Assessing Individual Differences in the Speed and Accuracy of Intersensory Processing in Young Children: The Intersensory Processing Efficiency Protocol

Lorraine E. Bahrick, Kasey C. Soska, James Torrence Todd

Florida International University

Author Note

This research was supported by National Institute of Child Health and Human Development Grants R01 HD05776 and K02 HD064943 to Lorraine E. Bahrick. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Health and Human Development or the National Institutes of Health. Portions of this work were presented at the 2013 meeting of the Society for Research in Child Development and the 2015 meeting of the Cognitive Development Society.

Correspondence concerning this article should be addressed to Lorraine E. Bahrick, Department of Psychology, Florida International University, bahrick@fiu.edu.

Abstract

Detecting intersensory redundancy guides cognitive, social, and language development. Yet, researchers lack fine-grained, individual difference measures needed for studying how early intersensory skills lead to later outcomes. The Intersensory Processing Efficiency Protocol (IPEP) addresses this need. Across a number of brief trials, participants must find a sound-synchronized visual target event (social, nonsocial) amidst five visual distractor events, simulating the "noisiness" of natural environments. Sixty-four 3- to 5-year-old children were tested using remote eye-tracking. Children showed intersensory processing by attending to the sound-synchronous event more frequently and longer than in a silent visual control, and more frequently than expected by chance. The IPEP provides a fine-grained, nonverbal method for characterizing individual differences in intersensory processing appropriate for infants and children.

Keywords: intersensory processing, audio-visual matching, eye-tracking, individual differences

Assessing Individual Differences in the Speed and Accuracy of Intersensory Processing in Young Children: The Intersensory Processing Efficiency Protocol

The world provides a richly structured dynamic flow of stimulation to all of the senses—far more than can be attended at any time. One of the biggest challenges for young perceivers is to make sense of this stimulation and determine which sights, sounds, and tactile impressions belong together and which do not. How do young children locate the speaker in a crowd or the object that is the source of a sound? Research on the development of intersensory processing, the ability to coordinate stimulation across the senses, has demonstrated that this fundamental skill develops in infancy and relies on detecting amodal information: information "redundant" or common across stimulation to different sensory systems, such as temporal synchrony, rhythm, tempo, and intensity changes (Bahrick & Lickliter, 2002, 2014; Gibson, 1969; Walker-Andrews, 1997). Selective attention to *intersensory redundancy*, stimulation temporally synchronized across the senses (e.g., common onset, offset, and temporal patterning) guides attention to unitary multimodal events (Bahrick & Lickliter, 2000; Lewkowicz, 2010). In turn, unitization of intersensory information provides a meaningful basis for perception, learning, and memory (Bahrick & Lickliter, 2012). These skills emerge in infancy and are refined across childhood (Bahrick, 2010; Bahrick & Lickliter, 2002; Kaganovich, 2016; Lewkowicz, 2000, 2014).

In this paper, we lay the foundation for a novel method for measuring individual differences in intersensory processing skills, capable of revealing new information about these skills in infants, children, and adults. This method, the Intersensory Processing Efficiency Protocol (IPEP), provides the first fine-grained assessment of selective attention and intersensory matching of naturalistic audible and visual stimulation, which does not require verbal instructions or responses. The IPEP indexes *intersensory processing efficiency*, the speed and

accuracy of locating the visual source of auditory stimulation in social and nonsocial events. Similar to measures designed for verbal participants (Foss-Feig et al., 2010; Ross et al., 2011; Woynaroski et al., 2013), this protocol assesses performance across a number of relatively brief trials and provides multiple indices characterizing individual differences in intersensory processing skills. The IPEP leverages familiar nonverbal measures of attention, including duration of looking time and frequency and latency to fixate a target, and averages them across multiple trials in a previously unstudied context (multiple concurrent dynamic audiovisual events) to provide a unique characterization of individual differences in accuracy and speed of processing naturalistic audiovisual events. We present evidence that the IPEP is suitable for young children, that children show evidence of intersensory processing in this challenging task, and that they show individual differences in intersensory skills as well as meaningful inter-correlations among measures. This manuscript serves both as a description of a new methodology and presents new findings and insights into the relations among intersensory speed and accuracy at both the group level and the individual level in young children.

