Examining the multimodal effects of parent speech in parent-infant interactions

Sara E. Schroer (seschroe@iu.edu)
Linda B. Smith (smith4@indiana.edu)
Chen Yu (chenyu@indiana.edu)
Department of Psychological & Brain Sciences, Indiana University
1101 East 10th Street, Bloomington, IN 47405 USA

Abstract

Parental input in the form of visual joint attention is hypothesized to serve a critical role in the development of infant attention, acting as a training ground by scaffolding an infant's ability to sustain visual attention in real-time. We extended this hypothesis by studying the effects of parent speech on infant visual and manual attention. Thirty-four toddlers and their parents participated in a free-play study while wearing head-mounted eye trackers. Infant multimodal behaviors were measured in four ways: visual attention, manual action, hand-eye coordination, and joint visual attention with their parent. Overall, we found that longer durations of attention were accompanied by parent speech. Moreover, sustained attention, defined as behaviors lasting 3s or more, almost always occurred with parent speech. Individual differences in parent-infant coordination were also explored. These results suggest that parent-infant interactions create multimodal opportunities for infants to practice sustaining attention.

Keywords: attention, children, cognitive development, eye-tracking, interactive behavior

Introduction

Infants are active learners – they seem to be self-motivated to explore and make predictions about their world. Early development is not solely an individual process, however – it is also embedded in a highly social context as young infants are taught and supported by caregivers. Parents provide scaffolding to their infants in many different ways and in many different contexts, such as recruiting the child's attention, reducing degrees of freedom, and providing demonstrations (Wood, Bruner, & Ross, 1976). Parent scaffolding has been shown to support the development of executive functioning (Bibok, Carpendale, & Müller, 2009) and verbal skills (Smith, Lamdry, & Swank, 2000). In early language learning, parents use infant-directed speech (Thiessen, Hill, & Saffran, 2005), intersensory redundancy (Gogate & Bahrick, 1998), and selective labeling of objects based on infant behaviors (Pereira, Smith, & Yu, 2014) to support word learning. The idea of parental scaffolding has even been adapted by robotics and AI researchers to build a robotic arm that can learn grasp affordances (Ugur, Nagai, Celikkanat, & Oztop, 2015). Understanding how the mature partner influences the sensorimotor experiences and actions of the young infant to support early development and learning is a key question in cognitive development.

Recent work by Yu & Smith (2016) revealed significant effects of parent behaviors on an infant's capacity for sustained attention. In the study, infants and their parents sat at a table while playing with a set of novel toys. Using head-mounted eye tracking, the authors identified moments when parents and infants jointly attended to (or shared attention to) an object and when infants sustained attention on the same object for at least 3s. When the dyad engaged in joint attention, the duration of the infant's sustained attention bout significantly increased, suggesting that 12-month-old infants' ability to sustain attention is scaffolded by parent attention.

Built upon this finding, a recent study (Suarez-Rivera, Smith, & Yu, in press) provided evidence that the social scaffolding effects from parents are not only limited to parent looking behavior. When parent visual attention was accompanied by other types of parent actions – such as talking and manual actions on objects – infants' sustained attention was further improved. Similarly, this redundancy of parent behaviors has been shown to promote joint attention with younger infants (3-to-11-months-old; Deák, Krasno, Jasso, & Triesch, 2018). In parent-infant interactions, both social partners generate various actions moment-by-moment to create multimodal dependencies of looking, talking, and touching, both within the infant's own system and between the two partners. If multimodal behaviors from parents have effects on infants' visual attention, then parent behaviors may also have effects on other, multimodal infant behaviors. The overarching hypothesis in the present study is that parent speech has cascading effects on not only infant visual attention but a suite of multimodal behaviors in parent-infant interactions.

We chose parent speech to study parent scaffolding because it plays a critical role in early communication and early language development. Hart & Risley (1995) famously demonstrated that the amount parents talk to infants is predictive of the varying language abilities of 3-years-olds in different socioeconomic strata. Subsequent studies show both quality and quantity of parent speech is predictive of later language outcomes (Tamis-LeMonda, Bornstein, & Baumwell, 2001; Hirsh-Pasek et al., 2015). While past research has focused on how parent speech and its linguistic properties, such as infant-directed speech and wh-questions

in speech, predict later child vocabulary size (e.g., Rowe, 2012; Weisleder & Fernald, 2013), the present study will examine the non-linguistic effects of parent speech.

