CHAPTER

Using head-mounted eye-trackers to study sensory-motor dynamics of coordinated attention

4

Chi-hsin Chen[a,*], Claire Monroy[a], Derek M. Houston[a,b], Chen Yu[c]

[a]*Department of Otolaryngology – Head and Neck Surgery, The Ohio State University, Columbus, OH, United States*

[b]*Nationwide Children's Hospital, Columbus, OH, United States*

[c]*Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, United States*

**Corresponding author: Tel.: +1-614-366-1028, e-mail address: chi-hsin.chen@osumc.edu*

Abstract

In this chapter, we introduce recent research using head-mounted eye-trackers to record sensory-motor behaviors at a high resolution and examine parent-child interactions at a micro-level. We focus on one important research topic in early social and cognitive development: how young children and their parents coordinate their visual attention in social interactions. We start by introducing head-mounted eye-tracking and recent studies conducted using this method. We then present two sets of novel analysis techniques that examine how manual actions of parents and children with and without hearing loss contribute to their attention coordination. In the first set of analyses, we investigated different pathways parents and children used to coordinate their visual attention in toy play. After that, we used Sankey diagrams to represent the temporal dynamics of parents' and children's manual actions prior to and during coordinated attention. These two sets of analyses allowed us to explore how participants' sensory-motor behaviors contribute to the establishment and maintenance of coordinated attention. More generally, head-mounted eye-tracking allows us to ask new questions and conduct new analyses that were not previously possible. With this new sensing technology, the results here highlight the importance of understanding early social interaction from a multimodal, embodied view.

Keywords

Infant eye-tracking, Head-mounted eye-trackers, Coordinated attention, Parent-child interactions, Pathway analyses

1 Eye-tracking methods in infant research

Over the past two decades, screen-based eye-tracking methods have been widely used in infant research to study learning, perception, representation, memory and other social-cognitive processes (Gredeback and von Hofsten, 2004; Johnson et al., 2003; McMurray and Aslin, 2004; Navab et al., 2012; Richmond and Nelson, 2009; Senju and Csibra, 2008). These studies have yielded considerable insight into infant development, including what infants perceive or know, what they can do, and how they learn. However, one drawback of this approach is that the young participants are usually required to sit (still) in front of a screen, look straight toward the screen, and watch what is being presented on the 2-D screen. This setup also requires minimal movement from the infants (for a few exceptions, see Corbetta et al., 2012; Wiener et al., 2018). Watching a screen while sitting still with minimal head movement is largely different from infants' and young toddlers' daily experiences, as they do not just *watch and learn*; they also *act and learn*. They usually learn through using their own actions to explore the environment and to interact with social partners. To address the limitation of screen-based eye-tracking, researchers have started using lightweight head-mounted eye-trackers to study infants' attention and behaviors while they are on the move (Franchak et al., 2011; Kretch et al., 2014), and when they are interacting with social partners (Yu and Smith, 2013, 2017). Head-mounted eye-tracking provides a way to investigate infant motor, social, cognitive, and language development in ecologically valid settings. Importantly, unlike traditional observational studies that record infant behaviors or parent-child interactions from third-person views, head-mounted eye-tracking studies provide infants' first-person perspective when they act and interact with the environment or with other people. This novel method allows researchers to ask new questions and conduct new analyses that were not previously possible.

The goal of this chapter is to provide an overview of the current state of knowledge on head-mounted eye-tracking in infant research and demonstrate how to use advanced techniques to analyze eye-tracking data recorded from head-mounted eye-trackers. In Section 1, we will first introduce the general setup of head-mounted eye-tracking systems (Section 1.1) and then review two lines of infant research that currently use head-mounted eye-tracking (Section 1.2). In the second section, we will use two separate sets of analyses as examples to demonstrate how to analyze eye-tracking data along with manual action data to answer critical research questions in parent-child play contexts. We will start by giving an overview of the study (Sections 2.1 and 2.2) and then present analyses on different pathways parents and children use to coordinate their visual attention in toy play (Section 2.3). After that, we will conduct novel analyses that examine the temporal dynamics of children's hand actions prior to and during coordinated attention (Section 2.4). We also suggest how these analysis techniques can be used to address other research questions. In the last section, we will discuss implications of the findings.

1.1 Head-mounted eye-tracking

Currently, there are several head-mounted eye-tracking systems that are commercially available. Some systems have headgears readily available for infants while some others can be modified to attach to a custom-made cap or headband, as shown in Fig. 1 (for more details on the selection of eye-tracking systems, see Slone et al., 2018). A head-mounted eye-tracker is composed of two cameras, a scene camera facing outward to record the participant's first-person view and an eye camera facing inward to record the participant's eye movements (Fig. 1). Some eye-trackers are wired to a computer while others can store data on a light-weighted recording device (e.g., a smart phone) and be placed on the child, such as in a pocket or a small backpack. Because head-mounted eye-trackers only capture the first-person view, additional third-person view cameras are usually used to obtain a wider view of the environment or social interaction. These third-person view cameras can be placed on the side or overhead. They help capture the interactions or actions that fall outside the participants' first-person camera views; and the recordings can be later used for data coding (e.g., coding of participants' actions).

Either before or after the experimental task(s), the experimenters must collect data for calibration. A common method with children is to draw participant's attention to several specific locations by using a small, attractive object or a laser pointer

