JOURNAL OF COGNITION AND DEVELOPMENT, 16(3):407–419

ISSN: 1524-8372 print/1532-7647 online
DOI: 10.1080/15248372.2014.933430

Routledge
Taylor & Francis Group

TOOLS OF THE TRADE

Contributions of Head-Mounted Cameras to Studying the Visual Environments of Infants and Young Children

Linda B. Smith and Chen Yu
Indiana University

Hanako Yoshida
University of Houston

Caitlin M. Fausey
Indiana University

Head-mounted video cameras (with and without an eye camera to track gaze direction) are being increasingly used to study infants' and young children's visual environments and provide new and often unexpected insights about the visual world from a child's point of view. The challenge in using head cameras is principally conceptual and concerns the match between what these cameras measure and the research question. Head cameras record the scene in front of faces and thus answer questions about those head-centered scenes. In this ''Tools of the Trade'' article, we consider the unique contributions provided by head-centered video, the limitations and open questions that remain for head-camera methods, and the practical issues of placing head cameras on infants and analyzing the generated video.

Precrawlers, crawlers, and walkers have different visual experiences of objects, space, and social partners (Adolph, Tamis-LeMonda, Ishak, Karasik, & Lobo, 2008; Bertenthal & Campos, 1990; Kretch, Franchak, Brothers, & Adolph, 2012; Soska & Adolph, 2014). Because the body's morphology and behavior change dramatically and systematically in early development, there are concomitant developmental changes in visual environments, changes that are likely to play an explanatory role with respect to development in many domains (see Byrge, Sporns, & Smith, 2014; Smith, 2013). However, we are at the earliest stages of understanding the specific

Correspondence should be sent to Linda B. Smith, Cognitive Development Lab, Department of Psychological and Brain Sciences, Indiana University, 1101 East Tenth Street, Bloomington, IN 47405-7007, USA. E-mail: smith4@indiana.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hjcd.

properties of children's environments and how they change with development. This article discusses how, by capturing a child-centered perspective on the visual world, head cameras may contribute to an understanding of the role of developmentally changing visual environments in the developmental process.

The central challenge in using head cameras to capture the ''child's view'' is conceptual and concerns the relevant scales at which environments may be measured. The conceptual problem derives from the fact that eyes and heads *typically* move together but do not *always* move together (see Schmitow, Sternberg, Billard, & von Hofsten, 2013). Because heads and eyes typically move together, there has been considerable interest in whether head cameras might provide useable data for studying looking behavior and visual attention; however, because heads and eyes do not always move together, there are also limitations as to what can be inferred from head-camera data alone (Aslin, 2008, 2012; Schmitow et al., 2013). In the first section of the article, we set the background by considering this larger conceptual issue. We then consider the unique role of head cameras in capturing visual scenes linked to the wearer's bodily posture and location. We then turn to open and theoretically important questions concerning heads, eyes, and their alignment that are also relevant to assessing the limits and potential contributions of head cameras. Finally, we consider the practical issues in using head cameras.

Before proceeding, it is helpful to make explicit the relation between head cameras and head-mounted eye trackers as measuring devices. Head-mounted eye trackers are just head-mounted cameras with an added camera directed at the eye to capture gaze direction. Algorithms are then used to estimate pupil orientation and corneal reflections from the eye camera and to project that information onto the *head-camera view* of the scene. There are many complexities in this step (see Aslin, 2012; Holmqvist et al., 2011; Nyström & Holmqvist, 2010; Wass, Smith, & Johnson, 2013). Further, although psychological significance of fixations has been studied in adults (e.g., Nuthmann, Smith, Engbert, & Henderson, 2010), little is known about the meaning of the non-adult-like frequencies and durations of infant and toddler fixations (see Wass et al., 2013). We do not consider these issues but instead focus on the unique contributions provided by the head-mounted camera whether used alone or as part of a head-mounted eye-tracking system. But keep in mind, with the exception of knowing the momentary direction of eye gaze, every contribution and every limitation concerning the video recorded from a head camera applies to head cameras used alone and as part of a head-mounted eye-tracking system.

THREE VIEWS ON DEVELOPMENT

Figure 1 shows the spatial scales of three perspectives on the visual environment: a third-person view, a first-person view, and fixations within the first-person view. Long before the invention of small head cameras or eye trackers, developmental researchers put video cameras on tripods and recorded third-person (observer) views of children's environments. Because much of this broad scene may be out of the view of the child at any moment, the room-size observer view may be considered a measure of the child's *potential* environment. Coded properties of these third-person views have repeatedly been shown to be predictive of developmental outcomes in many domains (e.g., Cartmill et al., 2013; Rodriguez & Tamis-LeMonda, 2011). However, cameras on tripods capture the same view regardless of the child's age and actions. For example,

