How do Infants Coordinate Head and Gaze?: Computational Analysis of Infant's First Person View in Social Interactions

Lars Schillingmann, Joseph M. Burling, Hanako Yoshida, Yukie Nagai

Infant eye gaze and head motion is critical to effective visual exploration and learning, and are highly correlated through the course of the child's physical development (Nakagawa and Sukigara, 2013). We hypothesize that this complex head-eye coordination might also be bootstrapped through social interaction.

Previous observations suggest that social input – e.g. parents handling objects – guides infants' eye gaze in a play context (Yoshida and Burling, 2014). In this work we analyzed the correlation between head motion and eye gaze of 6 and 18 month-olds in a naturalistic play to identify developmental changes. Our corpus consists of recorded social interactions of infant-adult dyads (Yoshida and Burling, 2013). The caretaker was instructed to freely play with the infant using a set of toys and to teach a set of words during a period of 5:30 Minutes. The infants' gaze and infants' field of view were recorded using a Positive Science eye tracker – consisting of head-mounted eye tracking and head view cameras. While the gaze coordinates are directly provided by the eye tracking system the head motion is not. This limitation offered an opportunity for us to develop the current method that neither requires manual frame-by-frame coding of the head position nor involves expensive external marker tracking systems that would require additional calibration and setup-time. Our method relies on open source video stabilization software (http://public.hronopik.de/vid.stab). A core functionality of video stabilization software is global motion estimation, i.e. given two consecutive video frames the software estimates the camera motion that occurred between these frames. In case of the aforementioned head-mounted camera, its motion corresponds to the head motion. We developed our own visualization and analysis tool which projects each video frame into an ego-centric view using the estimated global motion information (see Figure 1). The tool reads video frames and displaces them according to their global motion into a high resolution frame. Furthermore, it integrates the gaze data into this view. With this tool, we verified if the head motions were correctly estimated and performed further steps as for example noise filtering. The ego-centric view helps to see where the child is gazing at, even if the gaze coordinate does not lie within the current field of view of the head camera.

The head motion and gaze data were cross correlated for each participant. Figure 2 shows the correlation coefficient at the largest correlation peak on the y-axis over the infants' age. This result shows that the head-gaze coordination gets better as children get older. Previous results did not exhibit this trend. However they focused on a different age range (Nakagawa and Sukigara, 2013). The point color in Figure 2 indicates the lag between gaze and head motion where the maximal peak was observed. A trend in lag values is not observable yet but we speculate this might be caused by the parents scaffolding during early stage of development. With our tool we will analyze the data in more detail to identify the effects of social support on head and eye coordination.