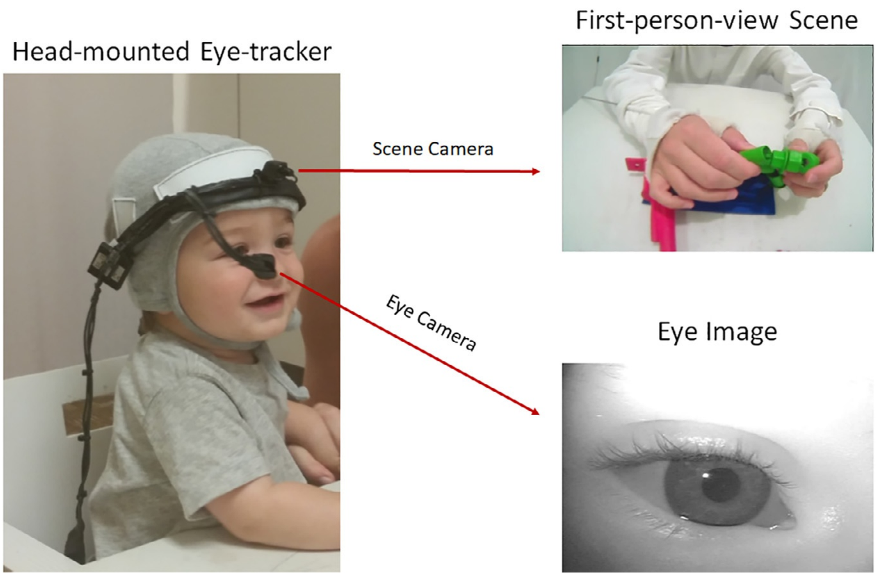

FIG. 1

A head-mounted eye-tracker is composed of a scene camera, which records the participant's first-person view, and an eye camera, which records the eye movements.

(for more details on the calibration procedure see Slone et al., 2018). Specialized software (e.g., Yarbus from Positive Science, LLC) is used that can map the changing positions of the pupil and corneal reflection recorded by the eye camera to corresponding locations in the first-person view scene. The calibrated videos are then used for data annotation and data analysis.

Head-mounted eye-trackers have been used with both typically developing infants and clinical populations, such as children with hearing loss (e.g., Chen et al., 2019a,b, 2020) or children with autism (Yurkovic et al., 2020). In terms of the eye-tracking system or technology, there is usually no specific requirement needed for clinical populations. However, for children with hearing loss who use cochlear implants or hearing aids, it is necessary to check whether the material or the position of the cap or headband to which the eye-tracker is attached would interfere with the placement or transmission of their hearing device(s).

1.2 Infant research using head-mounted eye-trackers

Up to now, head-mounted eye-tracking has been successfully used to advance two major lines of infant research. The first line focuses on infants' attention during locomotion or active exploration of the environment (Franchak et al., 2011, 2018; Hoch et al., 2019; Kretch and Adolph, 2015, 2017; Kretch et al., 2014). This line of research, for the first time, allows us to see what infants pay attention to when they are on the move. These studies show that infants do not just passively view the world; they *actively* use visual information to guide their own actions or movements (Franchak et al., 2011; Hoch et al., 2019; Kretch and Adolph, 2017; Rachwani et al., 2019). Infants' walking or crawling abilities not only affect how and how far they move, but also their social looking behaviors, such as looking to their caregiver's face (Franchak et al., 2018; Kretch et al., 2014). For example, compared to walking infants, crawling infants are less likely to have their caregiver's face in view. This is likely due to body constraints and motor costs, as it is more effortful for a crawling infant to lift his/her head to look at a caregiver's face than for a walking infant.

The second line of research focuses on infants' attention during social interactions, such as parent-child toy play (McQuillan et al., 2020; Slone et al., 2019; Suanda et al., 2019; Suarez-Rivera et al., 2019; Yu and Smith, 2013, 2016, 2017; Yu et al., 2019). For decades, people have been using video recordings or human observations to investigate different aspects of parent-child interactions (e.g., Bornstein et al., 1992; Masur and Gleason, 1980; Tamis-LeMonda and Bornstein, 1989). These studies do not only tell us what parents and children do during interactions, but also what behaviors observed during interaction associate with children's long-term language, cognitive, or social development (Bornstein et al., 1992; Masur and Gleason, 1980; Tamis-LeMonda and Bornstein, 1989). However, these studies record parent-child interactions from third-person perspectives. Head-mounted eye-tracking allows researchers to observe interactions and behaviors from the infants' first-person perspective for the first time. Using head-mounted eye-tracking, researchers have been able to show how infants' and young children's object-directed actions during play contribute to their learning of visual objects and object names (Bambach et al., 2018; McQuillan et al., 2020; Slone et al., 2019). For instance, infants' manual actions with objects create a clear and centered view of objects (McQuillan et al., 2020; Suanda et al., 2019) that reduces referential uncertainty and facilitates learning object labels (Yu and Smith, 2012). It has also been found that infant sustained visual attention to objects during play predicts their language growth (Yu et al., 2019). Importantly, infant sustained attention is supported by parental language input, manual actions, and gaze toward objects (McQuillan et al., 2020; Suarez-Rivera et al., 2019; Yu and Smith, 2016; Yu et al., 2019). These studies highlight the contribution of social interaction to infants' attention development, which has traditionally been considered endogenous rather than social in nature (Yu and Smith, 2016). They also underscore the importance of examining language learning from an embodied perspective. Children's bodily movements and actions influence the moment-by-moment visual input they receive; the visual input, in turn, affects what children learn about objects and their labels.

Another interesting finding from the second line of research is that parents and infants use multiple sensorimotor pathways to establish coordinated attention. Coordinated attention is defined as whenever social partners look at the same object at the same time (Yu and Smith, 2013, 2017). By analyzing the temporal dynamics of manual actions and gaze between typically developing toddlers and their parents, Yu and Smith (2013, 2017) found that gaze-following is not the most frequent pathway leading to coordinated attention. Instead, they tend to follow parents' attention through their manual actions, by looking at the object the parent is touching or handling. This line of research has been extended to clinical populations, including children with hearing loss (e.g., Chen et al., 2019a,b, 2020). Chen et al. (2020) found that, compared to their hearing peers, children with hearing loss used different pathways to achieve coordinated attention with their parents. Unlike typically developing children who mainly followed parents' attention through their manual actions, children with hearing loss relied on both parent gaze directions and manual actions.

1.3 Overview of the remainder of the chapter

In the following, we will use a subset of the data collected in Chen et al. (2020), which focused on the pathways leading to coordinated attention between hearing parents and children with and without hearing loss. We will demonstrate how to study coordinated attention at a fine temporal scale (i.e., a *micro-level*) and discuss different pathways to achieve coordinated attention. We will start by giving a brief overview of the study. Following that, we will first present results showing the pathways children with and without hearing loss used to follow parents' attention. Then we will describe novel analyses that examine the temporal dynamics of participants' manual actions prior to and during coordinated attention. Finally, in the last section, we discuss the implications of our results and future directions.

