Video Article

A View of Their Own: Capturing the Egocentric View of Infants and Toddlers with Head-Mounted Cameras

Jeremy I. Borjon*[1], Sara E. Schroer*[1], Sven Bambach[2], Lauren K. Slone[1], Drew H. Abney[1], David J. Crandall[2], Linda B. Smith[1]

[1]Department of Psychological and Brain Sciences, Indiana University

[2]School of Informatics, Computing, and Engineering, Indiana University

*These authors contributed equally

Correspondence to: Jeremy I. Borjon at jborjon@iu.edu, Sara E. Schroer at seschroe@iu.edu

Abstract

Infants and toddlers view the world, at a basic sensory level, in a fundamentally different way from their parents. This is largely due to biological constraints: infants possess different body proportions than their parents and the ability to control their own head movements is less developed. Such constraints limit the visual input available. This protocol aims to provide guiding principles for researchers using head-mounted cameras to understand the changing visual input experienced by the developing infant. Successful use of this protocol will allow researchers to design and execute studies of the developing child's visual environment set in the home or laboratory. From this method, researchers can compile an aggregate view of all the possible items in a child's field of view. This method does not directly measure exactly what the child is looking at. By combining this approach with machine learning, computer vision algorithms, and hand-coding, researchers can produce a high-density dataset to illustrate the changing visual ecology of the developing infant.

Video Link

The video component of this article can be found at https://www.jove.com/video/58445/

Introduction

For decades, psychologists have sought to understand the environment of the developing infant, which William James famously described as a "blooming, buzzing confusion[1]." The everyday experiences of the infant are typically studied by filming naturalistic play with social partners from a third-person perspective. These views from the side or above typically show cluttered environments and a daunting number of potential referents for any new word an infant hears[2]. To an outside observer, James's description is apt, but this stationary, third-person perspective is not the way an infant sees the world. An infant is closer to the ground and can move through their world, bringing objects closer for visual exploration. A third-person view of a parent-infant interaction is illustrated in **Figure 1**. Highlighted are the fundamental differences between their perspectives. Perhaps, the input that infants receive is not nearly as chaotic as anticipated by parents and researchers. The goal of methods with head-mounted cameras is to capture the infant experience from a first-person view in order to understand the visual environment available to them throughout development.

Head-mounted cameras, worn on a hat or headband, provide a window into the moment-to-moment visual experiences of the developing infant. From this perspective, the study of the structure and regularities in the infant's environment becomes apparent. Head-mounted cameras have revealed infants' visual experiences to be largely dominated by hands, both their own and their social partner's, and that face-looks, once considered imperative for establishing joint attention, are much scarcer than anticipated[3]. Head-mounted cameras have also shown that infants and their caregivers create moments when objects are visually dominant and centered in the infant's field of view (FOV), reducing the uncertainty inherent to object-label mapping[4].

Head-mounted cameras capture the infants' first-person view based on head movements. This view is not perfectly synchronous with, or representative of, infant eye movements, which can only be captured in conjunction with an eye-tracker. For instance, a shift of only the eyes while keeping the head stationary, or a shift of the head while keeping the eyes fixed on an object, will create a misalignment between the infants' actual FOV and the one captured by the head camera. Nonetheless, during toy play, infants typically center the objects they are attending to, aligning their head, eyes, and the location of the toy with their body's midline[5]. Misalignments are rare and are typically created by momentary delays between an eye shift and the accompanying head turn[3]. Therefore, head-cameras are not well suited to capturing the rapid dynamics of shifts in attention. The strength of head-mounted cameras lies in capturing the everyday visual environment, revealing the visual content available to infants.

The following protocol and representative results will demonstrate how head-mounted cameras can be used to study the visual environment of infants and toddlers.

Protocol

The following procedure to collect data on infant and toddler's visual experiences in the laboratory and at home was approved by the Indiana University Institutional Review Board. Informed consent was obtained from the infant's caregiver.

1. Choose a Head Camera

NOTE: There are numerous small, lightweight, and portable cameras readily available for purchase (**Figure 2**).

1. Choose a head camera that is unobtrusive and will not influence the scenes being recorded.
    1. Mount the camera onto a snug hat or headband using a temporary adhesive or securely mount the camera onto a small plastic plate attached to the headband. Position the camera at the child's brow (**Figure 2B**, left).
        1. Depending on the shape of the camera, mount the camera by sewing small cloth loops into a headband or hat (**Figure 2B**, center and right).
        2. Ensure that the hat or headband are adjustable to achieve a snug and comfortable fit on the child's head (**Figure 2A**).
        NOTE: A camera located directly between the eyes of the infant is theoretically ideal, however this is not quite possible with current technology. Small cameras that can be placed low on the forehead are becoming increasingly available.

