.g., Schillingman, et al., 2009; Wrede, et al., 2013) as well as adult language learning experiences.

Acknowledgements

Appendix

Table A1.

Mixed Effects Model Results: Onset Alignment

Fixed effects (intercepts, slopes)	Estimate *(SE)*	*p*	CI$_{95}$ Lower	Upper
Fixed Effects				

Fixed effects (intercepts, slopes)	Estimate *(SE)*	*p*	CI95 Lower	Upper
Intercept	4.682 (3.515)	0.183	−2.208	11.571
Session Length (covariate)	0.002 (0.000)	< 0.001	0.002	0.002
Age (covariate)	0.007 (0.005)	0.185	−0.003	0.018
Utterance Type	−0.720 (0.211)	<.010	−1.134	−0.306
Novelty	0.459 (0.158)	<.010	0.150	0.768
Receptive Vocabulary	−0.002 (0.006)	0.707	−0.013	0.009
Novelty * Utterance Type	0.174 (0.541)	0.747	−0.886	1.234
Receptive Vocabulary * Utterance Type	−0.009 (0.017)	0.585	−0.042	0.024
Receptive Vocabulary * Novelty * Utterance Type				
*Receptive Vocabulary * Novelty (Level: Non-Action Speech)*	0.043 (0.009)	<0.001	0.025	0.061
*Receptive Vocabulary * Novelty (Level: Action Speech)*	0.069 (0.034)	<0.050	0.002	0.135
Random Effects				
Residual Variance	24.678 (0.377)	<.001	23.950	25.429

Note: N = 56 dyads, 12,262 observations

Table A2.

Mixed Effects Model Results: Offset Alignment

Fixed effects (intercepts, slopes)	Estimate *(SE)*	*p*	CI95 Lower	Upper
Fixed Effects				
Intercept	4.552 (3.547)	1.283	0.199	−2.401
Session Length (covariate)	0.002 (0.000)	21.890	< 0.001	0.002
Age (covariate)	−0.002 (0.006)	−0..420	0.675	−0.013
Utterance Type	−0.574 (0.213)	−2.695	<.010	−0.991
Novelty	0.468 (0.159)	2.941	<.010	0.156
Receptive Vocabulary	−0.003 (0.006)	−0.499	0.618	−0.014
Novelty * Utterance Type	0.117 (0.546)	0.215	0.830	−0.952
Receptive Vocabulary * Utterance Type	−0.010 (0.017)	−0.577	0.564	−0.043
Receptive Vocabulary * Novelty * Utterance Type				
*Receptive Vocabulary * Novelty (Level: Non-Action Speech)*	0.039 (0.009)	4.217	<0.001	0.021
*Receptive Vocabulary * Novelty (Level: Action Speech)*	0.073 (0.034)	2.119	<0.050	0.005
Random Effects				
Residual Variance	25.134 (0.384)	<.001	24.392	25.899

Note: N = 56 dyads, 12,262 observations

References

Bahrick LE, & Lickliter R (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. Developmental Psychology, 36(2), 190–201. [PubMed: 10749076]

Bahrick LE, & Lickliter R (2002). Intersensory redundancy guides early perceptual and cognitive development. Advances in Child Development and Behavior, 30, 153–187. [PubMed: 12402674]

Baldwin DA, Baird JA, Saylor MM, & Clark MA (2001). Infants parse dynamic action. Child Development, 72(3), 708–717. [PubMed: 11405577]

Boersma PG (2001). Praat, a system for doing phonetics by computer. Glot international, 5, 341–345.

Brand RJ, Baldwin DA, & Ashburn LA (2002). Evidence for "motionese": Modifications in mothers' infant-directed action. Developmental Science, 5(1), 72–83.

Brand RJ, & Shallcross WL (2008). Infants prefer motionese to adult-directed action. Developmental Science, 11(6), 853–861. [PubMed: 19046154]

Brand RJ, Shallcross WL, Sabatos MG, & Massie KP (2007). Fine-grained analysis of motionese: Eye gaze, object exchanges, and action units in infant- versus adult-directed action. Infancy, 11(2), 203–214.

Brand RJ, & Tapscott S (2007). Acoustic packaging of action sequences by infants. Infancy, 11, 321–332.

