Selective Looking to Audiovisual Redundancy: A New Method Demonstrates Fine-Grained, Acoustically-Driven Visual Exploration in Young Children

Lorraine E. Bahrick, Kasey C. Soska, & James Torrence Todd
Florida International University

NIH R01-HD053776 &
K02-HD064943

bahrick@fiu.edu

Cognitive Development Society
October 2015, Columbus, OH

INTRODUCTION

Detection of intersensory redundancy—stimulation that is temporally synchronized across the sense modalities—guides attention to unitary multimodal events (e.g., sights and sounds of a person speaking), and in turn, provides a meaningful basis for perception, learning, and memory (Bahrick & Lickliter, 2014).

We developed the Intersensory Processing Efficiency Protocol (IPEP), the first fine-grained, individual difference measure of the efficiency (speed and accuracy) of intersensory processing, useable across the lifespan. Participants must locate an acoustically-synchronized target event amongst five competing visual events—simulating the "noisiness" of the natural environment.

Accurate intersensory processing on the IPEP should be evident in frequent selection and longer looking (audiovisual matching) of the sound-synchronized target event (*macro measures*, collapsed across multiple fixations) as well as in visual exploration strategies (spatial and temporal *micro patterns* of fixations).

Participants could show frequent *selection* of the target event in micro patterns by (a) quickly homing in on the sound-synchronous target on each trial and ignoring the competing events, or by (b) briefly looking to a few competing events before consistently fixating the sound-synchronous target each trial. Once they found the target event on each trial, they could show audiovisual *matching* for the target event by (a) continuing to fixate on it to the exclusion of the competing events, or by (b) visually exploring competing events but frequently returning to fixate the target event.

In the present study, we assess intersensory accuracy using macro measures on the IPEP. We then use micro measures to characterize the strategies children use to demonstrate selective looking to audiovisual redundancy by examining the spatial distribution and temporal sequencing of fixations during synchronous audiovisual and silent visual stimulation.

METHOD

We tested 64 3.5- to 5-year-old children. Visual fixations were recorded using a Tobii X120 eye-tracker after a 5-point calibration. Children saw either social events (six women reciting different stories; *N*=32) or nonsocial events (six objects striking a surface in varied patterns; *N*=32).

On each 6s trial, children viewed a 2x3 grid (*Figure 1*) consisting of one target event and five competing events (6 AOIs total)—across one block of 12 synchronous audiovisual trials and one block of 12 silent visual control trials (block order counterbalanced). On audiovisual trials the natural soundtrack was synchronized with the target event; on silent visual trials no soundtrack was played.

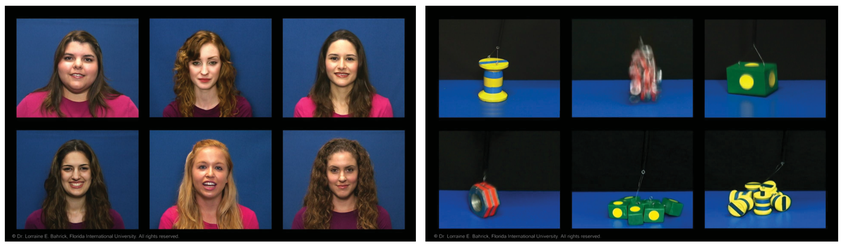

Figure 1. Static images of the dynamic social and nonsocial events in the IPEP.

Useable eye-tracking data across all trials was *M*=82.3% (range=42-99%). Raw gaze data were processed using the I-VT filter (Olson, 2012).

RESULTS

MACRO MEASURES OF INTERSENSORY ACCURACY

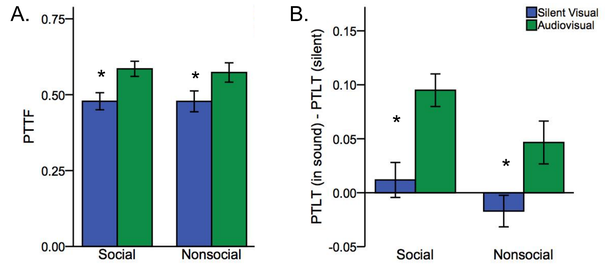

Figure 2. (a) Accuracy in selection and (b) accuracy in matching for social and nonsocial events in audiovisual and silent visual stimulation.

