

The Significance of Social Input, Early Motion Experiences, and Attentional Selection

Joseph M. Burling (jmburling@uh.edu)
and Hanako Yoshida (yoshida@uh.edu)
Department of Psychology, University of Houston
126 Heyne Bldg., Houston, TX 77204-5022 USA

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

Abstract—Before babies acquire an adult-like visual capacity, they participate in a social world as a human learning system which promotes social activities around them and in turn dramatically alters their own social participation. Visual input becomes more dynamic as they gain self-generated movement, and such movement has a potential role in learning. The present study specifically looks at the expected change in motion of the early visual input that infants are exposed to, and the corresponding attentional coordination within the specific context of parent-infant interactions. The results will be discussed in terms of the significance of social input for development.

I. INTRODUCTION

Babies are able to perceive and parse their visual environment and also able to move their eyes and head to select visual targets (objects or people) in space. Despite this seemingly primitive visual capacity, infants have the opportunity to continuously process complex visual input, and accumulate knowledge from the visual environment. Indeed, from day one, even without clear views of their scene, and well before they can walk or talk, babies actively contribute to their own learning experiences by observing the scenes available to them in the form of social interactions, and actively reciprocate as a social partner. Early emergence of any social intelligence can be found even at the earliest stage (e.g., facial recognition) and recent attempts with a baby robot simulating babies' visual constraints suggest how development (increase in visual acuity) optimizes learning, and how very early in development, limited visual capacity improves facial recognition [1]. New technological advancements further our understanding of visual input by taking a child's own perspective using head-mounted cameras and eye tracking devices, which also have begun to reveal a number of aspects regarding early attentional selection and its implication for learning [2]–[4]. These studies provide insight into how early visual input is systematic and constrained/supported by their own actions [3], and how such self-generated views has a direct impact on children learning words [5]. Furthermore, a previous study with 18-month-olds observed which social element is captured by a baby's own viewpoint, and demonstrated that early on, motion is generated most frequently under views containing hands. Thus, hands may help organize attentional resources.[3]. The most recent analysis of early motion (optical flow) experienced by babies provides supporting evidence that motion views of adult and child are similar when experiencing similar actions

[6]. Together, these results suggest that the child's selective attention is organized partly by their own actions [7], and that interestingly, these actions may generate unique patterns of motion in scenes from which attentional selection occurs. This raises the question of how their view selection may relate to actual moving scenes. Is attention selection similar across children due to the inherent characteristics of object selection (e.g., based on its saliency), yet changes through their development as a function of attentional development? Or, is the moment-to-moment selection of attention tightly linked to the moving scenes uniquely available to that child at a particular moment?

II. MOTHER-INFANT PLAY SESSIONS

One way to consider the relationship between early selective attention and social motion in children is to use a natural parent-child play environment to independently analyze the child's eye gaze behavior and quantify motion events presented in the different scenes. In the present study, we used the baby's perspective (via head-mounted eye-tracking device) during mother-infant play sessions (see Figure 1). From this perspective we obtained eye-tracking data for measuring selective attention and a first-person perspective for analyzing self-generated head motion. A wall-mounted camera was also used for capturing motion events of mother-infant interactions. As a first step toward understanding the potential similarities and differences between attentional selection (eye gaze) and scenes available to the child (motion), we studied the correspondences of data for two infants at their 6, 12, and 18 month play sessions, and where dramatic physical changes are also observed. Documentation of how attentional selection is tightly linked to the visual patterns presented to each child adds to the growing literature relating the significance of social interactions altering perception, and the role of actions in perception and learning.

III. METHODS FOR DETERMINING MOTION

Motion patterns generated by the social interactions between mother and child, and by the child's own view, were obtained by estimating optical flow using computer vision algorithms provided by the Open Source Computer Vision Library. Specifically, the Lucas-Kanade method with pyramids was implemented, which allows us to calculate the trajectory of motion at multiple points in space between subsequent

video frames. This approach was applied to both the third-person and child-centered videos, which does not involve additional sensors or attachments to the child—which may further restrict natural motion tendencies. The use of optical flow estimation allows us to observe the overall motion dynamics present among different perspectives and play sessions. Eye motion data was obtained using the Positive Science head-mounted system, and gaze direction was translated into x and y coordinates which can be mapped onto a 640 X 480 video frame recorded at 30 frames per second.