2 Coordinated attention in parent-child interactions

2.1 Background to the example study

Over the past few decades, a large body of work has shown that attention coordination between infants and caregivers is associated with their language, cognitive and social development (Carpenter et al., 1998; Mundy and Newell, 2007; Tomasello and Farrar, 1986). Gaze-following has been long viewed as an effective way to establish coordinated attention between social partners (Butterworth and Cochran, 1980; Slaughter and McConnell, 2003). As a result, numerous studies have used well-controlled experiments to study infants' ability to follow their social partner's gaze (Brooks and Meltzoff, 2005; Butterworth and Cochran, 1980; Flom and Pick, 2005; Gredeback et al., 2008, 2010; Slaughter and McConnell, 2003). However, data collected from naturalistic interactions has shown that infants rarely look at their parents' face during many daily activities, such as crawling, walking, and object play (Chang et al., 2016; de Barbaro et al., 2016; Deak et al., 2014, 2018; Franchak et al., 2011; Yu and Smith, 2013, 2016, 2017). This finding indicates that parents' gaze information is often not in view during infants' daily activities, and thus makes gaze-following less likely to be the optimal pathway for young infants to coordinate attention with parents. Alternatively, it has been proposed that another possible pathway for young children and infants to follow others' attention is through their hand actions, because our gaze directions usually align with our hand actions during daily activities (Deak et al., 2014, 2018; Yu and Smith, 2013, 2017). For example, Land and Hayhoe (2001) found that in everyday tasks, such as tea making or sandwich making, our visual attention is usually on the target object of our action. Recent evidence has supported this idea by demonstrating that young children and infants often use parents' hand actions to follow their attention in joint play (Chen et al., 2020; Deak et al., 2014, 2018; Yu and Smith, 2013, 2017).

In the following, we will demonstrate how to analyze behavioral and gaze data collected using head-mounted eye-trackers worn by hearing parents and toddlers with and without hearing loss. Classic studies annotate attention coordination at the *macro-level*, or at a temporal resolution of seconds to dozens of seconds (e.g., Tasker et al., 2010). In contrast, here we use high-resolution sensory-motor and gaze data to examine how parents and children coordinate attention at the micro-level—usually at a timescale of fractions of a second. We will use a subset of data collected in Chen et al. (2020) and present two sets of analyses. In the first set (Section 2.3), we examine parents' and toddlers' use of hand and gaze-following pathways—two pathways that have been shown to be used by infants and young toddlers in toy play (Deak et al., 2014, 2018; Yu and Smith, 2013, 2017). The data used for this set of analyses is available for readers interested in getting hands-on analysis experience (see Section 2.3.1). In a second set of analyses (Section 2.4), we will analyze the temporal flow of parents' and children's manual actions prior to and during coordinated attention moments and use Sankey diagrams to represent how they change *over time*. A Sankey diagram is a visualization method traditionally used to represent energy flows and their distributions in different states (Schmidt, 2008). Using this method, we will demonstrate how to define different states based on parent's and child's gaze and examine the temporal change of their hand actions *across different states*. This method showcases different approaches for analyzing gaze and manual data. Importantly, it illustrates the role of social partners' hand actions in coordinating their visual attention.

2.2 Description of the example study

2.2.1 Participants

The participants were 14 parent-toddler dyads. Seven of the toddlers (five female) had severe-to-profound hearing loss (HL group) and were between 24 and 37 months old. The children either had cochlear implants ($n=5$) or wore hearing aids ($n=2$). The other seven children had normal hearing (NH group) and were matched to the HL group in age.

2.2.2 Methods

Parents and toddlers sat across from each other and played with two sets of three toys in an alternating order (Fig. 2). The experiment lasted approximately 6 min. Parents were instructed to play with their children like they normally would at home. Both participants wore a head-mounted eye-tracker that recorded their first-person view and gaze directions. In addition to the eye-trackers, there were two additional cameras that recorded the interaction from third-person perspectives.

2.2.3 Data coding

We coded participants' gaze direction and hand contact with the objects frame-by-frame (30 frames/s). We used the calibrated eye-tracker recordings to code gaze direction (for details of gaze and hand coding, see Chen et al., 2020). For gaze coding,

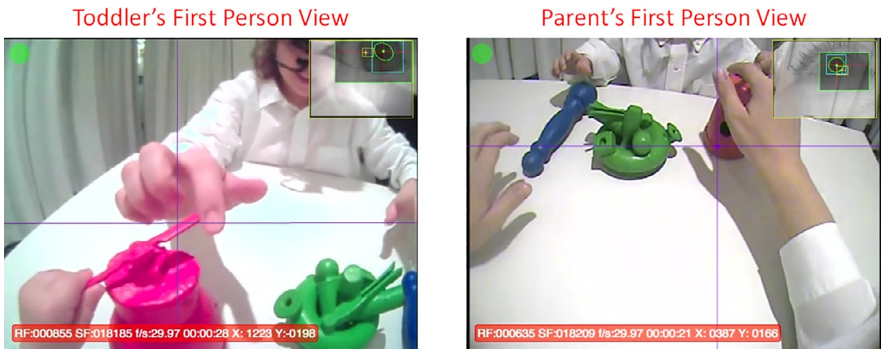

FIG. 2

Example frames from a child's and their parent's head-mounted eye-tracker. The cross-hairs indicate estimated gaze direction.

there were four regions of interest (ROIs): the three objects and the social partner's face. Trained coders coded each frame for the participants' gaze direction—whether it fell in any of the ROIs, and if so, which one. Because the eye-trackers recorded at a sampling rate of 30 Hz, each camera generated approximately 10,800 frames during the 6-min session. It took trained coders approximately 3–5 h to code each participant's gaze. A second coder coded the gaze for 10 participants in the whole dataset. Inter-rater reliability was good, with an average Cohen's kappa of 0.77 (Landis and Koch, 1977).