The Importance of an Individual Difference Approach to the Study of Intersensory Processing

Intersensory processing serves as a critical foundation upon which more complex social, cognitive, and language skills can develop. Rapidly shifting attention to locate the source of a sound allows children to unitize the sights and sounds of speech or object events, to pick out the speaker in a crowd, or attend to the object that is labeled. Focused attention, in turn, provides a basis for further processing these multimodal events. Bahrick and colleagues have proposed that individual differences in the speed and accuracy of attention to intersensory redundancy should predict social, cognitive, and language outcomes (Bahrick, 2010; Bahrick & Lickliter, 2002;

Bahrick & Todd, 2012). For example, intersensory processing has been proposed to underlie word mapping (Gogate & Hollich, 2010). Studies at the group level support this view. They reveal that synchronous, but not asynchronous, object movement and verbal labeling promotes object-label mapping (Gogate & Bahrick, 1998; Jesse & Johnson, 2016), and this provides a gateway for further processing of object-label relations (Gogate, 2010; Gogate & Maganti, 2016). Moreover, individual children show improved word learning if parents more often spontaneously synchronize object movement and labeling (Nomikou, Koke, & Rohlfing, 2017), and if children focus attention more often on the visual location of a verbally labeled object (Pereira, Smith, & Yu, 2014; Samuelson, Smith, Perry, & Spencer, 2011; Yu & Smith, 2012) and also switch attention between the object and the parent (Gogate, Bolzani, & Betancourt, 2006). These findings suggest that intersensory processing skills may promote a variety of downstream developmental improvements.

Individual differences in the accuracy and speed of intersensory skills, such as face-voice or object-sound matching, should predict individual performance in domains that rely on this foundation—from vocabulary growth, literacy skills, to social competence. However, to date, there are no commonly accepted measures of intersensory processing that are sufficiently fine-grained nor designed for assessing individual differences in infants or children. Without fine-grained individual differences measures, it has not been possible to determine if one child shows better intersensory processing skills than another, how these skills change across development, nor identify the pathways from these skills to later developmental outcomes.

In other areas of research, the creation of individual difference protocols utilizing traditional looking time measures has led to significant advances in our understanding of developmental processes and pathways. For example, Fernald and colleagues (2008) developed

the “Looking While Listening” procedure to assess individual differences in language processing efficiency (speed and accuracy of word recognition) in toddlers. In this method, static images of two familiar objects are shown side by side along with a verbal label for one. Faster speed and greater accuracy of word recognition at 18 months predicts vocabulary growth trajectories as well as cognitive skills years later (Fernald & Marchman, 2012; Marchman, Adams, Loi, Fernald, & Feldman, 2015). Similarly, more efficient visual attention (e.g., longer looks, faster encoding) and visual recognition memory in infancy predict cognitive skills at 2-3 years of age, which, in turn, predict later IQ (Rose, Feldman, & Jankowski, 2012a; Rose, Feldman, Jankowski, & Van Rossem, 2012b). Studies such as these highlight the value of individual difference measures—and the assessment of both speed and accuracy—allowing researchers to relate individual performance on basic, early-emerging skills with longitudinal changes in more complex, later-developing skills. Individual difference measures of intersensory processing speed and accuracy promise to yield similar benefits for the study of intersensory perception and its role in language, cognitive, and social development.

An individual differences approach can reveal typical developmental trajectories of intersensory processing skills in infants and children, pathways between these basic skills and later developmental outcomes, and in turn, help identify performance that is atypical and outside the normal range of variability. For example, children with Autism Spectrum Disorders (ASD), who show impairments in social and language functioning, also show impairments in intersensory processing (for reviews see Bahrick & Todd, 2012; Stevenson et al., 2016). Compared to typically-developing children, children with ASD fail to show a preference for synchronous over asynchronous audiovisual speech (Bebko, Weiss, Demark, & Gomez, 2006; Grossman, Schneps, & Tager Flusberg, 2009; Patten, Watson, & Baranek, 2014), have an