Studying the role of parent speech in the micro-level dynamics of parent-infant interactions is a crucial next step in the field. Although joint visual attention facilitates infant sustained attention (Yu & Smith, 2016), we know that joint attention during toy play does not result from infants following the gaze of their caregivers and does not require any overt bid for the partner's attention (Yu & Smith, 2017a; Deák et al., 2018). During play, adult object manipulations (often coupled with other behaviors, such as speech), are the most promotive of joint attention (Deák et al., 2018). However, maternal speech is tightly linked to object manipulation and occurs frequently in an interaction as a response to infants' visual attention to objects, handling of multiple objects, and vocalizations (Chang, de Barbaro, & Deák, 2017). Parents verbally respond to a suite of multimodal infant behaviors, potentially serving as scaffolding for not only joint attention but also other forms of sustained attention.

To test the multimodal effects of parent speech, we chose four infant behaviors from parent-infant interactions that have been shown to be important in early development: 1) visual attention; 2) manual action; 3) hand-eye coordination; and 4) joint attention. Visual attention was chosen because infant sustained visual attention predicts later language learning and cognitive development (Kannass & Oakes, 2008; Lawson & Ruff, 2004; Yu, Suanda, and Smith, 2018). Manual action was chosen because motor skills, including object exploration, are known to play a major role in early language development (Iverson, 2010). Hand-eye coordination was chosen because both infants and parents attend to their own actions and their partner's object manipulations in free play (Yu & Smith, 2017b). Lastly, joint attention between infant and parent was chosen because dyadic differences in the frequency with which parents and children engage in episodes of joint attention predict individual differences in child vocabulary size (Tomasello & Todd, 1983). For all of these behaviors, we will be looking at sustained attention, defined as when infants attend to an object for a long duration (e.g., greater than 3 seconds). While sustained visual attention is known to predict later outcomes (Kannass & Oakes, 2008; Lawson & Ruff, 2004, Yu, Suanda, and Smith, 2018), the ability to sustain attention in other modalities has not been explicitly studied.

The present study had two goals. In Study 1, we examined the multimodal effects of parent speech by measuring the durations of the four types of multimodal behaviors when they were accompanied by parent speech and comparing with when they were not. We hypothesized that parent talk increases infants' ability to sustain their multimodal behaviors. In Study 2, we focused on individual differences in parent speech, given that some parents generated more speech than others did in free play. We examined whether varying amounts of parent speech create different effects on infants' multimodal behaviors.

Methods

Thirty-four toddlers (mean age = 18.67mos [range: 12.3-24.3]; female = 16) and their parents participated in a study on naturalistic parent-infant interactions during free play. An additional 5 dyads were included in the experimental data set but were excluded from the current analyses due to missing parent eye-tracking (n = 2) and non-transcribable speech (n = 3).

Data Collection

Parents and infants played with 24 toys on a carpeted floor in a playroom for an average of 7.15 minutes (range 3.93-11.64). At the beginning of the play session, the toys were randomly spread out across the floor. Parents were instructed to play as they would at home and that they could sit in any orientation (behind, next to, in front of their infant), but were asked to keep their infant sitting on the floor due to the eye tacker's cable.

During the play session, both parent and infant wore a head-mounted eye tracker (Positive Science LLC). The eye tracker system used a scene camera on the participant's forehead to record images from the wearer's perspective with a visual field of 108°. A second, infrared camera pointed to the participant's right eye to record saccades and fixations. Both cameras sampled at a rate of 30Hz. The infant's eye tracker was affixed to a hat and the parent wore their eye tracker like a pair of glasses. Additional cameras were placed in the room to capture traditional third-person views of the dyad (Figure 1).

Figure 1: Experimental set-up (left) and the infant's first-person view, the cross-hair indicates infant gaze (right).