For hand contact, we used the participants' first-person view camera and third-person view cameras in the room to code whether the participants hands were in contact with any of the objects, and if so, which one. Participants' left hand and right hand were coded separately and then combined in the analyses. It also took trained coders approximately 3–5 h to code each participant's hand. A second coder coded the hand contact for eight participants. Inter-rater reliability was near-perfect with an average Cohen's kappa of 0.94 (Landis and Koch, 1977).

2.3 Gaze and hand-following pathways used by parents and children

In this section, we investigate what leads to coordinated attention by focusing on hand and gaze-following pathways. Coordinated attention (abbreviated as CAtt in Figs. 3 and 5) was defined as any temporal overlap between parent's and child's gaze toward the same object (Fig. 3, CAtt # 1–4). In order to examine what cues parents and children used to follow the other person's attention, we further categorized co-ordinated attention into parent-led and child-led episodes. A coordinated attention episode was defined as parent-led if the parent's gaze to an object started before the child's gaze to the same object (Fig. 3, CAtt # 2 and 4). Similarly, an episode

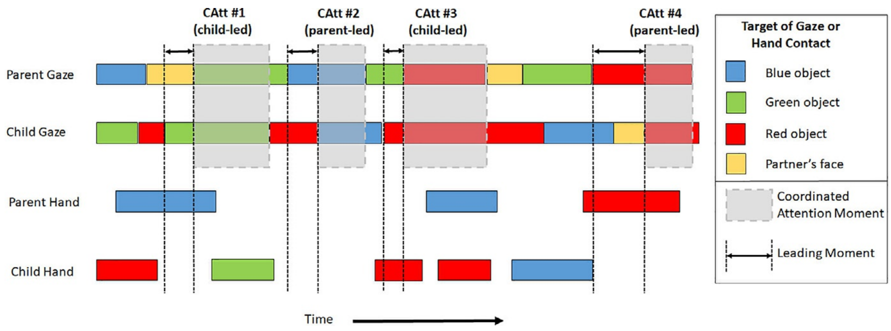

FIG. 3

Representative time series of parent's gaze, child's gaze, parent's hand contact, and child's hand contact. A coordinated attention (CAtt) moment was objectively defined as the temporal window in which the parent and the child look at the same object at the same time. A leading moment was defined as the time in between the coordinated attention leader's gaze onset and the follower's gaze onset.

was defined as child-led if the child's gaze to the target object started before the parent's gaze (Fig. 3, CAtt # 1 and 3). The time window in between the onset of the coordinated attention leader's gaze to the onset of the follower's gaze is termed the *leading moment* (the window marked by the double-headed arrows in Fig. 3). Since one needs to look at the other person's face to perceive their gaze direction, we defined a coordinated attention episode as being established through gaze-following if the follower looks at the leader's face during the leading moment (Fig. 3, CAtt #1 and 4). For coordinated attention episodes without face looking, if the leader's hand(s) touches the target object during the leading moment, we defined this episode as being established through hand-following (Fig. 3, CAtt # 3). It is noteworthy that in our definitions, the gaze-following pathway takes precedence over the hand-following pathway. As long as the follower looks at the leader's face during the leading moment, regardless of whether or not the leader's hand(s) is in contact with the target object, the episode is defined as being established through gaze-following (CAtt # 4). Coordinated attention may also be established when neither gaze nor hand cues are available (CAtt # 2).

2.3.1 Results and discussion

In the following, we will first report the overall coordinated attention patterns in the HL and NH groups. We will then examine parents' and children's use of gaze- and hand-following pathways.

In total, the HL group generated 346 coordinated attention episodes while the NH group generated 372 episodes. There was no group difference in the mean number of episodes from each group or the mean duration of episodes (HL: mean = 2.28 s, SD = 2.04, NH: 2.64 s, SD = 2.58). Both groups had over half of the coordinated attention episodes being led by children (HL: child-led = 61.2%, NH: child-led = 55.0%) and did not show significant group difference between parent-led or child-led proportions.

We next examined the different pathways parents and children used (for readers interested in getting hands-on experience with data analyses, the datasets used in the following analyses can be found at https://mfr.osf.io/render?url=https://osf.io/sg38u/?direct%26mode=render%26action=download%26mode=render). Here, we excluded ambiguous cases in which the follower's hands were already in contact with the target object during the leading moment. This is because the follower may look at the target object because of their own hand actions, rather than using the gaze or hand cues provided by the leader. As shown in Fig. 4A, parents in the HL and NH groups did not differ in the proportions of different pathways they used (Wald $\chi^2 = 1.47$, $P = 0.27$). Overall, parents in both groups preferred the gaze-following pathway over the hand-following pathway. In contrast, children in the HL and NH groups showed a significant difference in the proportions of pathways used (Fig. 4B, Wald $\chi^2 = 11.02$, $P = 0.001$). HL children were more likely to use the gaze-following pathway than NH children. For children in the HL group, they used gaze and hand-following pathways equally often, while children in the NH group used hand-following more frequently than gaze-following. In sum, children in the

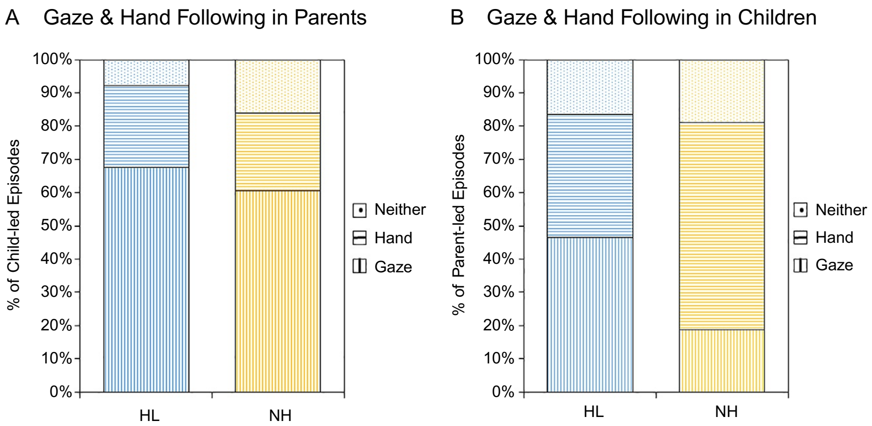

FIG. 4

(A) Gaze and hand-following pathways parents used. (B) Gaze and hand-following pathways children used.