enlarged temporal binding window for integrating visual and auditory speech (Stevenson et al., 2014; Woynaroski et al., 2013), and are less accurate in audiovisual speech in noise tasks (Foxe et al., 2015; E. G. Smith & Bennetto, 2007). Early disturbances of intersensory processing—especially of social events, which provide extraordinary amounts of intersensory redundancy across faces and voices —could induce a cascade leading to poor integration of faces and voices, piecemeal processing of multimodal events, and delayed social, cognitive, and language development (Bahrick, 2010; Bahrick & Todd, 2012; Patten, Labban, Casenhiser, & Cotton, 2016; Stevenson et al., 2017). Elucidating the potential cascades stemming from poor early intersensory processing requires fine-grained measures of individual differences in intersensory processing and systematically characterizing the typical development of these skills.

Group Level Approaches to the Study of Intersensory Processing: Findings and Limitations

The traditional approach for assessing intersensory functioning in nonverbal participants is a group-level approach—the intermodal preference method (Bahrick, 1983, 1988; Lewkowicz, 1992; Spelke, 1976). Infants view two side-by-side films, one synchronized with a soundtrack and the other out of synchrony. Usually a small number of trials (typically 2-4 within a condition), with trial times ranging between 20-120 s, are presented and the lateral position of the sound-synchronous film is varied across trials (Lewkowicz, 1992; Montague & Walker-Andrews, 2002; Patterson & Werker, 1999; though for a larger number of trials see Bahrick, 1988; Soken and Pick, 1992; Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991). A single dependent measure—the proportion of total looking time to the sound-synchronous event—is derived and is averaged across participants and taken as evidence of intersensory processing for the group. Another group-level approach has used the infant-controlled habituation procedure to

determine if infants detect a change in an audiovisual relation from habituation trials to test trials (e.g., Bahrick, 1994; Gogate & Bahrick, 1998).

These traditional procedures have revealed a great deal about intersensory processing at the group level. Even young infants show excellent intersensory skills. For example, newborns detect face-voice synchrony (Lewkowicz, Leo, & Simion, 2010), and infants of 2- to 4-months can match auditory vowel sounds with visual lip movements (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999, 2003). Further, infants of 4- to 5-months perceive audiovisual information for emotion (A. J. Caron, Caron, & MacLean, 1988; Flom & Bahrick, 2007; Walker, 1982) and the audiovisual rhythm and tempo of objects (Bahrick & Lickliter, 2000; Bahrick, Flom, & Lickliter, 2002; Lewkowicz, 1992). When learning novel object-label pairings, 8-month-olds differentiate the kinds of object motion patterns caregivers produce (Matatyaho-Bullaro, Gogate, Mason, Cadavid, & Abdel-Mottaleb, 2014), and 16-month-olds differentiate different vocal stress patterns (Curtin, Campbell, & Hufnagle, 2012). Although standard procedures can reveal intersensory skills in groups of children under different conditions, they are not designed to assess differences across individuals within a group. They provide only a course grain of analysis because an infant's score is typically derived from only a few trials (rather than averaged across a larger number of trials, providing a more stable mean and a measurable error variance for individual participants) and from a single dependent measure. Further, their psychometric properties are not known and, thus, they are not well suited for assessing individual differences in intersensory functioning.

Remote eye-tracking studies are capable of providing a fine-grained analysis of looking patterns to multimodal events. However, they have not yet been developed as an individual difference measure and typically rely on only a single dependent measure (duration of looking to

a region of interest). Group-level studies reveal that infants look longer to the mouth (the source of audiovisual redundancy) compared to other parts of the face while a woman is speaking (Tenenbaum, Shah, Sobel, Malle, & Morgan, 2012), when infants are first learning their native language (Lewkowicz & Hansen-Tift, 2012), and when there is a mismatch between visually- and acoustically-specified syllables (Tomalski et al., 2013). This methodology holds promise for future development as an index of individual differences in intersensory processing.