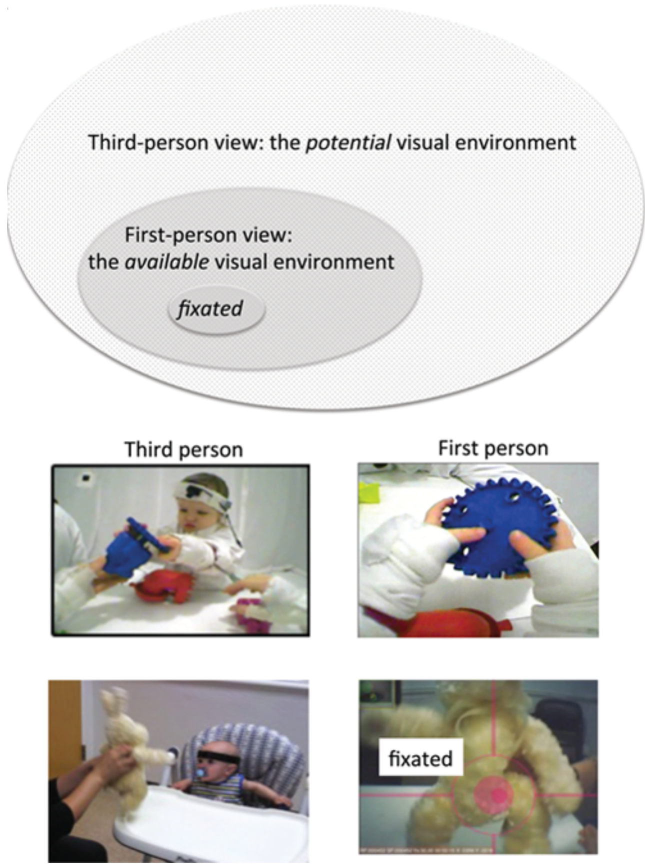

FIGURE 1 Three spatial scales for measuring the visual environment: the third-person view of the visual environment that may be potentially seen by the child; the first-person view of the available visual environment that is directly linked to the child's bodily location and posture; and the fixated elements of the first-person view.

the parent's face, the ceiling fan, small spots of dirt, and the toys on the rug may all be part of a recorded third-person view, and thus, all are within the potential environment for the studied child. However, the overhead fan is more visually available to a 3-month-old infant who is often in an infant seat on the table than it is to an 8-month-old who is often sitting or crawling on the floor. Likewise, the crawling 8-month-old has more visual access to the dirt spots on the rug than does the 3-month-old.

Head cameras replace the tripod with the child's own body and measure the *available* visual environment, or the scene that is in front of the wearer's face. This is a view that varies as a function of the child's location, posture, and activity. The evidence from head-camera studies to date indicates that the composition and statistical properties of these child views change considerably during the first 2 years of life. For example, for very young infants, the scene in front of their face is often full of other people's faces, whereas the scenes in front of older infants' faces contain many more views of hands-on objects (Frank, Vul, & Saxe, 2012; Jayaraman, Fausey, & Smith, 2013; Sugden, Mohamed-Ali, & Moulson, 2013). In one study, Kretch et al. (2012) compared the views of crawling and walking infants: Crawling infants—when crawling—had limited views of their social partners' faces and a limited view of potential goal objects. When infants were crawling, the head-camera images showed the floor and infants had to stop crawling to sit up and look at their social partners or the goal object. The head-camera images from walking infants, in contrast, showed continuous views of social partners and goal objects. In brief, the unique contribution of the head camera derives from the fact that it captures the region of the visual environment that is directly in front of the child, a moving region that changes in perspective, depth of field, and contents as the child's body, posture, and activities change moment to moment and over developmental time.

Eye trackers capture fixations within the recorded first-person view. By adding the measure of eye-gaze direction to the head-mounted camera, the researcher increases the spatial and temporal precision of the measured visual environment to determine just where in the head camera-captured scene the perceiver directed his or her gaze. Studies using head-mounted eye-tracking systems have yielded new insights into how infants and children use visual cues to reach for and grasp objects (Corbetta, Guan, & Williams, 2012), how they search for goal objects while moving in large physically complex spaces (Franchak & Adolph, 2012), how they coordinate head movements and eye gaze (Schmitow et al., 2013), and how they coordinate visual attention with a social partner (Yu & Smith, 2013).

All three of these perspectives on the visual environment—the *potential* information in the third-person view, the *availabl*e information in the first-person view, and the *fixated* information within the first-person view—are relevant to understanding the visual environments of developing children. But they provide different information that may be suited to different questions about the visual environment.

SCENES

The unique contribution of the head camera is that it measures scenes, *what wearers have the opportunity to see*. Researchers need these child-centered views in part because we—from our adult perspectives—do not have good intuitions about how the world looks to infants and toddlers and because these scenes may differ considerably from those available to adults. For example, Smith, Yu, and Pereira (2011; see also Yoshida & Smith, 2008) and Yu and Smith (2012) recorded head-camera videos from parents and toddlers as they played together with objects. The toddler view of a scene often contained a single object that was large and dominating (see Figure 1). In contrast, the parent view of a scene was broader and encompassed all the toys in play. In another study, Yurovsky, Smith, and Yu (2013) presented adults with scenes of parents naming objects for their toddlers. A beep replaced the name, and the adult's task was to guess—given the video clip—the object that was named. Adults were much better able to predict the named object from a series of child views than from a series of observer views, a result that confirms that child views contain unique information not available from other views.