HL group used both gaze direction and hand actions to follow parents' attention, whereas parents mainly relied on children's gaze direction and hearing children mostly relied on parent's hand actions.

The findings that parents mainly relied on children's gaze directions and children with normal hearing mainly relied on parents' hand actions are consistent with previous studies (Chang et al., 2016; de Barbaro et al., 2016; Deak et al., 2014, 2018; Franchak et al., 2011; Yu and Smith, 2013, 2017). In contrast, children with hearing loss relied on both gaze and hand actions as cues to follow. It has been suggested that children with hearing loss check their parents' faces for speech cues (Bergeson et al., 2005; Summerfield, 1992). This face-looking behavior may potentially give them information about their parent's head orientation and gaze direction, because gaze direction usually aligns with head orientation (Bambach et al., 2016; Yoshida and Smith, 2008). They may then use this information to coordinate attention with parents. This also suggests that children with hearing loss use various cues to coordinate attention with their parents.

Here, we have demonstrated how to use participants' gaze patterns to identify the moments in which they coordinate attention and then check back in time to the leading moments and examine the cues that the follower used to join the leader to look at the same object (Fig. 3). This technique requires aligning data from different data streams and using the temporal order of different events to determine the leading and following relationships of these events. Similar analytics techniques can be used to study other aspects of social interactions. For example, one question is whether infants can predict the target object of a goal-directed action, an ability that has been viewed as critical for social interactions (Sebanz and Knoblich, 2009). To do so, one can first use hand movement data to identify the target object of an action and then look back in time and check whether the infant's gaze reaches the target object before the hand (Monroy et al., 2019). If so, that suggests the infant has the ability to predict the goal of an action. Another example is to use similar techniques, but reversing the temporal order, to study whether parents' pointing gesture or utterances about an object *lead* children's attention to the pointed or referenced object. To do so, one can first use hand actions to identify the pointing gestures or speech data to identify the utterances and then check whether children's gaze to a target object follows the gesture or utterance. This type of analysis methods can be used not only to identify leading-following relationships of different events, but also potentially be used to study the causal relations of different events.

Head-mounted eye-tracking allowed us to precisely identify coordinated attention moments based on participants' gaze information and zoom into the leading moments to examine the cues infants and parents used to follow the other person's attention. This type of analyses requires data with a high temporal and spatial resolution, which is not obtainable from traditional observational methodologies. Instead of presenting an actor's action on a 2-D screen, head-mounted eye-tracking method also allowed us to see what cues infants' use in naturalistic interaction with a real social partner. This leads us one step closer to understanding infants' use of different types of social information in real-life interactions.

2.4 Temporal flow of hand actions prior to and during coordinated attention moments

Social interactions are continuous and multimodal. Where we look is guided by what we did in the previous moment and will guide what we do next (Land and Hayhoe, 2001). In this section, we analyze the relationship between where participants look and their ongoing hand actions. In the previous section, we looked at what happened right before children followed parents' gaze (and vice versa) by investigating the events occurring during the leading moments. However, gaze and behaviors emerge and flow from the preceding moment to the next. Coordinated attention does not occur "out of the blue," it emerges from a series of back-and-forth looking and manual behaviors. In this section, we expand our analyses to what happened *prior* to the leading moment. We will focus on parent-led coordinated attention episodes and use Sankey diagrams to demonstrate the change of participants' hand actions across 3 different time windows—1 s prior to the leading moment (which is subsequently termed the *pre-leading moment*), during the *leading moment*, and during *coordinated attention moment*. We examine, during these three time windows, whether the parent and/or child touched the target object of the coordinated attention episode. We will then compare the patterns in the HL and NH groups and examine any group differences.

As an example, in Fig. 5 the target object of a coordinated attention episode is the blue object. We check whether the parent and the child touch the blue object in the pre-leading moment, during the leading moment, and during the coordinated attention episode. In each time window, there are four possibilities: *both* parent and child touch the object (B), only the *child* touches the object (C), only the *parent*

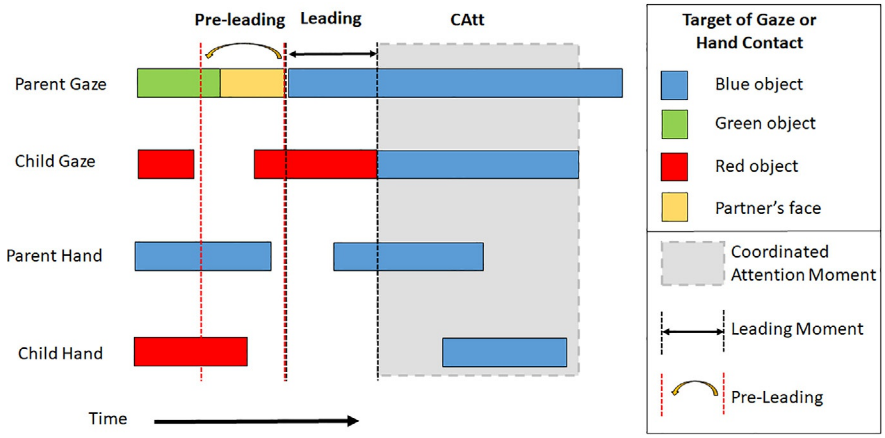

FIG. 5

Hand contact in three time windows. We examine whether the parent and/or the child touches the target object of the coordinated attention episode during the pre-leading moment (i.e., 1s prior to the leading moment), during the leading moment, and during the coordinated attention moment.

touches the object (P), and *neither* the parent nor the child touches the object (N). In this toy example, during the pre-leading moment, only the parent touches the blue object. During the leading moment, again, only the parent touches the object. During the coordinated attention moment, the child joins in and both participants touch the target object. Therefore, for this coordinated attention episode in these three time windows, the hand contact pattern goes from parent only (P), parent only (P), and then both (B), and is therefore noted as a PPB pathway. We then plotted the hand contact patterns in these three time windows for both HL group's and NH group's parent-led coordinated attention episodes and examined whether there were any group differences (Fig. 6).

