Developmental and postural changes in children's visual access to faces

Michael C. Frank, Kaia Simmons, Daniel Yurovsky, & Guido Pusiol
{mcfrank,kaias,yurovsky,pusiol}@stanford.edu
Department of Psychology
Stanford University

Abstract

The faces of other people are a critical information source for young children. During early development, children undergo undergoing significant postural and locomotor development, changing from lying and sitting infants to toddlers. We used a head-mounted camera in conjunction with a face-detection system to explore the effects of these changes on children's visual access to their caregivers' faces during an in-lab play session. In a cross-sectional sample of 4–20 month old children playing with their caregivers, we found substantial changes in face accessibility based on age and posture. These changes may translate into changes in the accessibility of social information during language learning.

Keywords: Social development; face processing; head-camera.

Introduction

A father offers his young daughter novel object: a bright yellow feather duster. A few moments after she accepts the toy, he remarks, "Isn't the zem funny?" Her father may still be talking about the feather duster, or he may be describing a new object, but to find out she has access to a simple and reliable method: she can look to his face to infer the direction of his gaze.

The ability to follow social signals like eye-gaze is an important part of early social cognition (Scaife & Bruner, 1975) and a strong predictor of children's early language development. For example, Brooks and Meltzoff (2005) found that children who followed an experimenter's gaze better before their first birthday had larger vocabularies at 18 months. Similarly, Carpenter, Nagell, and Tomasello (1998) found that children's level of joint engagement (as well as the degree to which mothers followed the child's focus of attention in their labeling) predicted vocabulary growth in both language production and comprehension. These studies suggest that children's social environment plays a powerful supportive role in language learning.

But at the same time as children are beginning to learn their first words, their view of the world is changing radically (Adolph & Berger, 2007). As speechless infants, they are unable to locomote independently and are moved from place to place and set in particular postures by their caregivers. Before their first birthday, they begin crawling; soon after, they begin to walk independently. These changes may have a profound effect on what children see.

A recent study suggests the possibility of links between motor milestones, social cognition, and language. Walle and Campos (under review) noted robust correlations between children's ability to walk and their vocabulary, both receptive and productive. On the basis of an observational study of parent input, they speculated that the emergence of walking may change the ability of the child to use and appreciate a number of social cues. One possible explanation for this pattern is that children's posture may mediate their access to social cues like eye gaze. Seeing more social information may in turn allow children to discover word meanings more effectively.

Recent methodological developments provide data that allow this hypothesis to be tested. The availability of head-mounted cameras and eye-trackers allows for the measurement of children's naturalistic environment in a way that was not previously possible. Yoshida and Smith (2008) gave the first demonstration of the radical differences between toddler and adult perspectives on the social world, with toddlers' visual field being dominated by hands and objects much more than that of adults. More recent work has used head-mounted eye-tracking methods to measure young toddlers' fixations (Franchak, Kretch, Soska, & Adolph, 2011), also finding that children look relatively infrequently at their mothers' faces in naturalistic play.

These methods are now being applied to understand inputs to language acquisition. Work by Yu, Smith, and colleagues suggests that word learning is facilitated when parents and children create moments in which the visual field is dominated by a single object (Smith, Yu, & Pereira, 2011; Yu & Smith, in press). Some data even suggest that young children's restricted viewpoint may be more effective for learning words than the comparable adult perspective (Yurovsky, Smith, & Yu, in press). Together, this body of evidence suggests that understanding infants' perspective during language use—and how it interacts with their development—is a critical part of understanding language learning.

In the current study we took a developmental approach to understanding the relationship between perspective and access to faces. We recorded head-camera data from a group of infants and children across a broad developmental range as they played with their caregivers during a brief laboratory visit. We then hand-annotated these data for the child's posture and used face-detection algorithms to measure the frequency of faces in the child's visual field. The resulting dataset allows us to analyze changes in access to faces according to children's age and posture.

Methods

Participants

Participants were 20 infants and children (N=4 each at 4, 8, 12, 16, and 20 months, 9 females total) in an ongoing large-scale study, recruited from the surrounding community via

Figure 1: Our light-weight, low-cost head-mounted camera with fisheye lens.

state birth records. Participants had no documented disabilities and were reported to hear at least 80% English at home. Success rates for children wearing the camera varied from 100% at 8 months to approximately 50% at 20 months.