We also need to measure these scenes for a well-founded account of visual development. In a recent review of gaps in developmental vision science, Braddick and Atkinson (2011) called for a description of the statistical structure of child-experienced scenes. They noted the considerable progress that has been made by studying the statistics of natural scenes (from third-person-perspective photographs of the physical world) and how properties of the mature visual system appear to be adaptations of the statistical regularities in those scenes (see Simoncelli, 2003). The developing visual system does not have access to all the kinds of scenes used to study natural statistics in adult vision. Instead, the visual scenes encountered by developing infants are more selective and are ordered in systematic ways across development. By recording the scenes in front of developing children's faces, head cameras provide a direct way to collect the developmentally appropriate scenes needed to determine their statistical properties. Although statistical analyses of the properties that characterize a large corpus of developmentally indexed head-camera scenes is just beginning (Jayaraman et al., 2013), this would seem to be a critical step toward understanding the role of visual environments in visual and cognitive development. The value of a *developmental* study of the natural statistics of scenes is supported by several recent studies using head cameras that have shown direct links between the contents of head-camera images and independent measures of performance in the domains of causality and agency (Cicchino, Aslin, & Rakison, 2011), object name learning (Pereira, Smith, & Yu, 2013; Yu & Smith, 2012; Yurovsky et al., 2013), and visual object recognition (James, Jones, Smith, & Swain, 2013). These studies provide direct evidence of the validity of head-camera images as measures of developmentally relevant properties of visual environments.

SCENES VERSUS FIXATIONS

Infants and adults typically turn their heads and eyes in the same direction to attend to a visual event (e.g., Bloch & Carchon, 1992; Daniel & Lee, 1990; Tatler, Hayhoe, Land, & Ballard, 2011; von Hofsten, Vishton, Spelke, Feng, & Rosander, 1998; Yoshida & Smith, 2008). The likelihood that both head and eyes move together may be particularly high in young children (Murray, Lillakas, Weber, Moore, & Irving, 2007; Nakagawa & Sukigara, 2013). Eyes typically lead infant heads by just fractions of a second (Schmitow et al., 2013; Yoshida & Smith, 2008). These facts foster the idea that head cameras by themselves might work as measures of attention and looking behavior (Aslin, 2008; Schmitow et al., 2013). The problem is that although both heads and eyes tend to move in the same direction, head movement undershoots eye movements at both horizontal and vertical extremes. Schmitow et al. (2013) measured eye and head movements in 6-month-olds and 12-month-olds. There were always fewer head movements than eye movements. The undershot was less than 5^{o} when the target was less than 30^{o} from the body-defined center, but it was more than 10^{o} when the target was laterally extreme (80^{o}). In light of the full pattern of their findings, Schmitow et al. concluded that head-mounted cameras are suitable for measuring horizontal looking direction in a task (such as toy play on a table) in which the main visual events deviate only moderately ($\pm 50^{o}$) from midline. In their view, in such geometrically constrained contexts, head movements are sufficiently correlated with eye movements to allow reasonable inferences. However, head movements are not sufficient in many contexts. Our view is that eye cameras, which are designed to measure gaze direction precisely, are the best method for measuring looking behavior. Head-mounted cameras capture the scenes in front of faces, and the research questions that head cameras can best answer are questions *at the level of scenes*, but not gaze within scenes.

For scene-level questions about the contents of the visual environment, the relevant methodological limit is not where the eyes are but whether the head camera captures the relevant scene information. This is a much more complicated question than it might first appear to be. We know that when viewers orient to a new target, head cameras miss those targets at the extremes. But orienting—that is, *turning* heads and eyes to a new target—is a momentary event, and those ''extreme'' targets, if attended to, do not remain in the periphery and outside of the central region of the head-camera image for long. As yet, there is no precise quantification of just how much or for how long information is missing from the head-camera view given directional shifts in eyes and heads.

We do know head cameras systematically miss available visual information because the lenses on current head cameras are not as broad the visual field. Visual fields are classically measured in terms of shifts in eye gaze to stimulus onsets in the periphery from a fixation at the center. The evidence suggests that infants detect onsets in the periphery up to 90° from the center, and by 16 months of age, they detect onsets in the periphery up to 170° horizontally and vertically (Cummings, Van Hof-Van Duin, Mayer, Hansen, & Fulton, 1988; Tabuchi, Maeda, Haruishi, Nakata, & Nishimura, 2003). Head cameras (with fields of view ranging from 60° to 100° diagonally as shown in Figure 2) do not capture the full visual field as defined. Again, the