The experiment was run by two researchers. The session began by one researcher placing the eye tracker on the parent and adjusting the scene and eye cameras, while the other researcher engaged with the infant. Afterwards, both researchers worked together to place the eye tracker on the infant. One researcher, and the parent, continued to distract the infant with exciting toys (e.g. a pop-up toy that played music) as the other researcher set up the eye tracker on the infant. After both members of the dyad were wearing their eye trackers, the researchers ran a brief calibration procedure. A large board that had lights and produced sounds was placed in front of the infant (approximately 30 cm away). One of the researchers controlled the board and lit up one of the lights

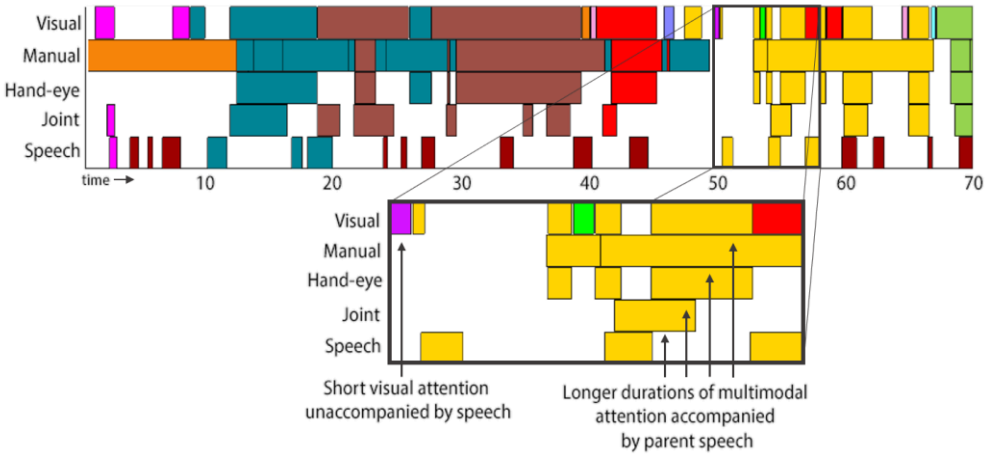

Figure 2: Data streams of infant visual attention, manual action, hand-eye coordination, joint attention, and parent speech over 70s of an interaction. Each block represents a behavioral event and each color represents a different object. The color of parent speech represents the object being named, dark red indicates no naming in that utterance.

until both the parent and infant shifted their gaze to that location. This procedure was repeated for 15 light locations.

The researchers monitored the experiment from an adjoining room. If the infant's eye camera was bumped or moved during play, the researchers reentered, adjusted the camera, and completed an abridged calibration procedure.

Coding and Analyses

Following the experiment, the eye tracking videos from the scene and eye cameras were synchronized and calibrated with a software program to generate a cross-hair that indicated where the participant was looking during each frame of the video (Figure 1). Parent and infant visual gaze were then coded manually using the first-person view (from the scene camera) with the cross-hair overlaid. Using an in-house program, the coder annotated which region of interest (ROI) the cross-hair overlapped with during a fixation. There were 25 ROIs – one for each toy and the social partner's face.

The scene cameras and third-person views were then used to annotate the objects being handled by a participant, frame-by-frame, in an in-house program. If a hand was touching an object, the object was considered "in hand". Participants' left and right hands were coded separately.

Parent speech was transcribed using Audacity at the utterance level. There was no minimum length for an utterance, but separate utterances had to be 400ms or more apart (otherwise they were collapsed together). All parent talk and vocal play (like saying "vroom-vroom" or making a crashing sound) were considered speech. Due to the 400ms criteria, chunks of speech that would be considered sentences could be split apart and separate sentences could be counted as one utterance.

In the current studies, we were interested in five behaviors: infant visual attention, manual action, hand-eye coordination, dyadic joint attention, and parent speech (Figure 2). **Visual attention** was defined as all infant fixations to the 25 ROIs. **Manual action** was similarly defined as all instances of the infant touching an object with either or both hands. **Hand-eye coordination** was defined as moments when the infant looked at and handled the same object, for any duration of time. **Joint attention** between the parent and infant was defined as any moment when the parent's and infant's visual attention fell on the same ROI. All parent utterances were counted as **speech**.

For all four multimodal behaviors (visual attention, manual action, hand-eye coordination, and joint attention) sustained attention was defined as a behavior lasting 3 seconds or longer (to match the previously used definition in Yu & Smith, 2016).

To test the effects of parent speech, we categorized each attention bout as "with speech", if the onset of a parent utterance began after the onset of the attention bout and before the offset of the bout. Other attention bouts, without any overlap with a parent utterance, were categorized as "without speech". With this definition, we can measure the effects of parent speech by comparing attention bouts in the two categories.