Head-mounted camera

Our head-mounted camera ("headcam") is composed of a small, inexpensive MD80 model camera attached to a soft elastic headband from a camping headlamp. An aftermarket fisheye lens intended for iPhones and other Apple devices is attached to increase view angle. The total cost of each camera is approximately $60. The camera captures 720x480 pixel images at approx. 25 frames per second, and has battery life of 60–90 minutes. With the fisheye lens affixed, it has a viewing angle of approx. 60 degrees of visual angle in the vertical axis. The device is pictured in Figure 1.

The vertical field of view of the camera was smaller than the child's approximate vertical field of view, which—even at 6-7 months—spans around 100–120 degrees (Mayer, Fulton, & Cummings, 1988; Cummings, Van Hof-Van Duin, Mayer, Hansen, & Fulton, 1988). We were therefore faced with a choice in the orientation of the camera. If we chose a lower or higher orientation, we would be at risk of truncating either the child's own hands and physically proximate objects, or the faces of the adults around the child. Yet if we chose the middle orientation, we would still be at risk of underestimating the proportion of faces viewed by the child. Thus, for the purposes of the current study—measuring visual access to faces—we chose to orient the camera towards the upper part of the visual field. While this orientation decreased our chances of recording the objects being manipulated by the child, it nevertheless allowed us to capture the majority of the faces in the child's visual field.

Procedure

After coming to the lab, families were seated in our waiting room where they signed consent documents and where children were fitted with the headcam. After a short period of play, they were escorted to a playroom in the lab where the free-play session (the focus of the current study) was conducted.

In the waiting room, the experimenter placed the headcam on children's heads after they had time to adjust to the environment. For children who resisted wearing the headcam, the experimenter used "distractor" techniques (presenting stimulating toys or engaging the children in hand-occupying activities) intended to keep children's focus elsewhere and prevent them from taking off the camera (Yoshida & Smith, 2008). Once the child was wearing the camera comfortably for a period of time, child and caregiver (or caregivers: in two cases, there were two adults present) were escorted to the playroom.

In the playroom, the experimenter presented the child's parent with a box containing three labeled pairs of objects, each consisting of a familiar and a novel object: a ball and a feather duster ("zem"), a toy car and an XYZ (blah) and an XYZ and a (blah). Parents confirmed that the child had not previously seen the novel toys. Parents were instructed to play with the object pairs with the child, one at a time, "as they typically would" and to use the novel labels to refer to the three toys. After giving these instructions, the experimenter left the room for a period of approximately 15 minutes. During this time, a tripod-mounted camera recorded video from a corner of the room and the headcam also captured video from the child's perspective.

Data Processing and Annotation

All headcam videos were cropped to exclude the period of entry to the playroom and were automatically synchronized with the tripod-mounted videos using FinalCut Pro Software. The final sample was approx. 5 hours of headcam video (M = 12 min, range: 2–21 min), for a total of roughly 400,000 frames.

Posture and Orientation Annotation One major goal of our study was to understand the relationship between children's posture and their access to information from the faces of their caregiver. To investigate this relationship, we created a set of hand annotations for the child's physical posture (e.g. standing, sitting) and orientation relative to the caregiver (e.g. in front of, behind, close, far away). For each headcam video, a coder used OpenSHAPA software to annotate both orientation and posture (Adolph, Gilmore, Freeman, Sanderson, & Millman, 2012).

Orientation was initially categorized as being in front of the caregiver, to the side, or behind, and close (defined informally as within arm's reach) or farther away. Because of data sparsity, we consolidated this scheme into three categories: close to the caregiver either in front or on the side, farther from the caregiver either in front or to the side, and a global category of behind the caregiver. Posture was categorized as being held/carried, lying on back, sitting, prone (crawling or lying), standing, or other. Data from when the child was out of view of the tripod camera was marked as uncodable and excluded from these annotations. A second coder coded XYZ

[1] Previous studies have shown that children's head movements in the horizontal dimension are approximated by (though are slightly lagged by) their head movements (Yoshida & Smith, 2008). Our own experience with the current apparatus ratifies these conclusions for the horizontal field but suggests that head movements in the vertical field are less reliable.