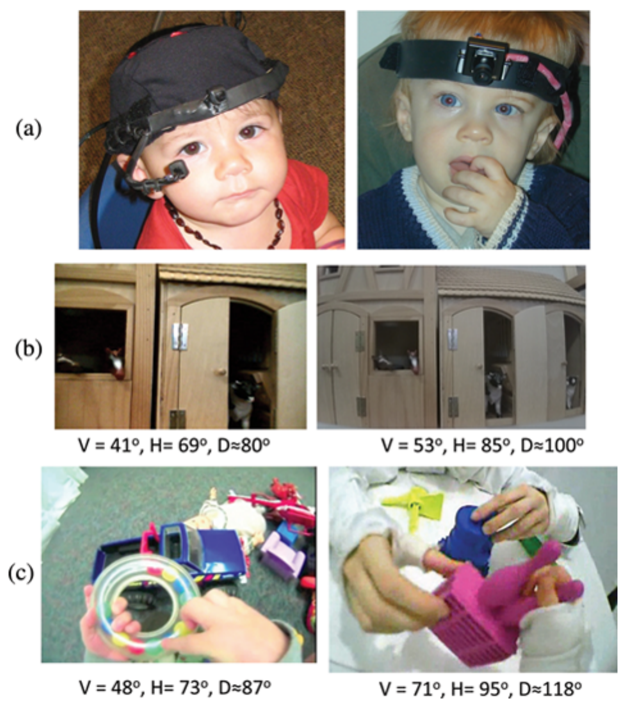

FIGURE 2 The panels in (a) show head and eye cameras on a hat (left) and a head camera on a band (right). The four panels in (b) and (c) show images from four different cameras with different vertical (V) and horizontal (H) fields of view (and the diagonal [D] measure of field of view). The two views in (b) were taken with each camera placed on a tripod 14 inches (14 cm) in front of a toy barn (a reachable distance for a toddler). The two views in (c) were taken while the head cameras were being worn by toddlers during toy play.

psychological relevance of the missed information is not clear because the *effective visual field* depends on the task (de Schonen, McKenzie, Maury, & Bresson, 1978; Ruff & Rothbart, 1996). In particular, the size of the effective visual field for an infant to detect a stimulus onset in the periphery will not be the same as that for discriminating objects, nor will it be the same in an empty field as in a crowded one (Farzin, Rivera, & Whitney, 2010; Whitney & Levi, 2011), when the perceiver is moving in three-dimensional space versus just watching events on a screen (Foulsham, Walker, & Kingstone, 2011), or when an attended object is held versus not held (Gozli, West, & Pratt, 2012). The developmental study of effective visual field sizes for different kinds of visual tasks is critical to understanding the utility and limitations of head-mounted cameras; it is also critical to understanding the development of visual processing. In sum, head cameras are imprecise in the timing of transitions between scenes and miss information at the edges of a scene; nonetheless, by measuring the scenes directly in front of infant faces, head cameras may capture the most important segment of the available information allowing researchers to study how the properties of visual scenes change with development and with activities.

ALIGNED HEADS AND EYES

One can have most confidence in the scenes captured by head cameras when the head and eyes are aligned. Critically, multiple lines of evidence also suggest that aligning the head and eyes is relevant to the effective *attentional* field. This idea is contrary to traditional approaches focused on eye gaze alone and that equate gaze direction and gaze duration with attention (Fantz, 1964). However, by both behavioral and neural measures, attention and looking are not the same (see Johnson, Posner, & Rothbart, 1991; Robertson, Watamura, & Wilbourn, 2012; Smith & Yu, 2013). Further, studies of adults indicate that aligning the head and eyes is better for visual processing than when the head and eyes are misaligned (e.g., Einhäuser et al., 2007; Jovancevic-Misic & Hayhoe, 2009; Thaler & Todd, 2009). If perceivers typically align their eyes and heads and if visual processing is optimal when their heads and eyes are aligned, then head cameras with their head-centered view may provide a measure of optimal views for attention and learning.

Consistent with this idea is evidence from research (using third-person camera views) to study infant visual attention during object play (e.g., Kannass & Oakes, 2008; Ruff & Capozzoli, 2003). These studies suggest that sustained attention is associated with minimal head movements and objects at midline, a posture consistent with aligned heads and eyes (Ruff & Capozzoli, 2003). If attention is optimal when the head and eyes are aligned and the attended object is at the child's midline, then head-camera images in which a target object is centered in the image should be indicative of optimal attention. Recent findings from head-camera studies support this prediction (Pereira, Smith, & Yu, 2013; Yu & Smith, 2012). In these studies, parents named objects as infants played with them. Subsequently, infants were tested to determine they had learned the names. The head-mounted camera images were analyzed to determine the properties of naming events that did and did not lead to learning. As shown in Figure 3, for learned object names, the named object was bigger in image size and more centered in the head-camera image compared with competitor objects. Moreover, for learned object names, the proximity and centering of the object were extended for several seconds before and after the parent had uttered the name. These results provide direct evidence for a role for joint head and eye direction in visual