Cicchetti DV (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6(4), 284–290.

Dunn LM & Dunn DM (2007). Peabody Picture Vocabulary Test – Fourth Edition: Manual. Minneapolis, MN: Pearson Assessments.

Friend M, & Pace A (2011). Beyond event segmentation: Spatial- and social-cognitive processes in verb-to-action mapping. Developmental Psychology, 47(3), 867–876. [PubMed: 21142370]

George NR, Göksun T, Hirsh-Pasek K & Golinkoff RM (2014). Carving the world for language: How neuroscientific research can enrich the study of first and second language learning. Developmental Neuropsychology, 39(4), 262–284. [PubMed: 24854772]

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(4), 878– 894. [PubMed: 11016554]

Gogate LJ, Bolzani LH, & Betancourt EA (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word–object relations. Infancy, 9(3), 259–288.

Gogate L, & Hollich G (2016). Early verb-action and noun-object mapping across sensory modalities: A neuro-developmental view. Developmental Neuropsychology, 41(5–8), 293–307. [PubMed: 28059566]

Gogate L, & Maganti M (2017). The origins of verb learning: Preverbal and postverbal infants' learning of word–action relations. Journal of Speech, Language, and HearingResearch, 60(12), 3538–3550.

Gogate L, Maganti M, & Bahrick LE (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. [PubMed: 25285369]

Gogate LJ, Prince CG, & Matatyaho DJ (2009). Two-month-old infants' sensitivity to changes in arbitrary syllable--object pairings: The role of temporal synchrony. Journal of Experimental Psychology: Human Perception and Performance, 35(2), 508–519. [PubMed: 19331504]

Grieser DL, & Kuhl PK (1988). Maternal speech to infants in a tonal language: Support for universal prosodic features in motherese. Developmental Psychology, 24(1), 14–20.

Hard B, Recchia G, & Tversky B (2011). The shape of action. Journal of Experimental Psychology, 140(4), 586–604. [PubMed: 21806308]

Hespos SJ, Grossman SR, & Saylor MM (2010). Infants' ability to parse continuous actions: Further evidence. Neural Networks, 23, 1026–1032. [PubMed: 20729035]

Hespos SJ, Saylor MM, & Grossman SR (2009). Infants' ability to parse continuous actions. Developmental Psychology, 45(2), 575–585. [PubMed: 19271840]

Hirsh-Pasek K, & Golinkoff RM (1996). The origins of grammar: Evidence from early language comprehension. Cambridge, MA/London, England: MIT Press.

Hollich GJ, Hirsh-Pasek K, Golinkoff RM, Brand RJ, Brown E, Chung HL, … & Bloom L (2000). Breaking the language barrier: An emergentist coalition model for the origins of word learning. Monographs of the Society for Research in Child Development, i-135.

Hollich G, Newman RS, & Jusczyk PW (2005). Infants' use of synchronized visual information to separate streams of speech. Child Development, 76(3), 598–613. [PubMed: 15892781]

Koterba EA, & Iverson JM (2009). Investigating motionese: The effect of infant-directed action on infants' attention and object exploration. Infant Behavior & Development, 32(4), 437–444. [PubMed: 19674793]

Kurby CA & Zachs JM (2008). Segmentation in the perception and memory of events. Trends in Cognitive Science, 12(2), 72–79.

Landis JR & Koch GG (1977). The Measurement of Observer Agreement for Categorical Data. Biometric, 33, 159–174.

Levine D, Buchsbaum D, Hirsh-Pasek K, & Golinkoff RM (2018). Finding events in a continuous world: A developmental account. Developmental Psychobiology. Advance online publication.

Liu H-M, Tsao F-M, & Kuhl PK (2009). Age-related changes in acoustic modifications of Mandarin maternal speech to preverbal infants and five-year-old children: A longitudinal study. Journal of Child Language, 36(4), 909–922. [PubMed: 19232142]

Meyer M, & Baldwin D (2011). Statistical learning of action: The role of conditional probability. Learning & Behavior, 39, 383–398. [PubMed: 21547528]

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(2), 154–162.

Newman RS (2003). Prosodic differences in mothers' speech to toddlers in quiet and noisy environments. Applied Psycholinguistics, 24(4), 539–560.