Figure 2: Sample frames from the headcam videos for a child from each age group, selected because they featured successful face detections (green squares).

videos; their categorizations were reliable at XYZ.

Face presence gold standard In order to evaluate our face detection algorithms (described below), we annotated a sample of frames from each video. DAN PARAGRAPH

Face Detection

A second goal was to measure the presence of caregivers' faces in the child's field of view (as approximated by the headcam). In order to avoid hand-annotating the size and position of faces in every frame of video, we tested two face detection systems. Sample frames from the video with successful detections are given in Figure 2.

The first algorithm was based on freely available tools (Bradski & Kaehler, 2008), described in depth in our previous work (Frank, 2012). This system had two parts. The first was the application of a set of four Haar-style face detection filters from the OpenCV library (Viola & Jones, 2004) to each frame of the videos independently. These detectors each provide information about whether a face is present in the frame as well as size and position for any detections. In a second step, these detections are then combined via a hidden Markov model (HMM), trained via annotated data. The HMM model (which performed nearly as well as the more complex and computationally-intensive Conditional Random Field model used in our previous work) attempted to estimate whether a face was truly present in each frame of the videos, using as its input the number of Haar detectors that were active in any given frame.

The second model we evaluated was a model based on ... GUIDO PARAGRAPH (Kalal, Mikolajczyk, & Matas, 2010)

We evaluated each algorithm on their precision (hits / hits + false alarms) and recall (hits / hits + misses), as well as F-score (the harmonic mean of these two measures), using the gold-standard data described above. Results are reported in Table 1. The HMM model obtained a relatively high level of performance for the random subsections, but performed poorly when there was a relatively high density of faces present. In contrast, LBP performed well on both samples, giving better performance especially in cases where there was partial occlusion.

Overall, we used face-detection algorithms was to provide a measurement technique that eliminated tedious and expensive hand-coding and provided acceptable results. We therefore selected the LBP model and report detections from this algorithm as an estimate of face presence in all further analyses.

Table 1: Model performance on gold standard generalization training set dataset. P, R, and F denote precision, recall, and F-score.

	High-density			Random sample		
Model	P	R	F	P	R	F
HMM	.55	.38	.45	.89	.74	.81
LBP	.86	.78	.81	.93	.76	.83

Results

We report results from three different sets of analyses. First, we explore developmental changes in posture and orientation in our dataset. Next, we explore how these changes affect access to faces, as measured using our face-detection algorithm.

Changes in Posture and Orientation

Our posture coding captured typical developmental milestones (Figure 3, left). Overall, sitting was the most common posture for interactions in the caregiver play session. The youngest infants in our sample mostly sat (with parental assistance), but also lay down and were carried a significant proportion of the time. The 12-month-olds were the only group who spent a large amount of time crawling, and the 16- and 20-month-olds sat and stood in equal parts.

Similarly, our coding of orientation revealed some significant developmental changes (Figure 3, right). Younger children more frequently had the caregiver behind them, often because the caregiver was supporting the child's sitting posture (for the 4-month-olds especially). In contrast, the 12–20 month olds were able to locomote independently and so were able to spend more time further from the caregiver.

Access to Faces

We next investigated the effects of the child's posture and orientation on the presence and size of the caregiver's face in the visual field. Figure 4 shows the proportion of frames with a positive face detection, plotted by the child's age, posture, and orientation relative to the caregiver. Detailed results of