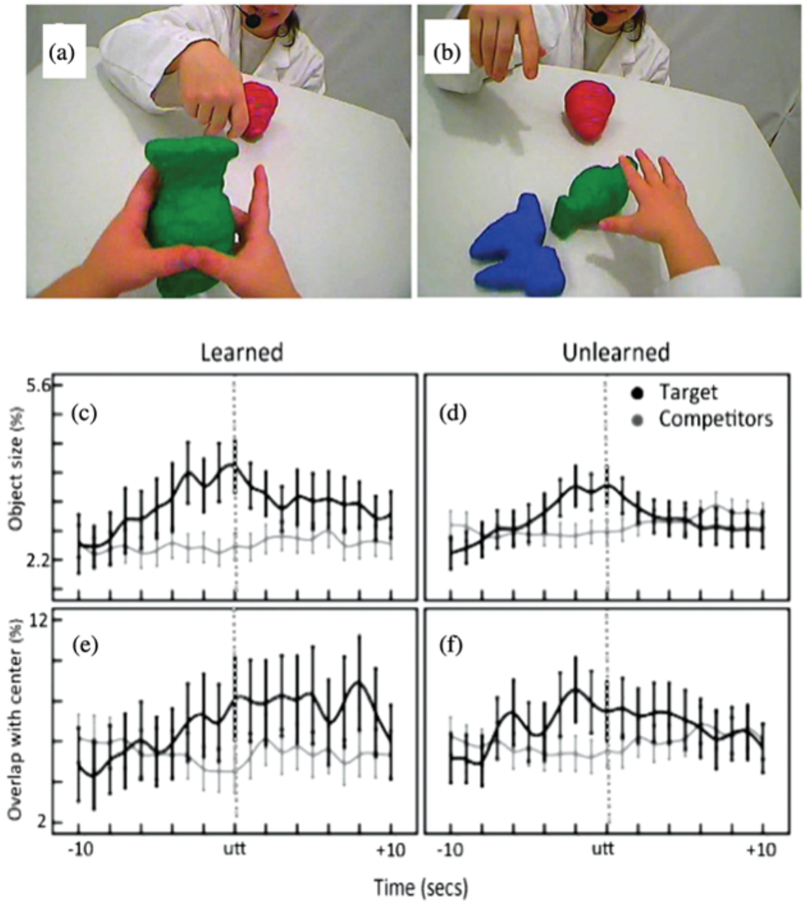

FIGURE 3 Results from head-camera studies linking visual size and centering of a named object to learning. Panels a and b show examples of head-camera images during two naming moments when later testing showed the child had learned the name (a) and not learned the name (b). Panels c and d show the image size (% pixels) of the named target object (black) and the mean of the other in-view, competitor objects (gray) for the 20-s window around the naming utterance (utt) for naming moments that led to the learning of the object name (c) or did not (d). Panels e and f show the overlap of the image of the named target (black) and the mean overlap of the images of the competitor objects (gray) with the center of the head-camera image for the 20-s window around the naming utterance (utt) for naming moments that led to the learning of the object name (e) or did not (f). See Yu and Smith (2012) and Pereira, Smith, and Yu (2013) for technical details and related graphs. Error bars represent standard errors.

processing and also illustrate how head cameras may provide insights beyond the contents of scenes and about the importance of the *stability* of those views.

In light of these issues, we have begun using head-mounted eye trackers to study how 13- to 24-month-old infants distribute eye gaze within the head-camera image (using the 118° diagonal head cameras as shown in Figure 2). The ‘‘heat maps’’ in Figure 4 show the gaze distribution within the head-camera image for infants for a 6-min session in which they were playing with

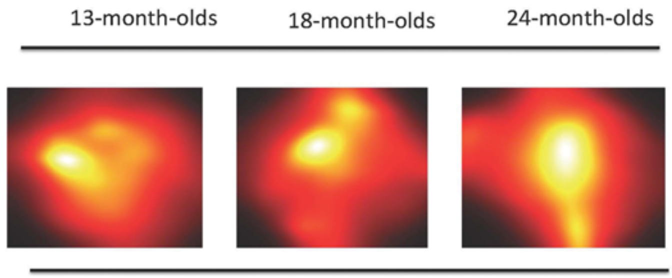

FIGURE 4 Gaze density as measured by an eye camera (low black, white high) within the head-camera images during a 6-min toy-play period for 13- ($n = 18$), 18- ($n = 18$), and 24-month-olds ($n = 16$).

toys on a table. The infants were free to move and they moved their heads often: More than 63% (*SD* = 10.9) of the time, head position was changing at a speed greater than 2 inches (5.08 cm) per second, and more than 71% (*SD* = 11.8) of the time, head rotation was changing at a speed of more than 30^{o} per second. Nonetheless, and as is evident in the heat maps in Figure 4, the distribution of eye gaze is organized in one region of the head-camera images, with more than 80% of gaze within the center (sized at 36% of the pixels) of the head-camera image. Although gaze distributions in broader contexts need to be measured, these data suggest that measures of the statistical properties of head-camera images may be sufficient to capture developmentally important contents. Comparisons of gaze distribution in adults wearing head-mounted eye trackers and acting in the world versus watching the same scenes on a screen also show that adults center their fixations when acting in a three-dimensional world; in contrast, when passively watching the same video on a screen, they distribute their eye gaze more widely (Foulsham et al., 2011). This is a reminder that what we know about gaze distributions from eye-tracking studies of infants looking at small screens while sitting in laboratories may not apply to gaze distributions when those same infants are acting in a three-dimensional world.

