Gaze is not Enough: Computational Analysis of Infant's Head Movement Measures the Developing Response to Social Interaction

Lars Schillingmann (lars@ams.eng.osaka-u.ac.jp)
Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan

Joseph M. Burling (jmburling@uh.edu)
Department of Psychology, University of Houston, 126 Heyne Building Houston, TX 77204-5022 USA

Hanako Yoshida (yoshida@uh.edu)
Department of Psychology, University of Houston, 126 Heyne Building Houston, TX 77204-5022 USA

Yukie Nagai (yukie@ams.eng.osaka-u.ac.jp)
Graduate School of Engineering, Osaka University, 2-1 Yamadaoka, Suita, Osaka, 565-0871 Japan

Abstract

Infant eye gaze is frequently studied because of its relevance as an indicator of early attention and learning. However, the coupling of eye gaze with an individual's head motion is often overlooked. This paper analyzes how head motion develops within a social interaction context. To measure this interaction, we developed an approach that can estimate infant head motion from ego perspective recordings as they are typically provided by eye-tracking systems. Our method is able to quantify infant head motion from existing social interaction recordings even if the head was not explicitly tracked. Therefore, data from longitudinal studies that has been collected over the years can be reanalyzed in more detail. We applied our method to an existing longitudinal study of parent infant interaction and found that infants' head motion in response to social interaction shows a developmental trend. Furthermore, our results indicate that this trend is less visible within gaze data alone. This suggests that head motion is an important element for understanding and measuring infants' behavior during parent-child interactions.

Keywords: head motion; gaze; computational analysis; parent infant interaction

Introduction

Head control is important developmental milestone for human infants. We move our head when we track objects and must be able to coordinate it with our body when reaching and grasping. The development of such capabilities can be seen in the first few months of infancy where infants' head movement patterns change and become more controlled around 3 months of age (de Lima-Alvarez, Tudella, van der Kamp, & Savelsbergh, 2014). When 6-month old infants follow a target, on average their head moves nearly as much as the object does (Jonsson & Von Hofsten, 2003). Furthermore, head control is an important factor in the development of reaching to provide a stable support for gazing at the target (Bertenthal & von Hofsten, 1998).

Head movements are also relevant for social communication. For children it is crucial to be able to follow their caretaker's social references. This ability does not only require following objects by eye movements alone, but also performing large shifts in head orientation in a short amount of time. For example, during gaze switching between the caretaker's face and an object. Eye gaze alone is not sufficient to reach this flexibility in all situations. What we do not yet know is how social interactions between the parent and infant play a role in facilitating the development of head movements. In this paper, we investigate how the development of these kinds of head movements emerges from specific social referencing contexts of parents naming and acting on objects. We use a corpus that was a part of a lager project conducted by Yoshida and Burling (2013). Infants from 6 to 24 months of age play with their mother who shows toys and objects from a predefined set. This allows analyzing infant head motion while differentiating interaction conditions—for example, when the parent is holding the object.

Figure 1: Visualization tool displaying a history of frames with corrected horizontal and vertical camera motion.

Like a number of observation studies aiming to track children's attention, this data set includes measurements of infants' gaze by using a head-mounted eye-tracking system, but does not contain instrumentation for detecting head position. The present study specifically considers this limitation. A trivial solution would be to conduct another study and include additional sensors or a tracking system to record the infants head pose. However, longitudinal studies are time and resource intensive. Fur-

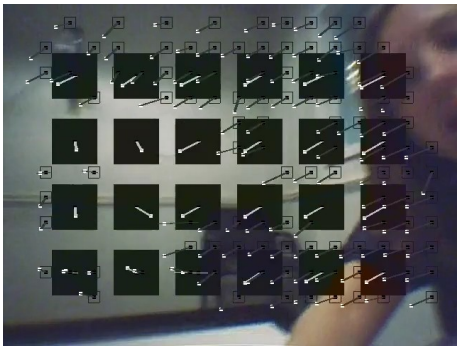

Figure 2: Debug visualization of the vid.stab library (Martius, 2014) showing blocks with sufficient contrast and estimated vectors.