Study 1: Multimodal Effects of Parent Support

In Study 1, we tested the relationship between parent speech and the four multimodal measures of infant behavior. Corpus-level analyses were used to compare the durations of all multimodal attentional bouts with and without speech.

Each modality was analyzed separately using mixed effects models to predict the duration of a bout by whether it was accompanied by speech, with subject and attended object as random effects. Each full model was then compared to a null model, with intercept and random effect of object only, using Chi-Square difference tests. All four multimodal behaviors were found to last longer when co-occurring with speech (Figure 3, Table 1).

To specifically test whether parent speech co-occurs with *sustained* attention, an infant behavior known to predict later outcomes (Yu et al., 2018), similar models were used to analyze the subset of sustained attention bouts lasting 3s or more. Bouts of sustained attention of each multimodal

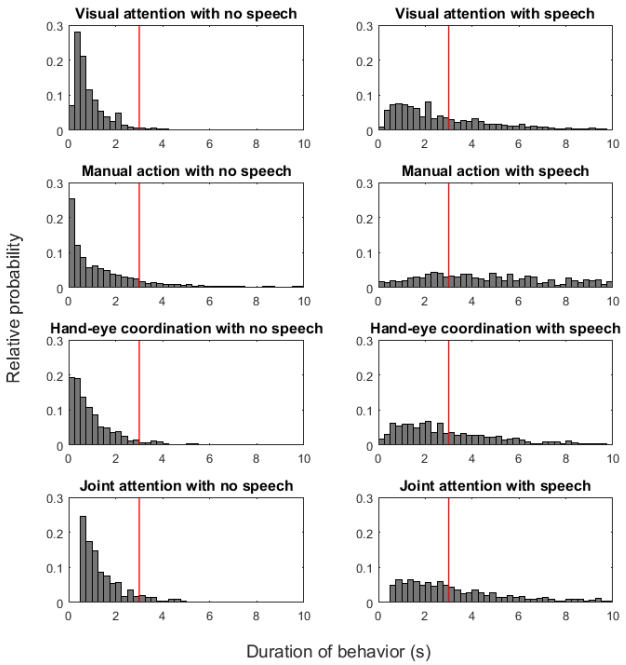

Figure 3: Durations of behaviors without speech (left column) and with speech (right column). Red line indicates the 3-second threshold for sustained attention.

behavior were also longer, and more likely to occur, with speech (Table 1).

The mean duration of **visual attention** bouts with speech was more than 3 times longer than the mean duration of bouts without speech ($M_{with\text{-}speech}$=3.769s, $M_{w/o\text{-}speech}$=1.000s). When we examined the subset of sustained attention bouts that were longer than 3s, sustained attention bouts increased in duration by 50% when accompanied by parent speech ($M_{with\text{-}speech}$=7.196s, $M_{w/o\text{-}speech}$=4.842s).

The mean duration of **manual action** bouts with speech was nearly 8 times longer than the mean duration of bouts without speech ($M_{with\text{-}speech}$=11.187s, $M_{w/o\text{-}speech}$=2.471s). Sustained manual action bouts that co-occurred with parent speech were close to double the duration of bouts without parent speech ($M_{with\text{-}speech}$=13.840s, $M_{w/o\text{-}speech}$=8.629s).

The mean duration of **hand-eye coordination** bouts with speech was nearly 4 times longer than bouts without speech ($M_{with\text{-}speech}$=4.004s, $M_{w/o\text{-}speech}$=1.084s). Sustained hand-eye coordination bouts increased in duration by 50% when accompanied by parent speech ($M_{with\text{-}speech}$=6.961s, $M_{w/o\text{-}speech}$=4.587s).

Lastly, the mean duration of parent-infant **joint attention** with parent speech was more than 2 times longer than joint attention without speech ($M_{with\text{-}speech}$=4.284s, $M_{w/o\text{-}speech}$=1.476s). As with the other behaviors, the duration of sustained attention events with parent speech was longer than sustained attention events without parent speech ($M_{with\text{-}speech}$=6.967s, $M_{w/o\text{-}speech}$=4.071s).