PRACTICAL MATTERS

Head cameras are not expensive (less than US$ 500 for everything excluding computers and servers for storage). There are a variety of small video cameras commercially available with different properties (see Frank, Vul, & Johnson, 2009; Smith et al., 2011; Sugden et al., 2013). The critical issues are field of view (in general, the larger the better; see Figure 2), distortion (more likely for wide-angle views), video storage (digital storage cards or cable to a computer are preferred as Wi-Fi and Bluetooth communication often fail), weight, and the ability to mount the camera in a way that infants tolerate.

Our success rate in placing head cameras on infants is about 75%. Success is very high with infants younger than 1 year and is more problematic in infants aged 15 to 18 months old. Placement is best done in one move; hesitation and multiple attempts increase the likelihood that the infant will refuse. However, experimenters who practice placing hats and devices on toddlers and parents (who have lots of experience putting hats on babies) can readily do this. Depending on the purpose of the experiment, we mount head cameras with and without eye trackers (as shown in Figure 2) on headbands or on hats. The critical issues for choosing how to mount the gear include: a) the ability to place the system on the child in a single move; b) placement low enough on the forehead for a front-of-face view; and c) no movement once placed. This last criterion is critical not just for the stability of the images captured, but it is also critical because if the headwear jiggles, toddlers notice and pull it off. We have found that anything that draws attention to the gear (including exploring the equipment or talking about it before placing it on the child) increases refusals. Placement is done in three steps: a) We desensitize the infant to hand actions near the head by asking the parent to lightly touch (or stroke) the child's head and hair several times. The experimenter who will be the ''placer'' does the same. b2) In the laboratory, we use three experimenters: one to place the head camera, the other to distract the child, and one to monitor the head-camera view. The experimenter places the head-mounted camera when the child is distracted with a push-button toy so that the child's hands are busy. The distracting experimenter or parent helps at this stage by gently pushing the child's hands toward the engaging toy so that they do not go to the head. c) When the child is clearly engaged with the toy, the placer tightens and adjusts the head camera. We adjust the camera so when the infant's hands make contact with the object, the object is centered in the head-camera field. For recording natural environments in the home, we fit a hat and camera to the individual infant, and then at home, parents put the hat on the child for recording.

DATA ANNOTATION

Head-mounted cameras, like traditional room cameras, yield a lot of data that have to be coded—a time-consuming task in which developmentalists are already experts. However, there are remarkable advances in computer-assisted hand-coding systems as well as more automatic analysis tools that may be able to help us with this task. We provide some leads here.

The **Datavyu Coding System** (originally Open-Shapa) is a free open-source event-based coding system that supports fine-grained dynamic and sequential hand coding of data and data analysis from very large data sets (see Adolph, Gilmore, Freeman, Sanderson, & Millman, 2012; Sanderson et al., 1994; http://datavyu.org).

There are a number of algorithm-assisted approaches to coding the contents of head cameras, including the coding of faces (Frank, 2012; Frank et al., 2009). One useful system is the **Visual Annotation Tool from Irvine California** (http://mit.edu/vondrick/vatic), a free, online, interactive video annotation tool for putting bounding boxes around objects to measure size and location (Vondrick, Patterson, & Ramanan, 2013). Advances in machine learning also make it possible to train automatic coding of specific classes of objects and their location in images (Fergus, Fei-Fei, Perona, & Zisserman, 2010; Smith et al., 2011).

The **Open Source Computer Vision** (http://opencv.org) library offers a whole toolbox for visual- and image-processing including measures of lower-level visual properties such as optical flow, motion vectors, and contrast. For relevant infant studies measuring optic flow patterns in head-camera images, see Burling, Yoshida, and Nagai (2013) and Raudies, Gilmore, Kretch, Franchak, and Adolph (2012). Finally, Itti, Koch, and Niebur (1998) proposed a procedure for creating **Salience Maps** from images that is widely used. Although their precise measures probably do not constitute a proper *psychological* description of stimulus salience for infants, the method provides a well-defined procedure through which to measure attention-getting properties of head-camera images.

SUMMARY

Head cameras measure the scene that is directly in front of the wearer. It seems highly likely that the statistical properties of these scenes play an important role in the development of the human visual system (Braddick & Atkinson, 2011). Moreover, vision is not just about eyes; because eyes are ''head-mounted,'' the coordination of the head and eyes plays a role in sustained attention and in learning. The unique contribution of head cameras is that they capture the head-centered child's view, *one* relevant view of the environment. However, there is still much we need to know to understand both the utility and limitations of this method. These open questions on limitations are also theoretically important questions about how heads, eyes, and bodies create visual environments.