thermore, some tracking systems can induce additional difficulties in conducting a study, as they might distract infants. A common option to solve this problem is to rely on third person view recordings of the infant and to apply video-based tracking methods (Delaherche et al., 2012). However, third person view recordings, if not taken from an ideal perspective, can contain segments where the tracking target is occluded. Another problem is that low resolution and the absence of markers can pose difficulties for accurate head tracking. Therefore, we will propose a method that is able to directly extract head movements from a head mounted eye-tracking system. The current approach also enables previously collected data to be reanalyzed using this new method. Usually, head mounted eye-tracking systems record an ego perspective video which is subsequently used to overlay the tracking data, so it can be understood where the person was looking. Head movements can be estimated from a head mounted camera by using optical flow to find shifts between subsequent frames. Similar techniques have already been applied in analyzing differences between parent and infant visual experiences (Raudies, Gilmore, Kretch, Franchak, & Adolph, 2012). We will demonstrate how open source video stabilization software can be used to estimate head motion from ego perspective recordings.

In summary, this paper will address the following questions. Does infant head motion change over the course of development under different interaction conditions? Can this trend also be found on gaze data alone? How can we measure head motion from ego perspective recordings?

Head Motion Extraction Based on Video Stabilization Techniques

In the following sections, we describe our method of estimating children's head motion from head-mounted camera recordings. The process is depicted in Figure 3.

Cropping and Calibration Before the videos were processed we cropped and calibrated the video data to remove black borders and lens distortion using standard

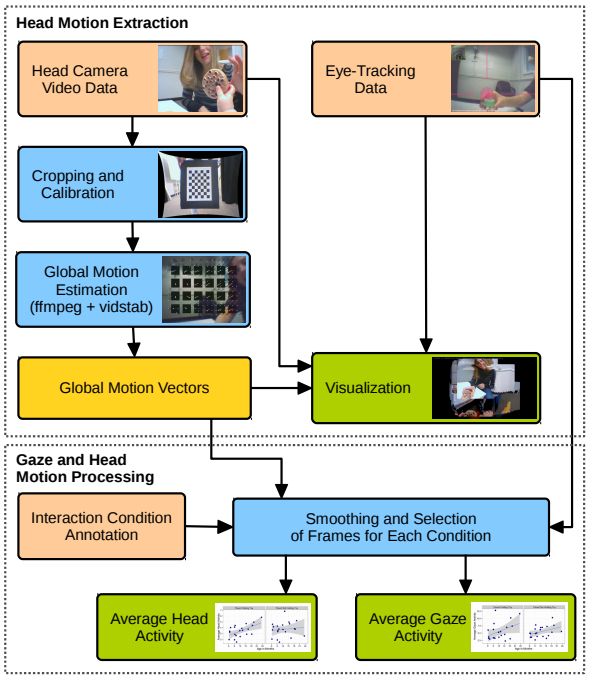

Figure 3: Process of estimating head motion and calculating the average head and gaze activity.

OpenCV methods (Bradski, 2000). The calibration information is obtained from a separately recorded calibration pattern.

Global Motion Estimation The child centered view recorded by the eye-tracking device reflects head motion during infant head rotations. Assuming an infant does not change location, the global motion from frame to frame provides an estimation of the horizontal and vertical shift due to head rotation. Stabilizing a shaky video requires frame-by-frame estimation of global motion to shift each frame to compensate the shaking motion accordingly. Open-source software can be used for solving this problem. We used the vid.stab library (Martius, 2014) that can be used by FFmpeg (*FFmpeg*, 2014) to motion compensate shaky video. Usually the main task of the library is to estimate global motion transformations in a first pass and then apply them in a second pass to render a new deshaked video. For our purpose, we use the library's debug output to retrieve the global motion vectors. The advantage of using a video stabilization library is that the implementation is already tuned to the task of global motion estimation. The vid.stab library uses several heuristics to minimize noise due to low contrast areas and moving elements in the video. The basic approach relies on block matching. First, a coarse-grained motion search using a large block size is performed. Subsequently, the search is refined towards local motion by using smaller blocks (see Figure 2). The global motion is estimated by finding a transformation that minimizes the error to the local motion fields. Outliers that potentially correspond to moving objects are excluded. A possible limitation

Figure 4: Exemplary third person view of the experimental setting.