Across all four multimodal behaviors, the duration of infant attention is extended when the bout is accompanied by parent speech. Moreover, when we specifically examined sustained attention bouts, we saw that not only is sustained attention substantially more likely to occur with parent speech, but that bouts of sustained attention with parent speech are significantly longer.

Study 2: Individual Differences

If we view the parent as a coach, training their infant to engage in sustained attention (Yu & Smith, 2016), then we should see differences in how the dyads practice, since different coaches may have different coaching styles. Parents may vary in the "drills", or amount of speech, they use in practice. If so, infants may react differently to parent's coaching which will influence how much they "score" in sustained attention. To understand the individual differences in the coordination of parent speech and infant attention, we examined whether more or less parent talk has different effects on the infant's ability to sustain attention.

Parents varied in how much they spoke to their infants. The average parent produced 16.819 utterances/minute (*SD* =3.844), though the quietest parent only spoke 9.597 times/minute and the most "talkative" parent generated 25.144 spoken utterances per minute. To test the relationship between parent speech and infant sustained attention, we divided the subjects into two groups based on a median split (median = 16.814 utterances/min). Parents in the high frequency speech group produced 19.905 utterances per minute while parents in the low frequency speech group produced on average 13.734 utterances per minute. The low frequency and high frequency groups did not differ in the mean duration of parent utterances (M_{low}=1.309s, M_{high}=1.330s, p = 0.871), suggesting low frequency parents

Table 1: Duration of multimodal behaviors with and without speech

		instances with speech			instances without speech			statistical comparison			
		# of bouts	mean dur	sd	# of bouts	mean dur	sd	beta	p-value	95% CI	null model comparison
visual	overall	2439	3.769	4.938	4053	1.000	1.283	2.741	< 0.001	[2.597 2.885]	$\chi2 = 1262.300$, $p < 0.001$
attention	sustained	986	7.196	5.227	171	4.842	3.971	2.654	< 0.001	[1.839 3.466]	$\chi2 = 40.181$, $p < 0.001$
manual	overall	1736	11.187	15.821	1788	2.471	4.989	8.468	< 0.001	[7.710 9.229]	$\chi2 = 448.450$, $p < 0.001$
action	sustained	1355	13.840	16.984	362	8.629	8.536	4.768	< 0.001	[3.047 6.491]	$\chi2 = 29.238$, $p < 0.001$
hand-eye	overall	920	4.004	4.463	1411	1.084	1.206	2.881	< 0.001	[2.640 3.121]	$\chi2 = 496.190$, $p < 0.001$
coordination	sustained	416	6.961	5.234	88	4.587	1.731	2.708	< 0.001	[1.617 3.791]	$\chi2 = 23.289$, $p < 0.001$
joint	overall	1033	4.284	4.430	865	1.476	1.047	2.796	< 0.001	[2.505 3.087]	$\chi2 = 325.500$, $p < 0.001$
attention	sustained	506	6.967	5.046	76	4.071	1.109	3.681	< 0.001	[2.583 4.767]	$\chi2 = 42.025$, $p < 0.001$

Table 2: Sustained attention in dyads with low frequency and high frequency parent speech

	low frequency group			high frequency group			statistical comparison			
	# bouts	mean dur	sd	# bouts	mean dur	sd	beta	p-value	95% CI	null model comparison
sustained visual attention	401	7.636	6.125	585	6.894	4.490	-0.523	0.121	[-1.185 0.136]	$\chi2 = 2.425$, p = 0.119
sustained manual action	604	14.946	18.215	751	12.950	15.881	-2.424	0.008	[-4.220 -0.619]	$\chi2 = 6.928$, p = 0.008
sustained hand-eye coordination	199	7.404	5.918	217	6.555	4.500	-0.822	0.144	[-1.838 0.195]	$\chi2 = 2.511$, p = 0.113
sustained joint attention	198	7.865	6.571	308	6.389	3.649	-1.165	0.009	[-2.035 -0.298]	$\chi2 = 6.911$, p = 0.009

were truly producing less speech, not just fewer, longer utterances. Therefore, the durations of spoken utterances in the two groups would not be a factor to influence infant's attention.

We then compared the durations of sustained attention bouts produced by infants in the low frequency and high frequency groups. To directly measure the effects of parent speech, we only analyzed sustained attention bouts that were accompanied by parent speech. As in Study 1, each type of multimodal behavior was analyzed separately using mixed effects models, with object attended to as a random effect, and then compared to a null model with intercept and random effect of object only.