ACKNOWLEDGMENT

We thank Sven Bambach for the gaze distribution data.

FUNDING

Support was provided in part by National Institutes of Health Grant 5R21HD068475.

REFERENCES

Adolph, K. E., Gilmore, R. O., Freeman, C., Sanderson, P., & Millman, D. (2012). Toward open behavioral science. *Psychological Inquiry*, *23*, 244–247.

Adolph, K. E., Tamis-LeMonda, C. S., Ishak, S., Karasik, L. B., & Lobo, S. A. (2008). Locomotor experience and use of social information are posture specific. *Developmental Psychology*, *44*, 1705–1714.

Aslin, R. N. (2008). Headed in the right direction: A commentary on Yoshida and Smith. *Infancy*, *13*, 275–278.

Aslin, R. N. (2012). Infant eyes: A window on cognitive development. *Infancy*, *17*, 126–140.

Bertenthal, B. I., & Campos, J. J. (1990). A systems approach to the organizing effects of self-produced locomotion during infancy. *Advances in Infancy Research*, *6*, 1–60. Retrieved from http://psycnet.apa.org/psycinfo/1990-30563-001

Bloch, H., & Carchon, I. (1992). On the onset of eye–head coordination in infants. *Behavioural Brain Research*, *49*, 85–90.

Braddick, O., & Atkinson, J. (2011). Development of human visual function. *Vision Research*, *51*, 1588–1609.

Burling, J. M., Yoshida, H., & Nagai, Y. (2013, August). The significance of social input, early motion experiences, and attentional selection. In *Proceedings of the 3rd IEEE International Conference on Development and Learning and on Epigenetic Robotics* (pp. 260–262).

Byrge, L., Sporns, O., & Smith, L. B. (2014). Developmental process emerges from extended brain-body-behavior networks. *Trends in Cognitive Science*, *18*(8), 395–403.

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences*, *110*, 11278–11283.

Cicchino, J. B., Aslin, R. N., & Rakison, D. H. (2011). Correspondences between what infants see and know about causal and self-propelled motion. *Cognition*, *118*, 171–192.
Corbetta, D., Guan, Y., & Williams, J. L. (2012). Infant eye tracking in the context of goal-directed actions. *Infancy*, *17*, 102–125.
Cummings, M. F., Van Hof-Van Duin, J., Mayer, D. L., Hansen, R. M., & Fulton, A. B. (1988). Visual fields of young children. *Behavioural Brain Research*, *29*, 7–16.
Daniel, B. M., & Lee, D. N. (1990). Development of looking with head and eyes. *Journal of Experimental Child Psychology*, *50*, 200–216.
de Schonen, S., McKenzie, B., Maury, L., & Bresson, F. (1978). Central and peripheral object distances as determinants of the effective visual field in early infancy. *Perception*, *7*, 499–506.
Einhäuser, W., Schumann, F., Bardins, S., Bartl, K., Böning, G., Schneider, E., & König, P. (2007). Human eye–head co-ordination in natural exploration. *Network: Computation in Neural Systems*, *18*, 267–297.
Fantz, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, *146*, 668–670.
Farzin, F., Rivera, S. M., & Whitney, D. (2010). Spatial resolution of conscious visual perception in infants. *Psychological Science*, *21*, 1502–1509.
Fergus, R., Fei-Fei, L., Perona, P., & Zisserman, A. (2010). Learning object categories from Internet image searches. *Proceedings of the IEEE*, *98*, 1453–1466.
Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*, 1920–1931.
Franchak, J. M., & Adolph, K. E. (2012). What infants know and what they do: Perceiving possibilities for walking through openings. *Developmental Psychology*, *48*, 1254–1261.
Frank, M. C. (2012). Measuring children's visual access to social information using face detection (N. Miyake, D. Peebles, & R. P. Cooper, Eds.). In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, Austin, TX.
Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*, 160–170.
Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the development of social attention using free viewing. *Infancy*, *17*, 355–375.
Gozli, D. G., West, G. L., & Pratt, J. (2012). Hand position alters vision by biasing processing through different visual pathways. *Cognition*, *124*(2), 244–250.
Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press
Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*, 1254–1259.
James, K. H., Jones, S. S., Smith, L. B., & Swain, S. N. (2013). Young children's self-generated object views and object recognition. *Journal of Cognition and Development*, *15*, 393–401. doi: 10.1080/15248372.2012.749481
Jayaraman, S., Fausey, C. M., & Smith, L. B. (2013). Visual statistics of infants' ordered experiences. *Journal of Vision*, *13*, 735.
Johnson, M. H., Posner, M. I., & Rothbart, M. K. (1991). Components of visual orienting in early infancy: Contingency learning, anticipatory looking, and disengaging. *Journal of Cognitive Neuroscience*, *3*, 335–344.
Jovancevic-Misic, J., & Hayhoe, M. (2009). Adaptive gaze control in natural environments. *Journal of Neuroscience*, *29*, 6234–6238.
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*, 222–246.
Kretch, K., Franchak, J., Brothers, J., & Adolph, K. (2012). What infants see depends on locomotor posture. *Journal of Vision*, *12*, 182.
Murray, K., Lillakas, L., Weber, R., Moore, S., & Irving, E. (2007). Development of head movement propensity in 4–15 year old children in response to visual step stimuli. *Experimental Brain Research*, *177*, 15–20.
Nakagawa, A., & Sukigara, M. (2013). Variable coordination of eye and head movements during the early development of attention: A longitudinal study of infants aged 12–36 months. *Infant Behavior and Development*, *36*, 517–525.
Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review*, *117*, 382–405.