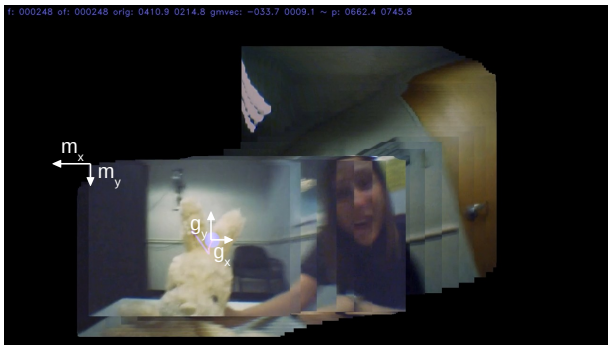

Figure 5: Schematic visualization of gaze and motion vector components.

of this approach is the limited search width. Thus, very fast head rotations cannot be captured. Another problem can be local object movement that covers most of the camera's viewing angle. In this case, object motion might be incorrectly registered as global motion.

Visualization To verify that the head motions were correctly estimated, we developed our own visualization and analysis tool which projects each video frame into an egocentric view using the estimated global motion information (see Figure 1). In contrast to the standard video deshaking approach, we specifically aimed to project a low-resolution frame onto a larger surface to maintain the frame's original resolution. Thus, the tool reads video frames and displaces them according to their global motion into a high-resolution frame. Rotations are ignored due to their relatively low accuracy. Furthermore, the tool integrates the gaze data into this view. With this tool, we verified if the head motions were correctly estimated. The accuracy is locally sufficient to render a scene as in Figure 1 which extends the camera's field of view by overlaying the current frame over the past renderings. Since the approach cannot measure the absolute head position, small errors accumulate over time. Therefore, this method is more useful to analyze head motion dynamics instead of absolute positions. A more detailed quantitative evaluation will be an important step to further develop the method but is omitted at this point.

Gaze and Head Motion Processing

In this part we use the head motion vectors ($\vec{m}$) that were extracted in the previous step and calculate the average head activity. The average gaze activity is calculated based on gaze vectors ($\vec{g}$) from the eye tracking data. Furthermore, an annotation is used to select frames that belong to different interaction conditions (see Figure 3). Both head motion and eye-tracking data is smoothed by fitting cubic splines to suppress high-frequency noise using the R software (R Development Core Team, 2011). The gaze coordinates are converted to relative gaze shift vectors by taking the first derivative before smoothing. To measure the development of infant's head motion activity we define a measurement that will be used to summarize each trial under the different annotated conditions. We use the Euclidean norm to calculate the magnitude of each motion vector $\vec{m}$ per frame at time step t (see Figure 5). The average head motion activity (HMA) per trial is the mean of these magnitudes:

$$\text{HMA} = \frac{1}{N}\sum_{t=1}^{N}\sqrt{m_x\left(t\right)^2 + m_y\left(t\right)^2} \tag{1}$$

Each infant's average gaze activity (GA) is measured analogously based on the gaze shift vector $\vec{g}$ (see Figure 5):

$$\text{GA} = \frac{1}{N}\sum_{t=1}^{N}\sqrt{g_x\left(t\right)^2 + g_y\left(t\right)^2} \tag{2}$$

If the number of frames matching a certain condition is below 100 the corresponding average activity is excluded from the analysis due to lack of sufficient samples.

Corpus

We used 21 sessions from data collected by Yoshida and Burling (2013). Infants from 6 to 24 months of age play with their mother with the goal of learning the names of objects (see Figure 4). The mother has access to a predefined set of objects where she uses toys of her choice based on the current word to be learned by the children, which is given by audio instruction. The mother is free to choose any of the objects at any point during the session, and is instructed to interact as naturally as possible. The child is wearing a Positive Science eye tracker, which records eye gaze and a video from the child's perspective. The method described in this paper operates on this video data. Furthermore, the corpus is annotated frame-by-frame providing a coding of the ongoing action. In this work, we focus on frames falling under the following conditions, which we define as social conditions: (1) the infant is looking at the mothers face; (2) the mother is holding a toy; (3) the mother is naming a toy.

Table 1: Regression analysis results for the relationship between head activity and age

Condition	R^2	F	p	
Parent Holding Toy	0.32	F(1,19)=8.92	0.01	**
Parent Not Holding Toy	0.01	F(1,19)=0.28	0.60	
Child Gazing at Parent	0.24	F(1,17)=5.28	0.03	*
Child Not Gazing at Parent	0.05	F(1,19)=1.01	0.33	
Parent Naming Toy	0.26	F(1,16)=5.72	0.03	*
Parent Not Naming Toys	0.01	F(1,19)=0.24	0.63	

Table 2: Regression analysis results for the relationship between gaze motion and age

Condition	R^2	F	p	
Parent Holding Toy	0.16	F(1,19)=3.57	0.07	
Parent Not Holding Toy	0.15	F(1,19)=3.45	0.08	
Child Gazing at Parent	0.14	F(1,17)=2.71	0.12	
Child Not Gazing at Parent	0.09	F(1,19)=1.90	0.18	
Parent Naming Toy	0.34	F(1,16)=8.25	0.01	*
Parent Not Naming Toys	0.05	F(1,19)=0.96	0.34	

For each condition the complementary non-matching condition is named non-social condition. The non-social conditions include frames where the mother is not interacting with the child (e.g., putting a toy away).