Accuracy in selection is the proportion of total trials on which the target was fixated (PTTF). *Accuracy in matching* is the proportion of total looking time (PTLT) to the target event when it was in sound versus when the same event was not in sound.

Children demonstrated intersensory accuracy according to macro measures of looking time. As shown in *Figure 2*, on the audiovisual compared to the silent visual block children fixated the target on a greater proportion of trials (greater PTTF: *Figure 2A*), $F(1, 60)=17.5$, $p<.001$, and gazed to it longer (greater PTLT: *Figure 2B*), $F(1, 60)=18.7$, $p<.001$. (These results also held when controlling for the number of AOIs fixated.) Results did not differ between social and nonsocial events.

MICRO PATTERNS OF VISUAL EXPLORATION CHARACTERIZING INTERSENSORY ACCURACY

The duration and variance in children's fixations did not differ between audiovisual and silent visual trial blocks nor did it differ between target and competing events. However, the strategies children used did differ between audiovisual and silent visual stimulation. These micro patterns were consistent across social and nonsocial events.

Spatial micro patterns. Children distributed their fixations more broadly—they fixated more AOIs—on audiovisual (*M*=3.13, *SD*=0.71) than silent visual trials (*M*=2.88, *SD*=0.73), $F(1, 62)=7.60$, $p<.01$. *Figure 3* shows exemplar spatial patterns from one child.

Temporal micro patterns. Children returned their gaze to the target event more often (they fixated the target on a greater proportion of total fixations on each trial) in audiovisual (*M*=19.1%, *SD*=7.3) than visual stimulation (*M*=13.7%, *SD*=5.8), $F(1, 62)=19.3$, $p<.001$. *Figure 4* shows exemplar temporal patterns from one child.

A. Audiovisual (top center is target) B. Silent Visual (top center is "target")

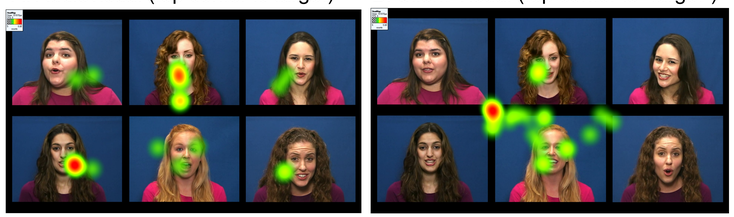

Figure 3. Spatial distribution heat-map of individual fixations across the 6 AOIs from (a) an exemplar audiovisual trial and (b) an exemplar silent visual trial. Red indicates more fixations.

A. Audiovisual (top left is target) B. Silent Visual (top left is "target")

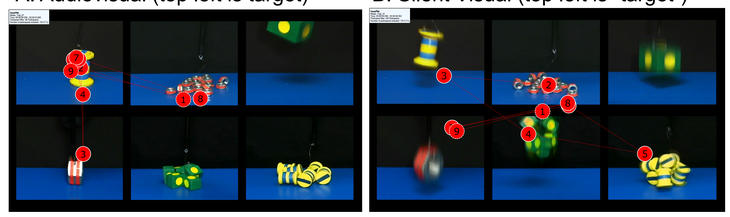

Figure 4. Temporal patterning gaze-plot of individual fixations to the target AOI amidst the five competing AOIs from (a) an exemplar audiovisual trial and (b) an exemplar silent visual trial. Numbers reflect the temporal order of fixations.

DISCUSSION

The present results reveal intersensory processing at the macro level (aggregated across fixations), as well as at the micro level in the real-time exploratory strategies. Exploratory strategies entailed widening spatial attention (fixating more areas) and returning gaze more frequently to the target. These data elucidate the real-time attentional capture of redundant audiovisual events.

The patterns of behavior in real time may reveal more about developmental origins and individual differences than aggregate data alone. The IPEP provides a novel method for characterizing fine-grained multimodal attention across different time frames. Developmental change (both typical and atypical) must play out across moment-to-moment behavior. Our micro measures are capable of characterizing the exploratory strategies that predict optimal development of social, cognitive, and language skills.