For manual actions and joint attention, we found that the duration of attentional bouts was longer for infants in the low frequency group (Table 2). The mean duration of sustained **manual action** bouts was 2 seconds longer in the low frequency group (M_{low}=14.946s, M_{high}=12.950s). The mean duration of sustained **joint attention** bouts was more than a second longer in the low frequency group (M_{low} =7.865s, M_{high} =6.389s). There were no differences between the low frequency and high frequency groups in the durations of sustained visual attention or hand-eye coordination.

We present evidence of two groups of dyads, classified by how much speech parents produced in an interaction. These two groups coordinate their attention in different ways – in the low frequency group, there are less occurrences of speech-attention overlap in all four types of behavior. But, when parent speech co-occurs with manual action or joint attention, infants in the low frequency group had significantly longer durations of sustained attention than infants in the high frequency group. This finding suggests two possible phenomena: 1) parents who talked less may be more selective in when they choose to talk; or 2) infants whose parents talked less are more responsive when their parent does talk.

Discussion

With the current studies, we examined the dynamics of parent-infant interactions, specifically the role of parent behaviors in influencing infant attention. We demonstrated that the duration of infants' visual attention is longer when accompanied by parent speech, extending prior work that focused primarily on parent's visual attention (Yu & Smith, 2016). Furthermore, we measured the relationship between parent speech and multiple infant sensory-motor behaviors beyond visual attention – manual action, hand-eye coordination, and joint attention – and found a similar coordination between parent speech and infant sustained attention. Sustained attention of each of these multimodal behaviors is more likely to occur, and lasts longer, when accompanied by parent speech.

There were, however, individual differences in the observed parent-infant coordination. Parents that spoke less during the interaction had infants with longer durations of sustained manual attention and dyadic joint attention, relative to their talkative peers. This relationship could have two (non-mutually exclusive) causes. One possible explanation is that infants with less talkative parents are more responsive to their parent's speech. Using the coaching analogy, those infants may not get coaching signals very often and therefore they respond to the signals better when they receive them. Another possible explanation is that parents that talk less are more selective in when they choose to talk. Rather than "coach" all the time, irrespective of their infant's attentional state, these parents may find optimal moments to support their infants. Regardless, it suggests that dyads with less talkative parents are still having high-quality practices. Parents that talk more can scaffold their infant's ability to sustain attention more frequently, creating more opportunities for the infant to score. Dyads with less talkative parents, however, appear to employ more effective drills during their practices – even though these infants "score" less, the durations of their sustained manual action and joint attention bouts are longer. Thus, there are two different pathways through which parents can support their infants. Future research needs to examine potential qualitative and quantitative differences between the two pathways used by more and less talkative parents, and how different dyads adjust and adapt to different interaction patterns based on the history of their experiences.

Our results present evidence of a multimodal sustained attention training ground. The coupling of parent speech and infant attention suggests that the more infants sustain their attention, the more parents respond to it, giving the infant even more time to practice. Coaching improves an infant's ability to sustain attention, increasing the time an infant can learn about the object's properties (Ruff, 1986) and creating more opportunities for the parent to talk about and label objects (Yu & Smith 2012; Pereira et al., 2014), fostering a developmental cascade yielding higher language outcomes (Yu et al., 2018). We are also among the first to study sustained attention beyond the visual modality. How

sustained manual attention, hand-eye coordination, and joint attention relate to later outcomes is still an open question to be investigated further, especially given the individual differences seen in manual attention and joint attention.

To better understand the parental scaffolding of sustained attention, we need to study the infant behaviors that elicit parent responses. It is unlikely that parents are randomly speaking during an interaction. Rather, they are responding contingently to certain infant behaviors and following non-linguistic cues like gaze, object manipulation, gesturing, smiling, and more. To create successful object labeling moments, a parent and infant need to couple their behavior so that they are attending to and naming the same object. Infants need to sustain their attention to the object long enough for the parent to provide a label, which requires the infant exhibiting behaviors indicating a readiness to learn (e.g. object-directed vocalizations; Goldstein, Schwade, Briesch, & Syal, 2010) and parents being able to follow these behaviors. One way to address this question is to analyze the temporal dynamics of parent-infant interactions. Measuring parent and infant behaviors seconds before a parent utterance and the subsequent behavioral changes after the utterance will provide further insight into how dyads coordinate their behaviors and influence one another. One possibility is that there are "signatures" that reliably predict whether a parent utterance leads to sustained attention and successful object-label mappings. Studying the temporal dynamics of infant looking and object handling before and after a naming moment revealed developmental changes from 4 to 9 months (Chang et al., 2017), positioning this form of analysis as a pertinent future direction.