Regression Analysis Results

To test for a developmental trend in head activity, linear models for the infant's age were estimated for each social and non-social condition (see Table 1). The individual models are visualized in Figure 6. In this figure each chart displays the relationship of age and average head activity per subject on frames within a given condition. The gray area indicates a 95% confidence interval of the regression line. All three conditions show significant correlations suggesting that head activity plays an increasing role in social interactions as infants develop. For the conditions that children look at the mothers face and the mother is naming a toy the correlations were significant. A strongly significant correlation was found for the condition that infants look at the toy. A comparison of each model's R^2 value (see Figure 7) shows that given parents hold a toy, the relation between age and activity can be best explained by a linear model. We were unable to show a significant linear relationship between age and head activity for the non-social conditions. This suggests that this developmental trend depends on social interaction contexts, which supports our initial hypothesis.

To test for a developmental trend in gaze activity, linear models for the infant's age were estimated given each condition, including models for the non-social conditions (see Table 2). The individual models are visualized in Figure 8 along with the average gaze activity per subject. The only significant correlation can be found for the condition that the parent is naming a toy. The linear model for the condition that parents are *not* naming a toy was not significant. To further analyze the estimated linear models, we use the R^2 statistic to compare how much of the variance of the underlying data is expressed by each linear model. A large difference in the goodness of fit was found when comparing the R^2 values (see Figure 9), suggesting that this developmental trend depends on the social interaction condition. This difference is not visible for when the parent-holding-toy condition and the child-gazing-at-parent condition.

In general, several outliers are visible (see Figure 8). Although, a minimum number of 100 frames matching the condition is required for the mean activity to be included in each linear model these outliers are caused to a relatively low number of frames matching the condition. In the child is gazing at the parent condition, the larger gaze activity values can be caused by children only quickly looking to the mother and back.

In Figure 6, a difference in head activity, if parents are holding a toy, is visible for very young infants around 6 months of age. The head activity is lower compared to the condition where parents are not holding a toy. We interpret this is due to scaffolding that parents perform by holding the object directly in front of the child. This effect is not visible for gaze activity (see Figure 8).

Discussion

In the present work, we used developmental data concerning children's visual experiences. We proposed a new alternative method using video stabilization techniques to extract commonly missing information such as head movement from this data. New methods for studying development using head-mounted eye-tracking have been gradually emerging over time (Franchak, Kretch, Soska, & Adolph, 2011; Kretch & Adolph, 2014). Our method can provide additional detail and new ways of analyzing these data. An important use-case is the reanalysis of existing longitudinal studies where repeating the study is costly. Our visualization module can also contribute to the analysis of eye gaze, since it helps to determine gaze locations outside the currently recorded frame. Furthermore, our approach is flexible, since it does not depend on the availability of unoccluded third person perspective recordings. The current limitations are the estimation of rapid head rotations and visual field covering object movements without enough peripheral view. However, their practical impact is minor.

The regression analysis of head motions and eye gaze linked to social interaction conditions revealed significant developmental trends for both head activity and gaze. Head motions and eye gaze for non-social conditions did not exhibit any significant developmental trends. This supports our hypothesis that both gaze and head activity