2.4.1 Results and discussion

In Fig. 6, the nodes represent who is touching the target object (i.e., both parent and child, child only, parent only, neither) in each time window, and the "rivers" represent the hand contact patterns from one time window to another. The widths of the nodes and the rivers represent the proportions of the flow quantities. In each time state, the proportions of all the nodes add up to 1. Similarly, in-between two states, the proportions of all the rivers add up to 1. In the following, we will first compare the hand contact patterns *within each time window* separately for the parent-led coordinated attention episodes in the HL and NH groups. Following that, we will compare the hand contact patterns *across the three time windows* in the two groups. The first set of analyses present the overall hand contact patterns in discrete time windows.

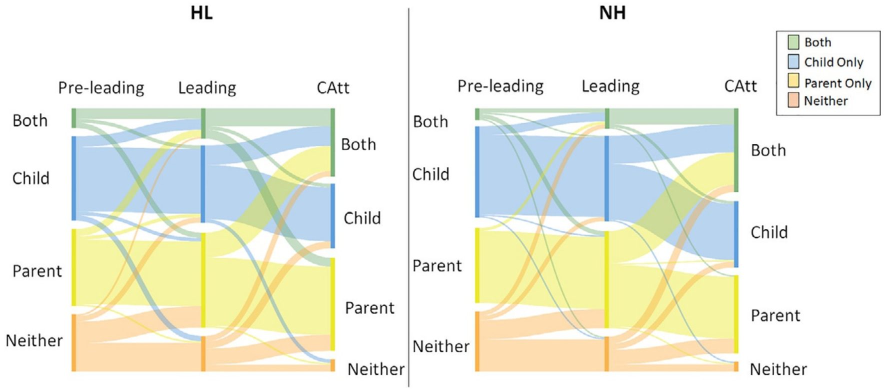

FIG. 6

Pathways leading to parent-led coordinated attention episodes. We examine who (i.e., both parent and child, child only, parent only, or neither) is touching the target object of the coordinated attention episodes in the three time windows: pre-leading moment, leading moment, and coordinated attention moment.

The second set of analyses focus on the dynamics of hand contact change across the time windows in a continuous way.

To test the overall hand contact patterns, we first calculated the proportions of both participants touching the object, child only, parent only, or neither in the three time windows separately and compared the distributions between the two groups (Fig. 6). There was no significant group difference in the hand contact distribution during any moment ($ps > 0.09$), suggesting that the groups had similar hand contact patterns *within each time window*.

As can be seen in Fig. 6, there were 64 (4 [both, child, parent, neither in pre-leading moment] * 4 [both, child, parent, neither in leading moment] * 4 [both, child, parent, neither during coordinated attention moment]) total possible hand contact patterns across the three time windows (e.g., CCC, indicating Child-Child-Child, or PPB, indicating Parent-Parent-Both). On average, dyads in the HL group demonstrated 10.86 hand contact pathways across the three time windows, while dyads in the NH group demonstrated 10.71 hand contact pathways. The top two pathways in both groups were CCC and PPP. These two pathways together accounted for 39.1% of all parent-led episodes in the HL group and 41.2% in the NH group. In fact, the top six pathways in the two groups were identical, albeit with slightly different orders (HL group: CCC, PPP, PPB, CCB, NNP, NPP; NH group: CCC, PPP, CCB, PPB, NNP, NPP). This set of results suggest that the two groups had similar hand contact patterns *across time*.

We next asked what changed across the three time windows and found two interesting patterns: one pattern related to the *neither* pathways and another pattern related to the *both* pathways. These two patterns can be seen in both HL and NH groups. First, during the pre-leading moments, *neither* the parent nor the child was touching the target object for approximately 25% of episodes. However, this percentage dropped greatly during the leading moments and further dropped to less than 5% during the coordinated attention moments. In other words, for over 95% of coordinated attention moments, at least one of the participants was touching the object. Second, in contrast, it was rare that *both* participants were touching the target object during pre-leading and leading moments. However, during coordinated attention, both participants were touching the target object for 30% of all parent-led episodes. These two patterns suggest that manual actions are associated with the establishment and maintenance of coordinated attention. One possibility is that manual actions attract attention in both dyad members, and continuous manual actions help with the maintenance of attention. The other possibility is that manual actions reflect interest, in that the participants touch, handle, or manipulate the object they are interested in. And, at the same time, they look at the object they are interested in. These two possibilities are not mutually exclusive. It is likely that manual actions contribute to and are associated with coordinated attention in different ways.

In this section of analyses, we used Sankey diagrams to represent how parents and children's hand contact with objects change *over time*. To use this method, one needs to identify different states across time and then map the factors of interest in each state or time point and plot how they change over time. This method is best used to describe how things change or do not change over time, either in the short term or long term. The analyses we presented in this section demonstrate how Sankey diagrams can be used with high-density data to track the dynamics of participants' manual actions at a fine timescale—with different temporal states being identified based on participants' gaze information. This method can also be used to track developmental changes. For example, one potential use is to conduct a longitudinal study and check whether children rely on different cues (gaze vs. hand) to coordinate attention with parents at different ages (e.g., from 12 to 36 months of age). One can identify the major pathway (e.g., gaze-following vs. hand-following) each child uses at different time points and then examine how the pathways they use change over time.

3 Implications

In this chapter, we reviewed recent developmental studies using head-mounted eye-trackers. We also presented two sets of analyses to demonstrate how to analyze gaze data along with action data to address different research questions. Prior studies suggest that parents' and children's hand contact with objects attract children's attention and support their sustained attention on objects (Suarez-Rivera et al., 2019; Yu and Smith, 2017; Yuan et al., 2019). The results reported in this chapter show that parents' and children's hand contact with objects are associated with coordinated attention in toy play. We also found that children with and without hearing loss differ in the pathways used to achieve coordinated attention, but not in the frequency or amount of attention behaviors. These findings also underline the importance of studying young infants' attention in naturalistic settings. Unlike screen-based settings, natural interaction settings are more complex. However, we are able to see how young children's attention relates to their own and their social partner's manual actions and bodily movements. Importantly, head-mounted eye-tracking records data from children's first-person perspective with a high temporal and spatial resolution, which was unobtainable using traditional observational methodologies. This new method gives us high-density data to quantify the temporal dynamics of different gaze or behavioral streams and provide a more complete picture of children's perception, action, and social interactions.