Conclusion

Previous work has shown that joint visual attention supports an infant's ability to sustain attention. We extended these findings by measuring the multimodal effect of parent speech on infant visual attention, manual action, hand-eye coordination, and joint attention. When multimodal attention is accompanied by parent speech, the infant sustains their attention for longer periods of time, creating a rich training ground for early development.

Acknowledgements

This research was funded by National Institutes of Health Grant R01HD074601 and R01HD093792 to CY. SES was supported by NSF GRFP 1342962.

References

Bibok, M. B., Carpendale, J. I., & Müller, U. (2009). Parental scaffolding and the development of executive function. *New directions for child and adolescent development*, 2009(123), 17-34.

Chang, L., de Barbaro, K., & Deák, G. (2016). Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: Longitudinal changes from 4 to 9 months. *Developmental neuropsychology*, 41(5-8), 342-361.

Deák, G. O., Krasno, A. M., Jasso, H., & Triesch, J. (2018). What leads to shared attention? Maternal cues and infant responses during object play. *Infancy*, 23(1), 4-28.

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of experimental child psychology*, 69(2), 133-149.

Goldstein, M. H., Schwade, J., Briesch, J., & Syal, S. (2010). Learning while babbling: Prelinguistic object-directed vocalizations indicate a readiness to learn. *Infancy*, 15(4), 362-391.

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday life of America's children.* Baltimore, MD: Paul Brookes.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., et al. (2015). The contribution of early communication quality to low-income children's language success. *Psychological science*, 26(7), 1071-1083.

Iverson, J. M. (2010). Developing language in a developing body: The relationship between motor development and language development. *Journal of child language*, 37(2), 229-261.

Kannass, K. N., & Oakes, L. M. (2008). The development of attention and its relations to language in infancy and toddlerhood. *Journal of cognition and development*, 9(2), 222-246.

Lawson, K. R., & Ruff, H. A. (2004). Early focused attention predicts outcome for children born prematurely. *Journal of developmental & behavioral pediatrics*, 25(6), 399-406.

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic bulletin & review*, 21(1), 178-185.

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child development*, 83(5), 1762-1774.

Ruff, H. A. (1986). Components of attention during infants' manipulative exploration. *Child development*, 105-114.

Smith, K. E., Landry, S. H., & Swank, P. R. (2000). Does the content of mothers' verbal stimulation explain differences in children's development of verbal and nonverbal cognitive skills?. *Journal of school psychology*, 38(1), 27-49.

Suarez-Rivera, C., Smith, L. B. & Yu, C. (in press) Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental psychology.*

Tamis-LeMonda, C. S., Bornstein, M. H., & Baumwell, L. (2001). Maternal responsiveness and children's achievement of language milestones. *Child development*, 72(3), 748-767.

Thiessen, E. D., Hill, E. A., & Saffran, J. R. (2005). Infant-directed speech facilitates word segmentation. *Infancy*, 7(1), 53-71.

Tomasello, M., & Todd, J. (1983). Joint attention and lexical acquisition style. *First language*, 4(12), 197-211.

Ugur, E., Nagai, Y., Celikkanat, H., & Oztop, E. (2015). Parental scaffolding as a bootstrapping mechanism for learning grasp affordances and imitation skills. *Robotica*, 33(5), 1163-1180.

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological science*, 24(11), 2143-2152.

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, 17, 89–100.

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, 125(2), 244-262.

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current biology*, 26(9), 1235-1240.

Yu, C., & Smith, L. B. (2017a). Multiple sensory-motor pathways lead to coordinated visual attention. *Cognitive science*, 41, 5-31.

Yu, C., & Smith, L. B. (2017b). Hand–Eye Coordination Predicts Joint Attention. *Child development*, 88(6), 2060-2078.

Yu, C., Suanda, S. H., & Smith, L. B. (2018). Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. *Developmental science*, e12735.