2. If cameras are being sent home with the parents, ensure that parents can use them without any technical help.
    1. Before the parents leave the laboratory, train them on how to operate the head camera.
    2. Send the parents home with the camera to use, the head-band or hat, and a hand-out describing the procedure to turn on the camera and record data.
3. Ensure that the cameras are light enough for infants to wear and forget about.
   NOTE: The recommended weight is less than 30 g[3]. Any chosen camera should not heat up during use and must be durable under heavy use.
4. If the experimental setup requires the children to be freely-moving, store recorded video to an internal storage card. Otherwise, use cameras tethered to the data collection computer.
5. Use a camera with a high-quality lens to better leverage recent advances in machine learning algorithms which segment visual scenes into regions of interest.
    1. If visual scenes will be manually coded by researchers, use a camera that can capture images at sufficient quality for manual inspection.
6. Ensure that any battery-powered camera is capable of recording for the desired amount of time.
7. Shortly before beginning the experiment or sending the camera home with the parents, test the head camera to ensure it is working appropriately.

2. Data Collection in the Laboratory

NOTE: Head-mounted cameras can be easily added to most experiments.

1. Have 2-3 experimenters place the camera onto the child's head: one experimenter places the head camera, one monitors the head-camera view, and, if needed, one distracts the child.
    1. Ask the parent to ensure that the child remains calm and to distract the child during the placement process.
2. Perform the camera placement in three steps as follows.
    1. Desensitize the infant to hand actions near their head.
        1. Ask the parent to lightly touch or stroke the infant's head and hair several times.
        2. Ask the experimenter placing the hat on the infant's head to do the same as in 2.2.1.1.

    2. Have the experimenter place the head-mounted camera when the child is distracted.
        1. Use push-button toys to keep the child's hands busy.
        2. Have the distracting experimenter or parent help at this stage by gently pushing the child's hands toward the engaging toy so that the child's hands do not go to the head.

    3. Tighten the hat on the child's head and adjust the head camera when the child is engaged with the toy.
        1. Adjust the camera so that when the infant holds an object in front of his/her face, the object is centered in the head camera FOV.
        2. If the child is sitting, adjust the camera so that it captures most of the child's lap when the child looks down.

3. After placing the camera on the child's head, ask the experimenter to leave the room and begin the recording.
4. In the event that a camera is moved out of place or removed, re-enter the room to correct the camera.
    1. Terminate the experiment if the child does not tolerate the camera being reapplied.
    NOTE: For recording natural environments in the home, first fit a hat and camera to the individual infant and show parents how to position the camera. The design and fit of the camera must ensure that parents will be able to put the hat on their child without technical help.

3. Data Collection for the Parent-Infant Study

NOTE: The following representative method for head-cameras uses naturalistic toy play in the lab to demonstrate the type of analyses that can be conducted on the egocentric views of infants and their parents (**Figure 3A**).

1. Outfit the parent and child with head-mounted cameras, as described in 2.1 and 2.2.
2. Use head cameras to capture videos with a resolution of 720 x 1280 pixels at 30 frame/s. Proceed as described in steps 2.3 and 2.4.
   1. Subsample the video stream at one frame every 5 s.
   2. Manually draw bounding boxes around each toy (**Figure 3B**) within view using commercial software or a program developed in-house (see **Figure 3C** for a sample frame).
      1. If only part of a toy is visible due to occlusion by other toys or truncation on the edge of the frame, only draw bounding boxes when a toy is easily identifiable and include all visible parts of the toy.
      2. For example, if only the doll's leg is visible, draw a bounding box around its leg. If an object occludes half of the doll, leaving the hair and legs visible, then draw a box that includes both hair and legs.

Representative Results

One simple, yet informative, analysis is to count the number of objects in view at each point in time. Since a head camera produces data at approximately 30 Hz (30 images/s), down-sampling the data to 1 image every 5 s helps to produce a more manageable dataset while maintaining a resolution appropriate for understanding the types of scenes children see. Prior research has demonstrated that visual scenes are slow-changing in infants[3]. A custom script was used to draw bounding boxes around the toys in view. **Figure 4** shows representative results for 1 parent-infant dyad. An independent t-test comparing the number of scenes with a given number of objects between the parent (**Figure 4A**) and the child (**Figure 4B**) revealed this child had a greater number of scenes with fewer objects in view compared to the parent ($t(78) = 4.58$, $p < 0.001$).