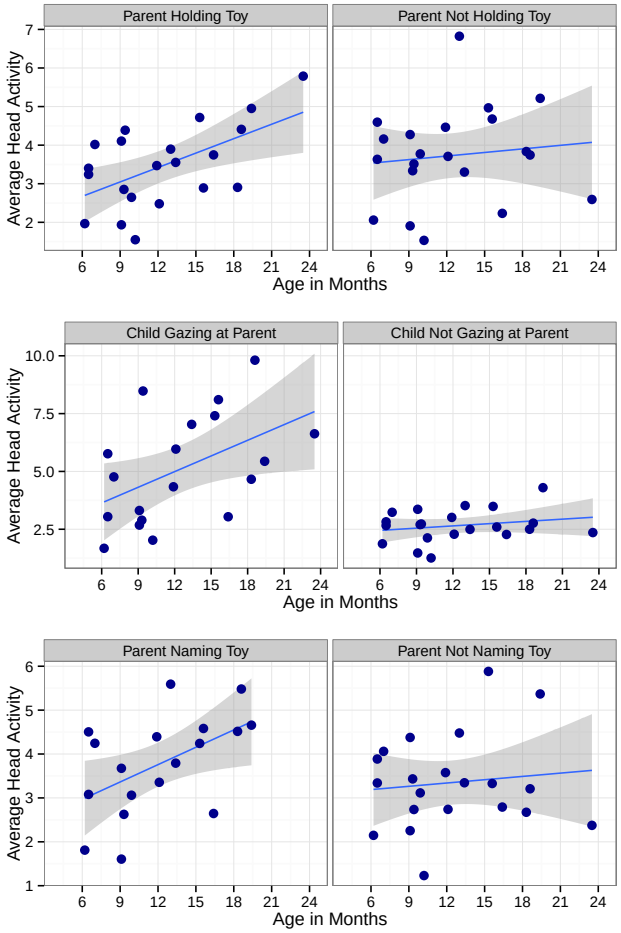

Figure 6: Relationship of age and average head activity per subject on frames falling under the given conditions. The gray area indicates a 95% confidence interval of the regression line.

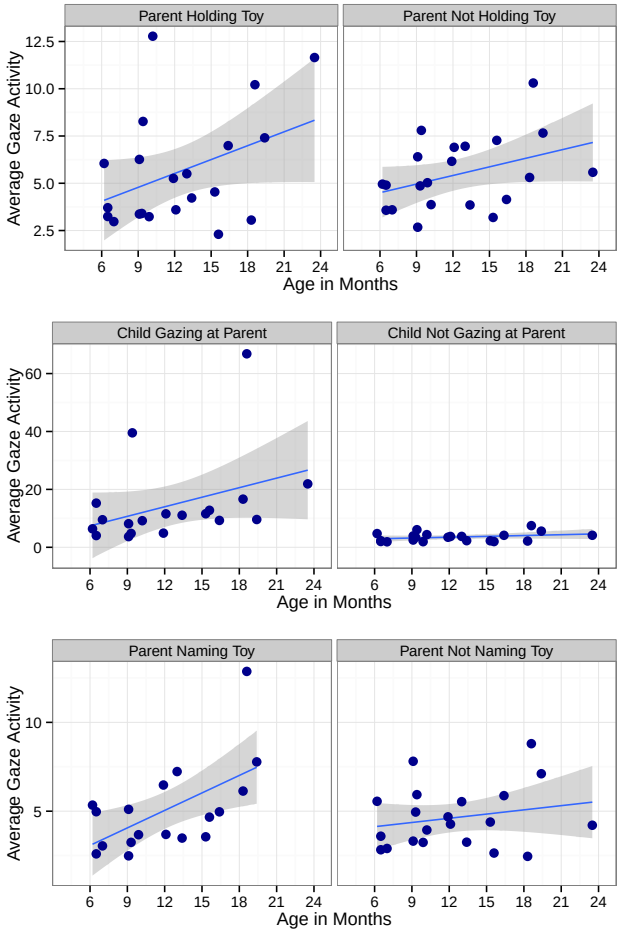

Figure 8: Relationship of age and average gaze activity per subject on frames falling under the given conditions. The gray area indicates a 95% confidence interval of the regression line.

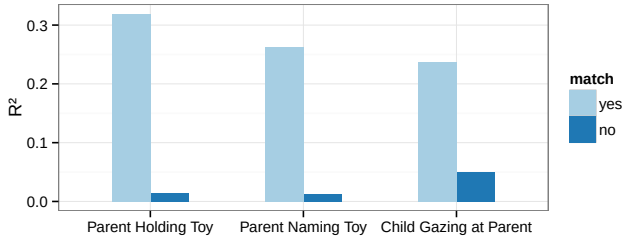

Figure 7: Comparison of R^2 for all head activity models. Match = yes / no indicates the result for frames matching / not matching the condition.