Newport EL (1977). Motherese: The speech of mothers to young children In Castellan NJ, Pisoni DB, & Potts G, (Eds.), Cognitive Theory, Vol. 2 Hillsdale, N.J.: Erlbaum.

Nomikou I, Koke M, & Rohlfing KJ (2017). Verbs in mothers' input to six-month-olds: Synchrony between presentation, meaning, and actions is related to later verb acquisition. Brain Sciences, 7(5), 52.

Rohlfing KJ, Fritsch J, Wrede B, & Jungmann T (2006). How can multimodal cues from child-directed interaction reduce learning complexity in robots? Advanced Robotics: The International Journal of the Robotics Society of Japan, 20(10), 1183–1199.

Rolf M, Hanheide M, & Rohlfing KJ (2009). Attention via synchrony: Making use of multimodal cues in social learning. IEEE Transactions on Autonomous Mental Development, 1(1), 55–67.

Saint-Georges C, Chetouani M, Cassel R, Apicella F, Mahdhaoui A, Muratori F, … Cohen D (2013). Motherese in interaction: At the cross-road of emotion and cognition? (a systematic review). PloS One, 8(10), e78103. [PubMed: 24205112]

Saylor MM, Baldwin D. a., Baird J. a., & LaBounty J (2007). Infants' on-line segmentation of dynamic human action. Journal of Cognition and Development, 8(1), 113–128.

Schillingmann L, Wrede B, & Rohlfing KJ (2009). A computational model of acoustic packaging. IEEE Transactions on Autonomous Mental Development, 1(4), 226–237.

Smith C (2007). Prosodic accommodation by French speakers to a non-native interlocutor In Barry WJ & Trouvain J (Eds.), Proceedings of the XVIth international congress of phonetic sciences (pp. 1081–1084). Saarbücken, Germany: Univiversity des Saarlandes.

Suanda SH, Smith LB, & Yu C (2016). The multisensory nature of verbal discourse in parent–toddler interactions. Developmental Neuropsychology, 41(5–8), 324–341. [PubMed: 28128992]

Tomasello M, & Kruger AC (1992). Joint attention on actions: Acquiring verbs in ostensive and non-ostensive contexts. Journal of Child Language, 19(2), 311–333. [PubMed: 1527205]

Verhagen J, & Leseman P (2016). How do verbal short-term memory and working memory relate to the acquisition of vocabulary and grammar? A comparison between first and second language learners. Journal of Experimental Child Psychology, 141, 65–82. [PubMed: 26340756]

Wrede B, Schillingmann L, & Rohlfing KJ (2013). Making use of multi-modal synchrony: A model of acoustic packaging In Gogate LJ & Hollich G (Eds.), Theoretical and Computational Models of Word Learning: Trends in Psychology and Artificial Intelligence (pp. 224–240). Hershey, PA: IGI Global.

Zacks JM, & Tversky B (2001). Event structure in perception and conception. Psychological Bulletin, 127, 3–21. [PubMed: 11271755]

Highlights

- Audiovisual synchrony in speech is a powerful cue for infant event segmentation

- It is unknown whether it persists after child gains top-down knowledge

- We demonstrate that acoustic packaging persists in speech to preschoolers

- Parents provisioning of this cue is related to child vocabulary and novelty of event