Another informative analysis is to calculate how visually large the objects are in each view. The proportion of the screen taken up by each object in view can be calculated and analyzed. For both parent and child, there is a negative correlation between the number of objects in view and the visual sizes of the objects in that view (**Figure 4C**, Spearman correlation $r = -0.19$, $p < 0.001$ and **Figure 4D**, Spearman correlation $r = -0.23$, $p < 0.001$). That is, if there are more objects in view, each object takes up less of the screen than if there are fewer objects in view. For this dyad, the child captured more scenes with less than 10 objects in view and the parent exhibits a larger number of objects in view. Similar results have been previously reported in the literature[3,4,5,8,9,10,11,12,13].
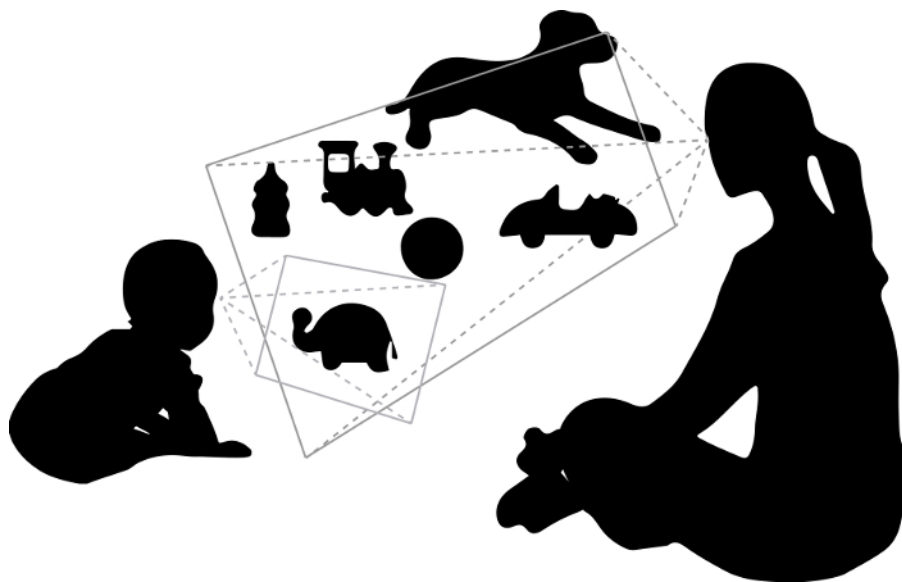

Figure 1: An illustrative schematic demonstrating the different views of a parent and their child during play. Please click here to view a larger version of this figure.

Figure 2: Examples of head-mounted cameras and their attachments. (**A**) Infants and toddlers wearing head-mounted cameras at home and in the lab. (**B**) Examples of ways to attach head cameras to headbands (left, middle) and hats (right). Please click here to view a larger version of this figure.