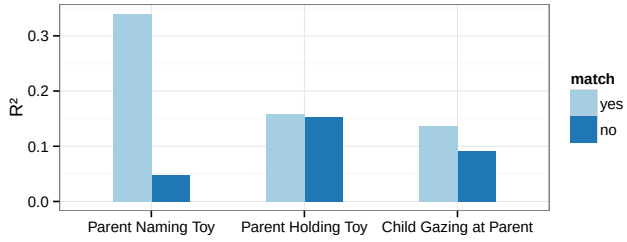

Figure 9: Comparison of R^2 for all gaze activity models. Match = yes/no indicates the result for frames matching / not matching the condition.

undergo developmental changes during social interactions. However, eye gaze activity only shows a significant and social context specific trend when the parent was naming an object, an event that is relatively short. In contrast, the most significant trend for head activity was found when parents hold an object, which is a comparably longer event. This suggests that eye gaze is coupled with social contexts when they require immediate attention. Head activity is required when children want to attend to a wider visual space. Parents might adapt to the children's increased motor capabilities and expand the space they use for their interplay. Thus, head activity reflects a property that is present for longer durations in social interactions. These differences in the development of head and gaze activity highlight that gaze analysis alone is incomplete in reflecting infants' developing response to social referencing contexts.

Furthermore, different attention shift patterns can be found in parent-child social interactions. For example, parents initiative in moving an object might create a bottom-up visual cue the child reacts to. In contrast, child's initiative might originate from intentional playing with toys and thus result in top-down attention shifts. Doshi and Trivedi (2012) showed that bottom-up visual cues result in different eye-head movement latencies compared to top-down initiated attention shifts. Thus, additional analysis of head movements has the potential to identify different types of social interaction patterns automatically. Gaze analysis alone might not be sufficient to analyze these patterns.

The present attempt not only indicates the potential early psychological significance of head motion, but it may also provide a new insight into how early head motion can offer systematic cues for caregivers. Parents tend to respond to infants' attention, gesture, and facial affect, in timely manner. Responses to these cues can foster early learning but we know little how parents do so. Thus, further developing and applying this new approach to the domain of development of social cognition could help to understand the underlying mechanism of parental responsiveness within parent-infant interactions.

Using our new analysis method, we were successfully able to extract head motion and to measure head activity. Based on the results we showed that head activity in addition to gaze activity robustly reflects important developmental trends that indicate possible links to social cognition.

Acknowledgments

This work has been supported in part by MEXT KAKENHI "Constructive Developmental Science" (24119003). The data used for the present analysis is part of a project founded by the National Institutes of Health grant (R01 HD058620), Foundation for Child Development: Young Scholars Program, and the University of Houston's Grants to Enhance and Advance Research (GEAR) program. We especially want to extend our gratitude to the families who participated in the original study.

References

Bertenthal, B., & von Hofsten, C. (1998, March). Eye, Head and Trunk Control: The Foundation for Manual Development. *Neuroscience & Biobehavioral Reviews*, *22*(4), 515–520.

Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D. (2012, July). Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines. *IEEE Transactions on Affective Computing*, *3*(3), 349–365.

de Lima-Alvarez, C. D., Tudella, E., van der Kamp, J., & Savelsbergh, G. J. P. (2014, January). Early development of head movements between birth and 4 months of age: a longitudinal study. *Journal of motor behavior*, *46*(6), 415–22.

Doshi, A., & Trivedi, M. M. (2012, January). Head and eye gaze dynamics during visual attention shifts in complex environments. *Journal of vision*, *12*(2).

FFmpeg. (2014). Retrieved from `ffmpeg.org`

Franchak, J. M., Kretch, K. S., Soska, K. C., & Adolph, K. E. (2011, January). Head-mounted eye tracking: a new method to describe infant looking. *Child development*, *82*(6), 1738–50.

Jonsson, B., & Von Hofsten, C. (2003). Infants' ability to track and reach for temporarily occluded objects. *Developmental Science*, *6*, 86–99.

Kretch, K. S., & Adolph, K. E. (2014). Active vision in passive locomotion: real-world free viewing in infants and adults. *Developmental Science*, 1–15.

Martius, G. (2014). *vid.stab - Video stabilization library.* Retrieved from `github.com/georgmartius/vid.stab`

R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing.* Vienna, Austria. Retrieved from `http://www.r-project.org`

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

Yoshida, H., & Burling, J. M. (2013). An Embodied Perspective of Early Language Exposure. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 1635–1640). Austin, TX: Cognitive Science Society.