Head-mounted eye-tracking allows us to examine looking behaviors while children are actively participating in social interactions or navigating the environment. We are able to investigate how their attentional patterns interact with other behaviors and lead to successful within-individual eye-hand coordination and between-individual social coordination (Yu and Smith, 2013). We are thus one step closer to understanding how they process and learn while actively engaging in activities in the real world. It is clear that infants and toddlers do not just passively perceive the world. Instead, they actively switch gaze to things they are interested in, seek information from social partners, and gather information for action and movement planning (Chen et al., 2020; Franchak et al., 2011; Hoch et al., 2019; Kretch and Adolph, 2017; Rachwani et al., 2019; Yu and Smith, 2013, 2017). Their attention and behaviors are dynamic. They constantly switch their gaze, adjust their movements, and change their actions based on incoming information, on different action outcomes, and on how their social partners respond to their behaviors. These types of studies may seem "messy," in that they do not follow a pre-determined script or a well-controlled design. However, they are one step closer to capturing infants' and toddlers' behaviors "in the wild." With new technology and analysis methods, we can quantify detailed sensorimotor pathways and study complex behaviors in children's natural interactions with their environment and with social partners, opening doors for innovative ways of studying development and learning.

References

Bambach, S., Smith, L.B., Crandall, D.J., Yu, C., 2016. Objects in the center: How the infant's body constrains infant scenes. In: 2016 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics, pp. 132–137.

Bambach, S., Crandall, D.J., Smith, L.B., Yu, C., 2018. Toddler-inspired visual object learning. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (Eds.), Paper Presented at Neural Information Processing Systems (NIPS), pp. 1209–1218.

Bergeson, T.R., Pisoni, D.B., Davis, R.A., 2005. Development of audiovisual comprehension skills in prelingually deaf children with cochlear implants. Ear Hear. 26, 149–164.

Bornstein, M.H., Tamis-LeMonda, C.S., Tal, J., Ludemann, P., Toda, S., Rahn, C.W., Pecheux, M.-G., Azuma, H., Vardi, D., 1992. Maternal responsiveness to infants in three societies: the United States, France, and Japan. Child Dev. 63, 808–821.

Brooks, R., Meltzoff, A.N., 2005. The development of gaze-following and its relation to language. Dev. Sci. 8, 535–543.
Butterworth, G., Cochran, E., 1980. Towards a mechanism of joint visual attention in human infancy. Int. J. Behav. Dev. 3, 253–272.
Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., Moore, C., 1998. Social cognition, joint attention, and communicative competence from 9 to 15 months of age. Monogr. Soc. Res. Child Dev. 63, 1–174.
Chang, L., de Barbaro, K., Deak, G.O., 2016. Contingencies between infants' gaze, vocal, and manual actions and mothers' object-naming: longitudinal changes from 4 to 9 months. Dev. Neuropsychol. 41, 342–361.
Chen, C., Castellanos, I., Yu, C., Houston, D.M., 2019a. Parental linguistic input and its relation to toddlers' visual attention in joint object play: a comparison between children with normal hearing and children with hearing loss. Inf. Dent. 24, 589–612.
Chen, C., Castellanos, I., Yu, C., Houston, D.M., 2019b. Effects of children's hearing loss on the synchrony between parents' object naming and children's attention. Infant Behav. Dev. 57, 101322.
Chen, C., Castellanos, I., Yu, C., Houston, D.M., 2020. What leads to coordinated attention in parent-toddler interactions? Children's hearing status matters. Dev. Sci. 23, e12919.
Corbetta, D., Guan, Y., Williams, J.L., 2012. Infant eye-tracking in the context of goal-directed actions. Inf. Dent. 17, 102–125.
de Barbaro, K., Johnson, C.M., Forster, D., Deak, G.O., 2016. Sensorimotor decoupling contributes to triadic attention: a longitudinal investigation of mother-infant-object interactions. Child Dev. 87, 494–512.
Deak, G.O., Krasno, A.M., Triesch, J., Lewis, J., Sepeta, L., 2014. Watch the hands: infants can learn to follow gaze by seeing adults manipulate objects. Dev. Sci. 17, 270–281.
Deak, G.O., Krasno, A.M., Jasso, H., Triesch, J., 2018. What leads to shared attention? Maternal cues and infant responses during object play. Inf. Dent. 23, 4–28.
Flom, R., Pick, A.D., 2005. Experimenter affective expression and gaze-following in 7-month-olds. Infancy 7, 207–218.
Franchak, J.M., Kretch, K.S., Soska, K.C., Adolph, K.E., 2011. Head-mounted eye tracking: a new method to describe infant looking. Child Dev. 82, 1738–1750.
Franchak, J.M., Kretch, K.S., Adolph, K.E., 2018. See and be seen: infant-caregiver social looking during locomotor free play. Dev. Sci. 21, e12626.
Gredeback, G., von Hofsten, C., 2004. Infants' evolving representations of object motion during occlusion: a longitudinal study of 6-to 12-month-old infants. Inf. Dent. 6, 165–184.
Gredeback, G., Theuring, C., Hauf, P., Kenward, B., 2008. The microstructure of infants' gaze as they view adult shifts in overt attention. Inf. Dent. 13, 533–543.
Gredeback, G., Fikke, L., Melinder, A., 2010. The development of joint visual attention: a longitudinal study of gaze-following during interactions with mothers and strangers. Dev. Sci. 13, 839–848.
Hoch, J.E., Rachwani, J., Adolph, K.E., 2019. Where infants go: real-time dynamics of locomotor exploration in crawling and walking infants. Child Dev. 91, 1001–1020.
Johnson, S.P., Amso, D., Slemmer, J.A., 2003. Development of object concepts in infancy: evidence for early learning in an eye-tracking paradigm. Proc. Natl. Acad. Sci. U. S. A. 100, 10568–10573.
Kretch, K.S., Adolph, K.E., 2015. Active vision in passive locomotion: real-world free viewing in infants and adults. Dev. Sci. 18, 736–750.