Figure 3: The 24 toys used in the representative method. (**A**, left) Representative frame from the head camera of a child illustrating a smaller number of objects in view and at visually larger sizes. (**A**, right) A representative frame from the head camera of a parent illustrating their typical view: many objects at visually smaller sizes. (**B**) Toys are consistent in size, ranging from 2-7 inches on the longest dimension, and between 2-3 inches on the shorter dimensions. (**C**) Boxes are drawn around each toy, or part of toy that is visible and identifiable, using an in-house graphical user interface. Please click here to view a larger version of this figure.
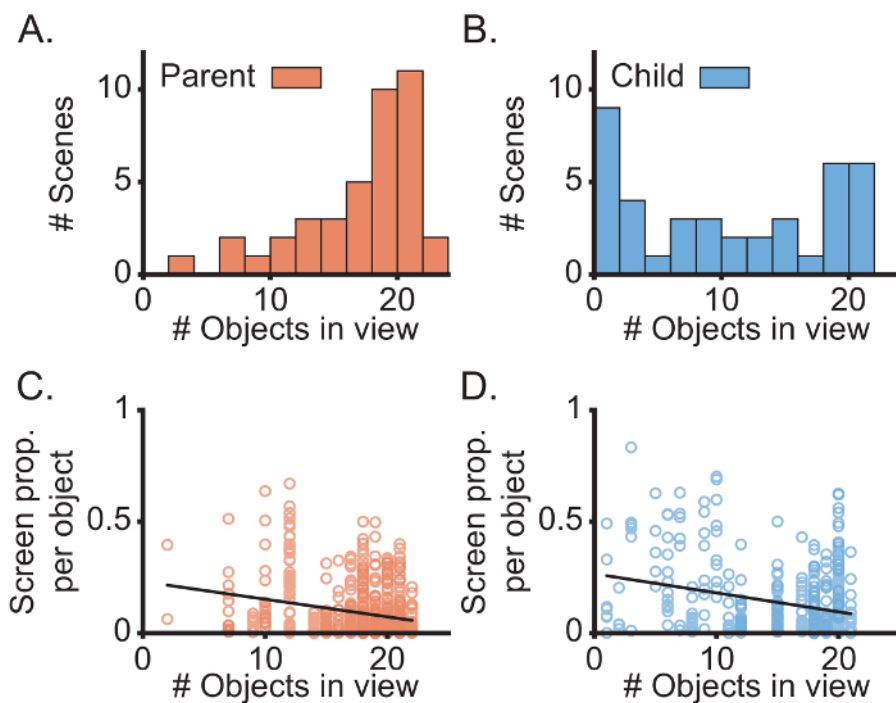
Figure 4: Representative results from a single dyad participating in toy play. Histograms grouping the number of scenes based on the number of objects in view for the parent (**A**) and child (**B**). The proportion of the screen taken up by each object in view *versus* the number of objects in view for the parent (**C**) and the child (**D**). The black line is the line of best fit. Please click here to view a larger version of this figure.

Discussion

This paper outlines the basics for applying head-mounted cameras to infants to capture their egocentric visual scene. Commercially available head cameras are sufficient for the vast majority of studies. Small, lightweight, and portable cameras should be incorporated into a soft fabric hat or headband and applied to the child's head. Once successfully designed and implemented, a variety of experiments can be run, both in laboratory settings as well as in the home environment. From the videos gathered, aggregate data about the developing infant's visual ecology can be compiled and analyzed.

The most critical step with this method is the application of the head camera onto the child. If done incorrectly, the head camera will be poorly placed and data quality will be diminished or unusable. An incorrect placement could also spur the child to reject the camera and halt the experiment. We will briefly discuss suggestions to ensure success with the application of the head camera. Cameras should be placed on the infant in one move without hesitation. If the researcher is apprehensive to place a camera on the child's head, or if multiple attempts are made, the likelihood of refusal becomes much higher. Experimenters should practice placing hats and camera devices on willing toddlers or mannequins beforehand. When placing the camera, it must be placed low enough on the forehead to ensure a clear view of the scene in front of the face. Slightly angling the camera downwards will guarantee view of the infant's hands during active manipulation. The camera should also be stable and secure on the infant's head. Stable cameras mean stable and clear images. If the headwear jiggles, toddlers can notice this and pull the camera off. For children under 18 months of age, anything drawing attention to the gear increases refusals. This includes having the infant handle the equipment or talking about it before placing it on the child. For children over 18 months of age, talking about the camera beforehand and asking the child's permission to put it on may be more effective. With a trained researcher, success rates in placing head cameras on infants, without the infant fussing out of the experiment, can reach around 75%.

When sending a head-camera home with the families, take considerable time to design the cap/headband and camera placement. The way a parent places a camera on their child's head will not always be at the same level of precision as a trained researcher. Ensure the cap is easy to apply by the parents and ensure the research questions do not demand exacting specifications. If experimental needs require precise placement of the camera on the head, consider running the study in a laboratory setting instead of at home.

Head-cameras will have limitations in what they can capture. Given the location of the camera on the head of the infant, the horizontal views of the infant as they move their head from left to right will be extensively captured. Vertical displacements of the camera, when the infant looks up and down, will be unable to capture the very extremes of the visual scene. This is especially true if the camera is angled slightly downwards on the head of the infant, in order to capture the infant's hands.