- Results suggest audiovisual synchrony informs event structure beyond infancy

Figure 1.
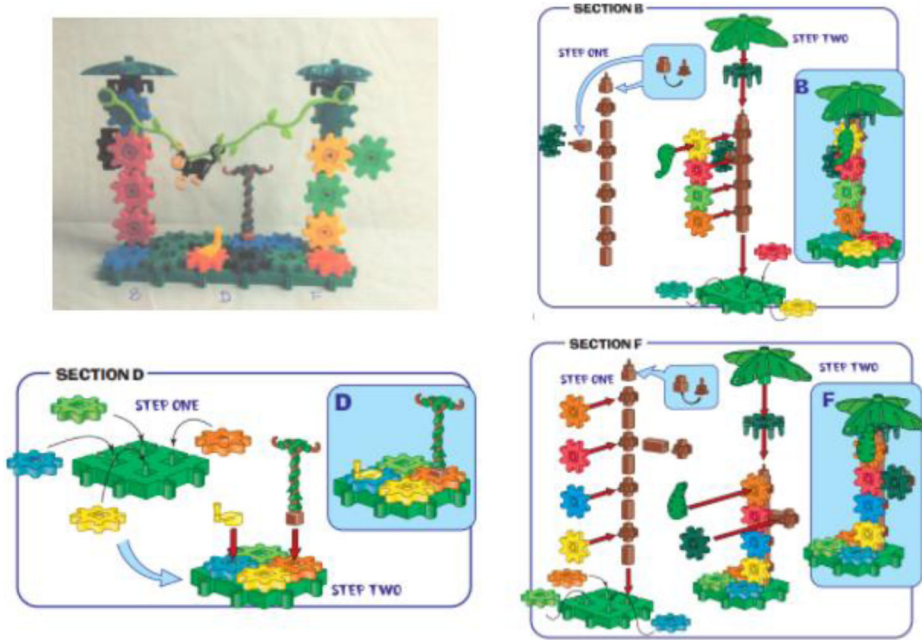Photographs of assembled simple toys for use by parents as a reference.

Figure 2.
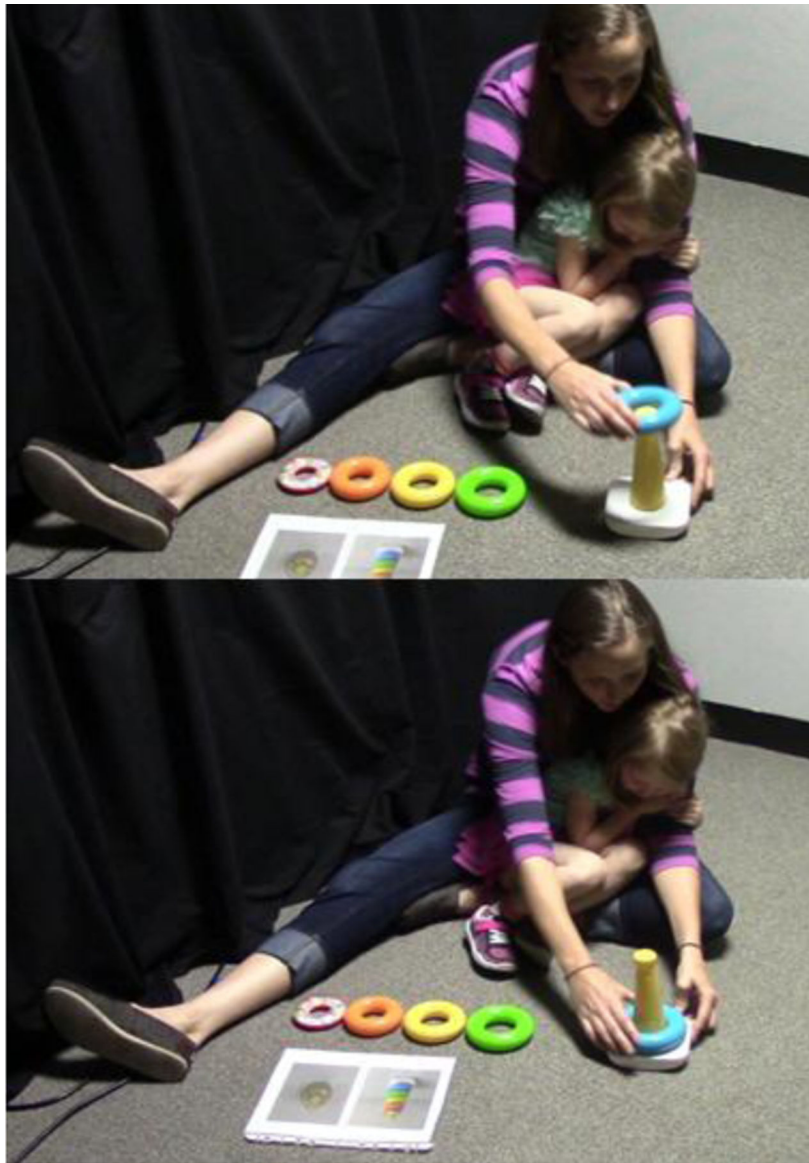Photographs of instructions and assembled complex toy, for use by parents as a reference.