Several studies have attempted to use traditional group-level methods for predicting developmental outcomes. One study (Montague & Walker-Andrews, 2002) relating intersensory matching of face-voice emotion with infant experience interacting with different caretakers had mixed success (finding relations for mother but not father or stranger). Some have grouped children into high and low intersensory performance groups (Altvater-Mackensen & Grossmann, 2015; Altvater-Mackensen, Mani, & Grossmann, 2016; Eppler, 1995). A few have found relations between performance of individual children on nonverbal intersensory tasks and language outcomes or Autism symptom severity (Alvatar-Makensen & Grossman, 2015; Alvater-Mackensen, et al., 2016; Patten, Watson, & Baranek, 2014) while others found no relations (Bebko et al., 2006). Fine-grained individual difference measures for intersensory skills have been successfully used with older children and adults (Ross et al., 2011; Woynaroski et al., 2013), however, they rely on language and thus are not appropriate for preverbal children or those with language impairments. There is, therefore, a critical need for a fine-grained, nonverbal individual difference measure of intersensory processing useable across age from infancy—the period during which intersensory skills develop most rapidly—through childhood and beyond.

The Intersensory Processing Efficiency Protocol: A New Approach

The Intersensory Processing Efficiency Protocol (IPEP) leverages traditional looking time measures to derive indices of speed and accuracy (using remote eye-tracking) in a context of multiple, concurrent events, both social and nonsocial. In the IPEP, participants must locate an acoustically-synchronized target event amidst five competing visual distractors (see Figure 1). The display is comprised of six concurrent events (rather than two), and relative to traditional measures, it averages a larger number of shorter trials to create a more fine-grained and potentially more sensitive protocol than currently available. The IPEP thus allows assessment of individual differences in intersensory processing as well as aggregate scores for groups of children. It can also provide rich detail about individual scanning patterns through eye-tracking.

Individual difference approach. By deriving indices from multiple trials (24, with 12 trials per condition in the present protocol; 48, with 24 trials per condition in our refined audiovisual version; see Discussion for details), rather than only a few trials, the IPEP can generate potentially more fine-grained and stable measures based on mean performance and variability for a single participant. This permits a more sensitive assessment of an individual's intersensory processing abilities, complementing the sensitivity of measures of individual differences in social, cognitive, and language outcomes.

Multiple measures. The IPEP assesses three indices of intersensory processing reflecting both speed and accuracy: 1) the *accuracy in selecting* the audiovisual-synchronous event, represented as the frequency of trials on which participants look to the target event amidst the five visual distractors, 2) the *accuracy in matching* the audiovisual-synchronous event, represented as the proportion of looking time to the target event, and 3) the *speed* in selecting the audiovisual-synchronous event, represented as the latency to look to the target event. For each participant, mean performance and variability is calculated for each measure across multiple

trials and target locations. Although looking time measures such as these are common across various methods testing non-verbal participants, they have not previously been used together to assess intersensory skills. These three complementary measures reflect critical components of attention to intersensory redundancy. Rapidly finding the source of intersensory information leaves more time to process the event. Frequently finding and selecting sound-synchronous events leads to more frequent sampling of amodal information, and attending longer to each intersensory event allows deeper processing of the multimodal event.

Simulating the complexity of the multimodal environment. The IPEP indexes intersensory processing in the context of multiple competing, naturalistic events—providing a meaningful basis for generalizing intersensory skills to natural, multimodal learning contexts. In traditional methods only one or two events are shown together, often with simple repetitive sounds, limiting their relevance to complex, real-world learning situations. In the IPEP, participants see six concurrent, dynamic events (see Figure 1), while hearing the synchronous and appropriate soundtrack to one of them, simulating the "noisiness" of the natural world of overlapping events. The audiovisual events are rich and varied, depicting women speaking fluid, child-directed speech and objects of various shapes and compositions striking a surface in varied temporal patterns. The events provide both macro-synchrony (onset and offset of head and large lip movements or object impacts against a surface) and micro-synchrony (specific speech sounds and fine-grained lip movements, or fine-grained temporal structure of object impacts). The protocol resembles the task of picking out a speaker in a crowd or the sounding object among a group of dynamic events.

A nonverbal measure. Finally, the IPEP does not require verbal responses or understanding language, and thus can be administered at any age across the lifespan. Participants

view the displays while their eye gaze patterns are recorded with a remote eye-tracker. The IPEP allows individual differences in the speed and accuracy of intersensory processing to be compared on the same metric across development.