Kretch, K.S., Adolph, K.E., 2017. The organization of exploratory behaviors in infant locomotor planning. Dev. Sci. 20, e12421.
Kretch, K.S., Franchak, J.M., Adolph, K.E., 2014. Crawling and walking infants see the world differently. Child Dev. 85, 1503–1518.
Land, M.F., Hayhoe, M., 2001. In what ways do eye movements contribute to everyday activities? Vision Res. 41, 3559–3565.
Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.
Masur, E.F., Gleason, J.B., 1980. Parent-child interaction and the acquisition of lexical information during play. Dev. Psychol. 16, 404–409.
McMurray, B., Aslin, R.N., 2004. Anticipatory eye movements reveal infants' auditory and visual categories. Infancy 6, 203–229.
McQuillan, M.E., Smith, L.B., Yu, C., Bates, J.E., 2020. Parents influence the visual learning environment through children's manual actions. Child Dev. 91, e701–e720.
Monroy, C., Chen, C., Houston, D., Yu, C., 2019. Action prediction during real-time social interactions in infancy. In: Goel, A.K., Seifert, C.M., Freksa, C. (Eds.), Proceedings of the 41st Annual Conference of the Cognitive Science Society. Cognitive Science Society, Montreal, QB, pp. 836–841.
Mundy, P., Newell, L., 2007. Attention, joint attention, and social cognition. Curr. Dir. Psychol. Sci. 16, 269–274.
Navab, A., Gillespie-Lynch, K., Johnson, S.P., Sigman, M., Hutman, T., 2012. Eye-tracking as a measure of responsiveness to joint attention in infants at risk for autism. Inf. Dent. 17, 416–431.
Rachwani, J., Herzberg, O., Golenia, L., Adolph, K.E., 2019. Postural, visual, and manual coordination in the development of prehension. Child Dev. 90, 1559–1568.
Richmond, J., Nelson, C.A., 2009. Relational memory during infancy: evidence from eye tracking. Dev. Sci. 12, 549–556.
Schmidt, M., 2008. The Sankey diagram in energy and material flow management: part I: history. J. Ind. Ecol. 12, 82–94.
Sebanz, N., Knoblich, G., 2009. Prediction in joint action: what, when, and where. Top. Cogn. Sci. 1, 353–367.
Senju, A., Csibra, G., 2008. Gaze-following in human infants depends on communicative signals. Curr. Biol. 18, 668–671.
Slaughter, V., McConnell, D., 2003. Emergence of joint attention: relationships between gaze-following, social referencing, imitation, and naming in infancy. J. Genet. Psychol. 164, 54–71.
Slone, L.K., Abney, D.H., Borjon, J.I., Chen, C., Franchak, J.M., Pearcy, D., Suarez-Rivera, C., Xu, T.L., Zhang, Y., Smith, L.B., Yu, C., 2018. Gaze in action: head-mounted eye tracking of children's dynamic visual attention during naturalistic behavior. J. Vis. Exp. 141, e58496.
Slone, L.K., Smith, L.B., Yu, C., 2019. Self-generated variability in object images predicts vocabulary growth. Dev. Sci. 22, e12816.
Suanda, S.H., Barnhart, M., Smith, L.B., Yu, C., 2019. The signal in the noise: the visual ecology of parents' object naming. Infancy 24, 455–476.
Suarez-Rivera, C., Smith, L.B., Yu, C., 2019. Multimodal parent behaviors within joint attention support sustained attention in infants. Dev. Psychol. 55, 96–109.
Summerfield, Q., 1992. Lipreading and audio-visual speech perception. Philos. Trans. R. Soc. Lond. B Biol. Sci. 335, 71–78.

Tamis-LeMonda, C.S., Bornstein, M.H., 1989. Habituation and maternal encouragement of attention in infancy as predictors of toddler language, play, and representational competence. Child Dev. 60, 738–751.

Tasker, S.L., Nowakowski, M.E., Schmidt, L.A., 2010. Joint attention and social competence in deaf children with cochlear implants. J. Dev. Phys. Disabil. 22, 509–532.

Tomasello, M., Farrar, M.J., 1986. Joint attention and early language. Child Dev. 57, 1454–1463.

Wiener, R.F., Thurman, S.L., Corbetta, D., 2018. Directing gaze on a scene before reaching for an object: changes over the first year of life. J. Motor Learn. Dev. 6, S105–S125.

Yoshida, H., Smith, L.B., 2008. What's in view for toddlers? Using a head camera to study visual experience. Infancy 13, 229–248.

Yu, C., Smith, L.B., 2012. Embodied attention and word learning by toddlers. Cognition 125, 244–262.

Yu, C., Smith, L.B., 2013. Joint attention without gaze-following: human infants and their parents coordinate visual attention to objects through eye-hand coordination. PLoS One 8, e79659.

Yu, C., Smith, L.B., 2016. The social origins of sustained attention in one-year-old human infants. Curr. Biol. 26, 1235–1240.

Yu, C., Smith, L.B., 2017. Multiple sensory-motor pathways lead to coordinated visual attention. Cognit. Sci. 41, 5–31.

Yu, C., Suanda, S.H., Smith, L.B., 2019. Infant sustained attention but not joint attention to objects at 9 months predicts vocabulary at 12 and 15 months. Dev. Sci. 22, e12735.

Yuan, L., Xu, T.L., Yu, C., Smith, L.B., 2019. Sustained visual attention is more than seeing. J. Exp. Child Psychol. 179, 324–336.

Yurkovic, J.R., Lisandrelli, G., Shaffer, R.C., Dominick, K.C., Pedapati, E.V., Erickson, C.A., Yu, C., Kennedy, D.P., 2020. Examining sustained attention in child-parent interaction: A comparative study of typically developing children and children with autism spectrum disorder. In: Proceedings of the 42nd Annual Meeting of the Cognitive Science Society.