Nyström, M., & Holmqvist, K. (2010). An adaptive algorithm for fixation, saccade, and glissade detection in eyetracking data. *Behavior Research Methods*, *42*, 188–204.

Pereira, A. F., Smith, L. B., & Yu, C. (2013). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*, 178–185.

Raudies, F., Gilmore, R. O., Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2012, November). Understanding the development of motion processing by characterizing optic flow experienced by infants and their mothers (pp. 122–124). In *Proceedings of the 2nd IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL), 2012 IEEE International Conference*.

Robertson, S. S., Watamura, S. E., & Wilbourn, M. P. (2012). Attentional dynamics of infant visual foraging. *Proceedings of the National Academy of Sciences*, *109*, 11460–11464.

Rodriguez, E. T., & Tamis-LeMonda, C. S. (2011). Trajectories of the home learning environment across the first 5 years: Associations with children's vocabulary and literacy skills at prekindergarten. *Child Development*, *82*, 1058–1075.

Ruff, H. A., & Capozzoli, M. C. (2003). Development of attention and distractibility in the first 4 years of life. *Developmental Psychology*, *39*, 877–890.

Ruff, H. A., & Rothbart, M. K. (1996). *Attention in early development*. New York, NY: Oxford University Press.

Sanderson, P. M., Scott, J. J. P., Johnston, T., Mainzer, J., Wantanbe, L. M., & James, J. M. (1994). MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA). *International Journal of Human-Computer Studies*, *41*, 633–681. doi:10.1006/ijhc.1994.1077

Schmitow, C., Stenberg, G., Billard, A., & von Hofsten, C. (2013). Using a head-mounted camera to infer attention direction. *International Journal of Behavioral Development*, *37*, 468–474.

Simoncelli, E. P. (2003). Vision and the statistics of the visual environment. *Current Opinion in Neurobiology*, *13*, 144–149.

Smith, L. B. (2013). It's all connected: Pathways in visual object recognition and early noun learning. *American Psychologist*, *68*, 618–629.

Smith, L. B., & Yu, C. (2013). Visual attention is not enough: Individual differences in statistical word-referent learning in infants. *Language Learning and Development*, *9*, 25–49.

Smith, L. B., Yu, C., & Pereira, A. F. (2011). Not your mother's view: The dynamics of toddler visual experience. *Developmental Science*, *14*, 9–17.

Soska, K. C., & Adolph, K. E. (2014). Postural position constrains multimodal object exploration in infants. *Infancy*, *19*, 138–161.

Sugden, N. A., Mohamed-Ali, M. I., & Moulson, M. C. (2013). I spy with my little eye: Typical, daily exposure to faces documented from a first-person infant perspective. *Developmental Psychobiology*, *56*, 249–261.

Tabuchi, A., Maeda, F., Haruishi, K., Nakata, D., & Nishimura, N. (2003). Visual field in young children. *Neuro-Ophthalmology*, *20*, 342–349.

Tatler, B. W., Hayhoe, M. M., Land, M. F., & Ballard, D. H. (2011). Eye guidance in natural vision: Reinterpreting salience. *Journal of Vision*, *11*.

Thaler, L., & Todd, J. T. (2009). The use of head/eye-centered, hand-centered and allocentric representations for visually guided hand movements and perceptual judgments. *Neuropsychologia*, *47*, 1227–1244.

Vondrick, C., Patterson, D., & Ramanan, D. (2013). Efficiently scaling up crowdsourced video annotation. *International Journal of Computer Vision*, *101*, 184–204.

von Hofsten, C., Vishton, P., Spelke, E. S., Feng, Q., & Rosander, K. (1998). Predictive action in infancy: Tracking and reaching for moving objects. *Cognition*, *67*, 255–285.

Wass, S. V., Smith, T. J., & Johnson, M. H. (2013). Parsing eye-tracking data of variable quality to provide accurate fixation duration estimates in infants and adults. *Behavior Research Methods*, *45*, 229–250.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*, 160–168.

Yoshida, H., & Smith, L. B. (2008). What's in view for toddlers? Using a head camera to study visual experience. *Infancy*, *13*(3), 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. doi: 10.1371/journal.pone.0079659

Yurovsky, D., Smith, L. B., & Yu, C. (2013). Statistical word learning at scale: The baby's view is better. *Developmental Science*, *16*, 959–966.