Current Study

The primary goal of the current study was to demonstrate the feasibility of the IPEP as a novel method for assessing individual differences in intersensory processing skills in young children. We tested 3- to 5-year-old children, since children of this age are adept at intersensory matching and reliably detect sound-synchronous events in simple, more traditional methods (Bebko et al., 2006; Lewkowicz & Flom, 2014). However, it was not known if they could show intersensory matching in this more difficult task with 6 competing events. Further, developmental disorders such as autism are commonly diagnosed around this age. We included both social (women speaking) and nonsocial (objects impacting a surface) events because they serve as a foundation for language, cognitive, and social development and children with autism show selective impairments in perception of social events, but relatively spared performance with nonsocial events (Dawson, Meltzoff, Osterling, Rinaldi, & Brown, 1998; Patten et al., 2016; Swettenham et al., 1998). Pre-term infants may demonstrate a similar pattern of impairments in intersensory matching for social but not nonsocial events (Gogate, Maganti, & Perenyi, 2014; Pickens et al., 1994; Provasi, Lemoine-Lardennois, Orriols, & Morange-Majoux, 2017).

Another goal was to reveal new information about children's intersensory processing skills. Little research has assessed speed of intersensory processing, nor characterized differences in scanning patterns and strategies used by children in attending to multisensory events. It is not known how measures of intersensory speed and accuracy relate, nor what strategies children use to selectively attend to synchronous audiovisual events in the midst of other concurrent visual

events. Do children show longer look durations or more frequent looks to the sound-synchronous event? Do they show different scanning patterns to audiovisual than silent dynamic events?

We evaluated evidence for intersensory processing through two complementary avenues of analyses: 1) Audiovisual stimulation versus within-participant visual control: By controlling for differences in overall amount and spread of attention across conditions, we ensured that differences in gaze between conditions would uniquely reflect acoustically-driven visual exploration (i.e., intersensory processing). 2) Audiovisual stimulation alone: At the *group level*, accuracy of selecting and attending to each target event relative to attention for that same event when it served as a distractor—provided a within-condition control for chance looking. At the *individual level*, we characterized individual differences in speed and accuracy of intersensory processing and assessed the relations among these measures. Convergent findings across the two analytic approaches provide evidence for the IPEP as a viable index of intersensory processing efficiency. They also provide a basis for us to refine the protocol in future studies (see Discussion) to optimize efficiency and maximize the amount of data collected from a child.

Method

Participants

We tested 64 children (37 boys, 27 girls) between 3 and 5 years of age (M = 45.72 months, SD = 3.26). An additional 10 children were tested but their data were excluded because of fussiness (n = 3), inattentiveness to the display (n = 2), experimenter error (n = 3), and useable eye-tracking data from either eye less than 33.3% of the time (n = 2). Assuming β of .80 and a two-tailed p value of .05, a sample size of approximately N = 46 is needed for detecting a Cohen-defined medium effect size of d = .50 and r = .40. Thus, our sample of N = 64 children is sufficient to detect these effects.

Participants were recruited through county birth records and parents were contacted via publicly available phone records. Participants were from suburban and urban areas in Miami-Dade County and were predominantly middle class. Eighty-seven percent of the children were of Hispanic ethnicity (the remainder were non-Hispanic). Ninety-one percent of the sample were Caucasian-American, 3.1% were African-American, and 4.7% were Asian-American. Fifty-nine percent of children spoke predominantly English and 40.8% spoke predominantly Spanish. (Some parents did not disclose demographic and exact birth date information about their children as reported in the Supporting Information.) Families received a small monetary gift, a certificate of appreciation, and a small toy to thank them for their participation. The research protocol ("Development of Intermodal Perception of Social Events: Infancy to Childhood"; protocol number 051900-04) was approved by the Institutional Review Board at Florida International University.

Stimulus Events

The stimulus display consisted of a computer-generated 2 (rows) × 3 (columns) grid of six dynamic events (see Figure 1). The entire grid was displayed at 1920 · 1080 pixels and covered 104.1 cm (73.2° visual angle). Each square of the grid contained a single dynamic event, was 31.3 · 24.1 cm (25.2° visual angle) and was separated by 2 cm of black space horizontally and 3.2 cm vertically. Stimuli were delivered to a widescreen monitor using Tobii Studio on a Mac Pro computer with a 3.33 GHz processor, 16 GB of RAM, and a 400 MHz graphics card.