Figure 3.
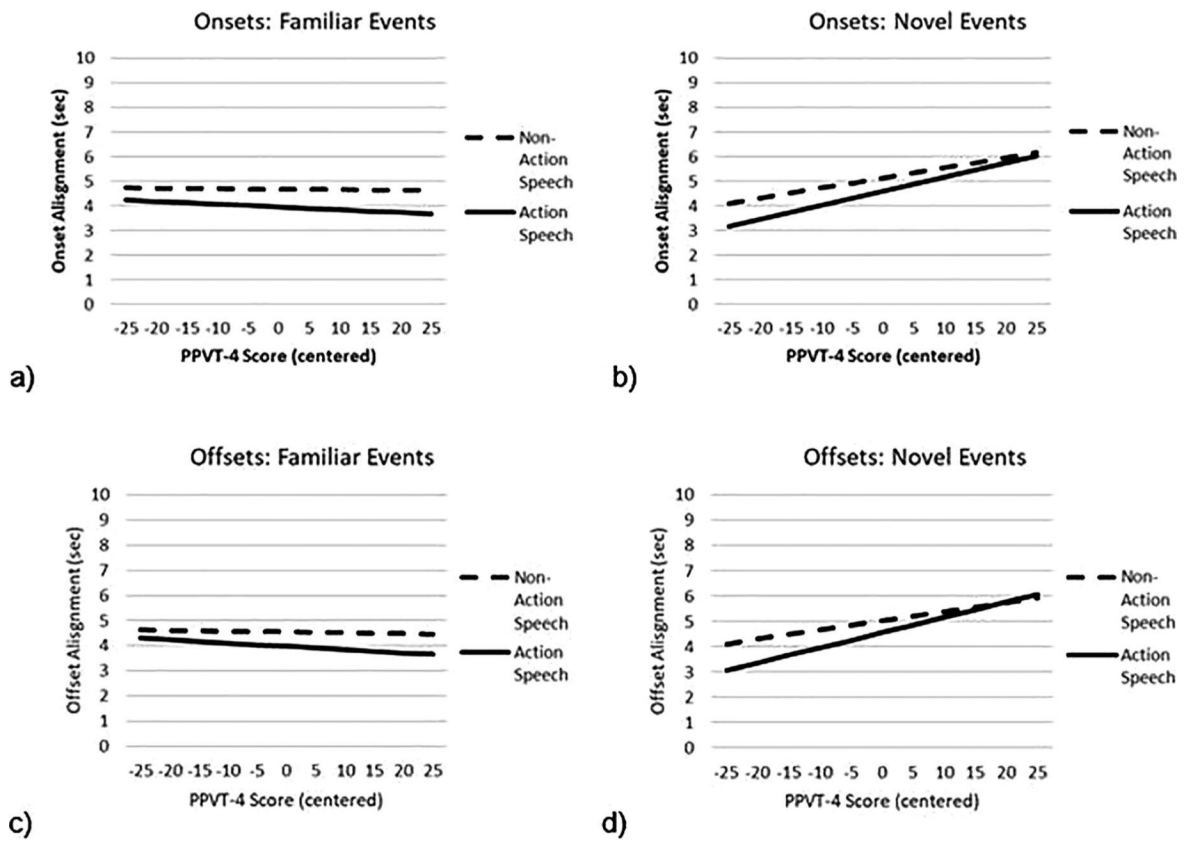Depictions of an onset (top) and offset (bottom) of action.

Figure 4.
Effect of Receptive Vocabulary (PPVT) on onset and offset alignment as a function of novelty and speech type.

Table 1

Percentages of speech segments in each category across tasks

Utterance Type	Example	Preschoolers (Current Study)		Infants (Meyer et al., 2011)
		Jungle Toy	Cups/Rings	
Action Description	*Now turn it over*	8.4%	9.1%	57.3%
Goal Setting	*So we'll put this right here*	10.9%	7.5%	7.3%
Completion	*Good job*	2.2%	1.1%	1.9%
Attention Getting	*Ohh look at that*	4.2%	4.3%	21.4%
Mixed/all others	*I don't know about that*	74.4%	78.1%	12.2%