Head-mounted cameras have revealed that children have a view of their own. On a fundamental level, toddlers and infants view the world differently than their parents. Toddlers shape their visual experience with their hands: holding and manipulating objects close to their face[4,5,8]. Given the very short arms of toddlers, the object is held close and appears large in the field of view. Such scenes with a clear focal object are often long-lasting, around four seconds in duration, and coincide with the lessening of head movements by the infant[4]. It is important to note, however, that head-mounted cameras do not provide any information as to where the participant is looking. Instead, this protocol can quantitatively describe the range of visual scenes available to children. Across an entire play session, there is a high probability that the child's eyes are typically centered in the middle of the visual scene in front of the child[5]. Head cameras allow us to investigate the aggregate of the

scenes available to a child. For example, how often is a face available for them to look at? How persistent are these scenes with faces? How often are children viewing cluttered (a pile of toys on the floor) *versus* uncluttered (the ceiling or a blank wall) scenes? This egocentric view methodology is best suited for data at the macro-scale, 100 million images collected over days. If the research question needs more fine-grained resolution than these aggregate-level questions, head-mounted eye-tracking may be better suited to capture the exact dynamics of infant vision.

Just as the toddler and adult have different visual experiences, the visual experience of infants and toddlers is not developmentally static. As children grow, the available visual scenes change dramatically, and there is developmental structure in the people and objects visually available to infants of different ages. For instance, when infants are very young, their visual environment is dense with the faces of a very small number of people[10]. From this non-uniform sampling of a few faces, infants can extrapolate and learn to recognize and discriminate between faces they encounter. At around 8-10 months of age, infants are beginning to sit steadily, to crawl, and to play with objects, but their manual skills are still quite limited when compared to older infants. As a result, these infants experience a higher frequency of visual scenes with few objects in view, compared to older infants. Nevertheless, mealtime scenes from these same 8- to 10-month-olds also reveal times of clutter[13], with each mealtime scene containing many different objects. Despite this clutter, there is a predictable structure to the objects in view: a very small set of objects appear repeatedly. These repeated objects belong to categories encompassing the first words learned by infants[13]. Thus, although it may be easy to look upon children's environment and argue their world is a "blooming, buzzing confusion," head-camera data showing infants' egocentric views reveal that predictable statistical regularities exist in their field of view to dampen the din and confusion.

Disclosures

The authors declare no conflicts of interest.

Acknowledgements

References

1. James, W. *The Principles of Psychology.* Henry Holt and Co. New York. (1890).
2. Quine, W.V.O. *Word and object: An inquiry into the linguistic mechanisms of objective reference*. The MIT Press. Cambridge, MA. (1960).
3. Yoshida, H., & Smith, L. B. What's in view for toddlers? Using a head camera to study visual experience. *Infancy.* **13** (3), 229-248 (2008).
4. Yu, C., & Smith, L. B. Embodied attention and word learning by toddlers. *Cognition.* **125** (2), 244-262 (2012).
5. Bambach, S., Smith, L. B., Crandall, D. J., & Yu, C. Objects in the center: How the infant's body constrains infant scenes. In *Joint IEEE International Conference on Development and Learning and Epigenetic Robotics 2016.* 132-137 (2016).
6. Adolph, K. E., Gilmore, R. O., Freeman, C., Sanderson, P., & Millman, D. Toward open behavioral science. *Psychological Inquiry.* **23** (3), 244-247 (2012).
7. Sanderson, P. M., Scott, J. J. P., Johnston, T., Mainzer, J., Wantanbe, L. M., & James, J. M. MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA). *International Journal of Human-Computer Studies.* **41** (5), 633-681 (1994).
8. Pereira, A. F., Smith, L. B., & Yu, C. A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review.* **21** (1), 178-185 (2014).
9. Yu, C., & Smith, L.B. Joint Attention without Gaze Following: Human Infants and Their Parents Coordinate Visual Attention to Objects through Eye-Hand Coordination. *PLOS ONE.* **8** (11), e79659 (2013).
10. Jayaraman, S., Fausey, C. M., & Smith, L. B. The faces in infant-perspective scenes change over the first year of life. *PlOS ONE.* **10** (5), e0123780 (2015).
11. Fausey, C. M., Jayaraman, S., & Smith, L. B. From faces to hands: Changing visual input in the first two years. *Cognition.* **152**, 101-107 (2016).
12. Jayaraman, S., Fausey, C. M., & Smith, L. B. Why are faces denser in the visual experiences of younger than older infants? *Developmental Psychology.* **53** (1), 38 (2017).
13. Clerkin, E. M., Hart, E., Rehg, J. M., Yu, C., & Smith, L. B. Real-world visual statistics and infants' first-learned object names. *Philosophical Transactions of the Royal Society B, Biological Sciences.* **372**, 20160055 (2017).