IV. CORRESPONDENCES BETWEEN MULTIPLE PERSPECTIVES

Preliminary results indicate similar developmental trajectories between the child-centered perspective and the third-person view. The range of motion extracted from the third-person view indicates an increase in dynamic interactions across multiple play sessions. It is possible that the social partner plays a major role during the early stages of development by generating and engaging in actions aimed specifically for the child. Increases in optical flow from this perspective during later stages of development are partly attributed to increased child interaction with the objects and the parent. Based on this, additional perspectives were considered to infer whether or not the child is receptive to the parent’s actions. Optical flow measurements taken from the child’s head-centered view indicate correspondences between motion generated by the parent and motion generated by movement of the child’s head. Head motion is most constrained during initial play sessions, in which less dynamic head turns are observed, and the flow of motion is generated by the parent’s actions centered primarily at the child’s own perspective. This is also expected given the proposed shift in social dynamics seen from the third-person view. More specifically, the optical flow measured from the child’s view taken at later stages of development indicates dynamic shifts in their perspective as the child becomes more actively engaged in determining their own optimal view. As a result, this also increases the complexity of the social dynamics and flow of motion between mother and child. However, the motion correspondences between third-person and first-person views may be less pronounced in terms of gaze patterns, which might be due to individual differences in attentional selection. Shifts in gaze, as measured by length of vector coordinates, frequency of shifts, and speed, seem to better explained by individual differences between children. This variation in attentional selection may be driven by internally motivated factors exhibited by the child, or by the unique style of parental instruction and interaction. Ongoing investigation into the differences in eye-gaze patterns among children and across development offers systematic documentation of linkages between moment-to-moment selective attention, self-generated head motion, and dynamic social interactions—all of which are relevant for developing children learning through social dynamics.

The present work is a first step toward investigating the potential role and relationship between self-generated motion,

organizing early attention, and parental scaffolding during complex visual scenes. These attempts at applying recent technologies leads to our understanding of real-time sensitivity and responses to social cues that emerge through complex bodily experiences during natural learning, such as in language learning, motor learning, and the development of social intelligence.

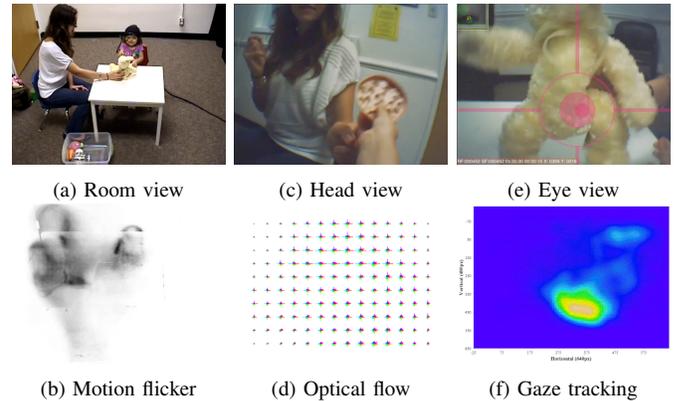


Fig. 1: *Left to right.* Views from three different perspectives. *Top row.* Video data taken at 18 months for a single child. *Bottom row.* [1b], Average motion flicker for the third-person view, [1d], Average optical flow of the child-centered view, [1f], gaze trajectory of a single child.

ACKNOWLEDGMENT

This research was supported by a National Institutes of Health grant (R01 HD058620), the Foundation for Child Development, and University of Houston’s Grants to Enhance and Advance Research (GEAR) program. We especially wish to thank the families who participated in this study. We thank the undergraduate research students in the Cognitive Development lab for their support in coding for the present study.

REFERENCES

- [1] Y. Nagai, “Joint attention development in infant-like robot based on head movement imitation,” *Proceedings of the International Symposium on Imitation in Animals and Artifacts*, 2005.
- [2] A. Pereira, L. Smith, and C Yu, “A bottom-up view of toddler word learning,” *Psychological Bulletin and Review*, no. 8, p. 8, under revision.
- [3] H. Yoshida and L. Smith, “What’s in view for toddlers? using a head camera to study visual experience,” *Infancy*, vol. 13, pp. 229–248, 2008.
- [4] C. Yu, L. Smith, H. Shen, A. Pereira, and T. Smith, “Active information selection: visual attention through the hands,” *IEEE Transactions on Autonomous Mental Development*, vol. 2, pp. 141–151, 2009.
- [5] C. Yu and L. Smith, “Embodied attention and word learning by toddlers,” *Cognition*, in press.
- [6] F. Raudies, R. O. Gilmore, K. S. Kretch, J. M. Franchak, and K. E. Adolph, “Understanding the development of motion processing by characterizing optic flow experienced by infants and their mothers,” *Developmental Science*, 2012.
- [7] H. Yoshida and J. M. Burling, “A new perspective on embodied social attention,” *Cognition, Brain, Behavior: An Interdisciplinary Journal*, vol. 15, pp. 535–552, 2011.