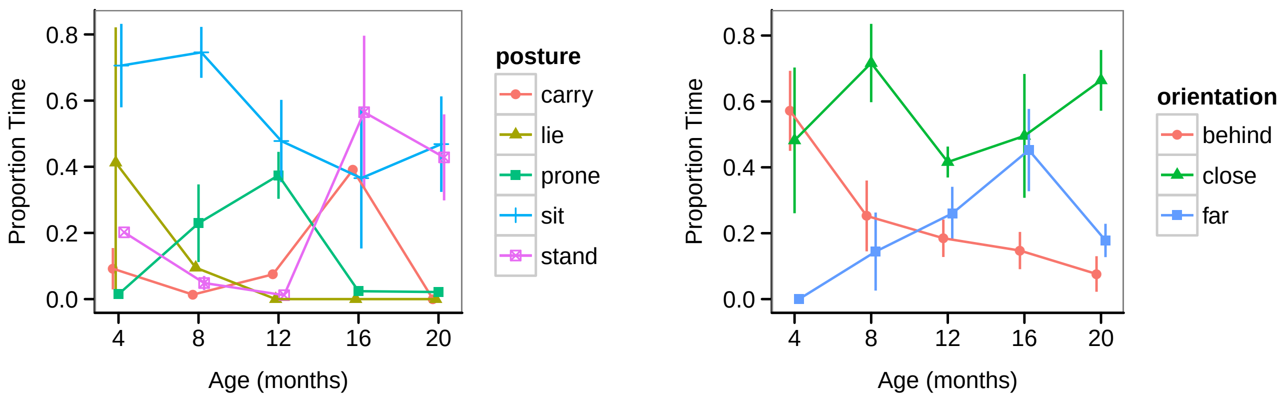

Figure 3: Proportion time in each posture, plotted by child's age (left panel). Proportion time in each orientation relative to the caregiver, again plotted by child's age (right panel). For clarity, the "other" code is not plotted in either figure. Error bars show standard error of the mean across participants.

face detection by posture and orientation are given in Figure **??**.

Overall, there were very large differences in access to faces across age. The four-month-olds saw almost no faces–their parents were behind them most of the time, supporting them as they could not sit themselves. In contrast, the eight-month-olds, who could sit independently, typically sat across from their caregiver and saw many faces in both the sitting and prone postures. The twelve-month-olds spent a large amount of time in the prone position (typically crawling after the ball, for example) and saw almost no faces in that posture. The sixteen- and twenty-month-olds saw many faces, often because they were standing while their parents were sitting.

Across ages, the carrying and prone postures resulted in the smallest number of faces seen, while standing and sitting resulted in far more. These postures both presented opportunities for seeing faces in large part because parents were sitting or lying on the floor with children. Although far fewer faces were seen when the caregiver was behind the child,[2] both the close and far positions resulted in approximately equal proportions of face detections.

General Discussion

Using a head-mounted camera, we explored the relationship between infants' postural and locomotor development and their visual access to social information. The use of automated annotation tools from computer vision allowed us to measure the prevalence of caregivers' faces in their children's visual field. We found systematic differences in the visual accessibility of faces based on the child's posture, orientation, and age.

Our study complements a wide variety of work that has investigated the relationship between children's perspective and their access to social and referential information (Yoshida & Smith, 2008; Franchak et al., 2011; Smith et al., 2011; Yu & Smith, in press; Walle & Campos, under review).

The measures developed here have broad applicability to the study of individual and cultural differences. Since the physical circumstances of child rearing vary widely across households and across cultures with different material circumstances, there may be predictable differences in children's visual experience. As suggested by the correlations between walking and vocabulary (Walle & Campos, under review), shifts in how infants are placed in particular postures by strollers or carriers (Zeedyk, 2008) or how much exercise they are given (Bril & Sabatier, 1986).

While infants' visual field is often subject to the whims of their caregivers, toddlers determine their own input to a much greater degree. Toddlers live in a world populated by knees.

[2] Since orientation was coded via body posture, faces seen while the caregiver was behind were due to caregivers turning their heads to look at children who have run behind them.

Acknowledgments

Thanks to Ally Kraus, Kathy Woo, Aditi Maliwal, and other members of the Language and Cognition Lab for help in recruitment, data collection, and annotation. This research was supported by a John Merck Scholars grant to MCF.

References

Adolph, K., & Berger, S. (2007). Motor development. *Handbook of child psychology*.

Adolph, K., Gilmore, R., Freeman, C., Sanderson, P., & Millman, D. (2012). Toward open behavioral science. *Psychological Inquiry*, *23*(3), 244–247.

Bradski, G., & Kaehler, A. (2008). *Learning opencv: Computer vision with the opencv library*. O'Reilly Media.

Bril, B., & Sabatier, C. (1986). The cultural context of motor development: Postural manipulations in the daily life of bambara babies (mali). *International Journal of Behavioral Development*, *9*(4), 439–453.

Brooks, R., & Meltzoff, A. (2005). The development of gaze following and its relation to language. *Developmental Science*, *8*(6), 535–543.

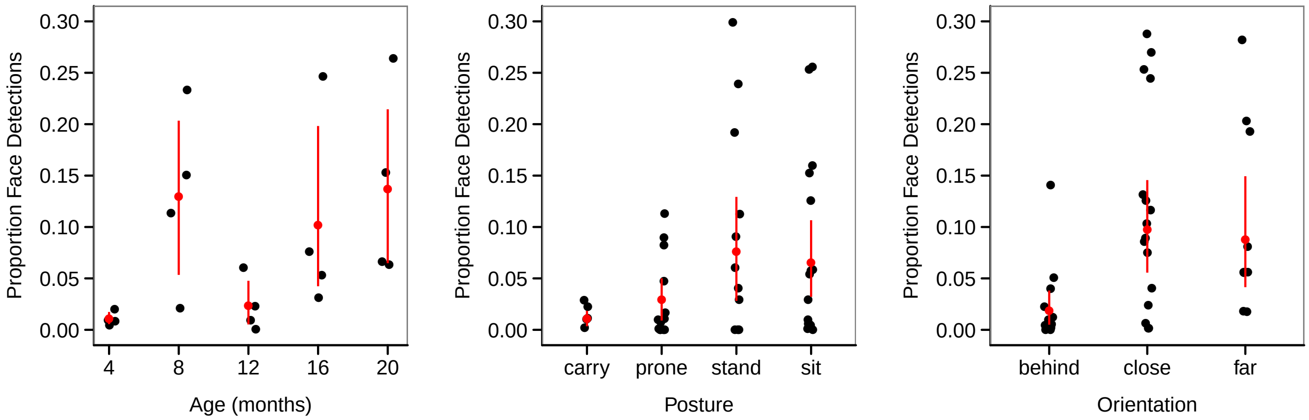

Figure 4: Proportion face detections, split by age group (left panel), posture (middle panel), and caregiver's orientation (right panel). Black points show individual participants and are jittered slightly on the horizontal, red lines show means and 95% confidence intervals.

Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the society for research in child development*, *63*(4).

Cummings, M., Van Hof-Van Duin, J., Mayer, D., Hansen, R., & Fulton, A. (1988). Visual fields of young children. *Behavioural brain research*, *29*(1), 7–16.

Franchak, J., Kretch, K., Soska, K., & Adolph, K. (2011). Head-mounted eye tracking: A new method to describe infant looking. *Child development*.

Frank, M. C. (2012). Measuring children's visual access to social information using face detection. In *Proceedings of the 33nd Annual Meeting of the Cognitive Science Society.*

Kalal, Z., Mikolajczyk, K., & Matas, J. (2010). Forward-backward error: Automatic detection of tracking failures. In *Pattern recognition (icpr), 2010 20th international conference on* (pp. 2756–2759).

Mayer, D., Fulton, A., & Cummings, M. (1988). Visual fields of infants assessed with a new perimetric technique. *Investigative ophthalmology & visual science*, *29*(3), 452–459.

Scaife, M., & Bruner, J. (1975). The capacity for joint visual attention in the infant. *Nature*.

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.

Viola, P., & Jones, D. H. (2004). Robust real-time face detection. *International Journal of Computer Vision*.

Walle, E. A., & Campos, J. J. (under review). Infant language development is related to the acquisition of walking.

Yoshida, H., & Smith, L. (2008). What's in view for toddlers? using a head camera to study visual experience. *Infancy*, *13*, 229–248.

Yu, C., & Smith, L. B. (in press). Embodied attention and word learning by toddlers. *Cognition*.

Yurovsky, D., Smith, L., & Yu, C. (in press). Statistical word learning at scale: The baby's view is better. *Developmental Science*.

Zeedyk, M. (2008). *Whats life in a baby buggy like?: The impact of buggy orientation on parent-infant interaction and infant stress* (Tech. Rep.). National Literacy Trust.