The social events consisted of six women speaking (Figure 1A) in child-directed speech–each woman recited a different story. The nonsocial events consisted of six objects on strings striking a blue wooden surface (Figure 1B) in varied, erratic temporal patterns (M = 50.5 impacts/min, SD = 9.9). The stimulus events were each recorded separately using a digital video

camera and unidirectional microphone. Actresses were instructed to match an exemplar to ensure that prosody, quality of child-directed speech, and range of facial expressions were similar across all actresses. The events lasted 20-30 s each and two 6-s long clips were selected from different portions of each event to serve as stimulus events (one clip in each trial block of the procedure) to produce distinct visual events for each block (woman reciting different parts of the story, object impacts depicting different temporal patterns). The videos were arranged in a 2 × 3 grid of six social events and six nonsocial events using Adobe Premiere software. Example videos of the audiovisual events are on Databrary (https://nyu.databrary.org/volume/336).

Procedure

Children sat in a chair (alone or on their parents' laps) 70 cm away from a 119.4-cm widescreen monitor. The chair's height was adjustable so that each participant could be positioned with the display midpoint at eye level. Children were positioned 60 cm away from a Tobii X120 eye-tracker that was angled upward toward the eyes at 20°. A video camera positioned above the monitor captured the participants' faces, so an experimenter could ensure the child was optimally positioned for comfortable viewing.

To calibrate the infrared corneal reflection-to-pupil tracking system for each participant, we used Tobii Studio's "Infant" calibration procedure: An experimenter presented animated objects with an accompanying attention-getting sound at five locations (top left, top right, center, bottom left, bottom right) on the monitor. This was repeated as often as needed at specific locations to obtain minimal deviation from the center of each calibration point. Calibration accuracy for the Tobii X120 is within 2°.

Participants received 24 6-s trials from the IPEP. Half the participants (n = 32) received the social events (women speaking) and half (n = 32) received the nonsocial events (objects

striking a surface). Each participant experienced 2 consecutive blocks of 6 *audiovisual test* trials (12 trials total) and 2 consecutive blocks of 6 *silent, visual control* trials (12 trials total), in a counterbalanced order across the sample. Between each 6-s trial, a 2-s looming and receding smiley face was presented to recapture children's attention to the center of the screen. Six different faces, each of a different primary color, were used in pseudo-random order across trials. For each participant, the audiovisual and visual control trial blocks were identical, except there was no sound during the control blocks. On each audiovisual test trial, the natural synchronous soundtrack from one target event was played and it was unsystematically asynchronous with the visual stimulation from the other five distractor events. The soundtrack was played at M = 58.4 dB (SD = 5.6) from two lateral speakers equidistant to the monitor midpoint. On each silent, visual control trial, the event designated as the "target" was yoked to the corresponding trial in the audiovisual block, making the trial blocks identical in every way except for the soundtrack.

Within each block of 6 trials, every event served as the sound-synchronous target on one trial and as a silent distractor on the other five trials. The order of which event was the sound-synchronous target on each trial changed from one block to the next (e.g. in the first block, the top middle event was the target on the first trial, but in the second block, the top middle was the target on the fourth trial) and was counterbalanced across the sample. Between the first and second blocks of 6 trials in each condition, a new clip from the same event was presented: Each woman recited a different part of the story and each object struck the surface in a new, unpredictable rhythmic pattern.

Eye-Tracking and Data Processing

The Tobii X120 system sampled eye gaze at 120Hz. Samples with useable gaze data during trial times averaged 82.3% (SD = 13.4). Trials with no useable data (either inattention to

the screen or missing data) were removed from analyses. This resulted in the loss of 0-3 trials per participant—producing a total of 21-24 trials each. From the raw gaze data, we derived fixations using an I-VT filter (Olsen, 2012) with a moving averaged, 20-ms window of gaze data used to calculate velocity, minimum fixations defined as sequential gaze data of greater than or equal to 50 ms, and a velocity threshold of 30° per s delineating fixations and saccades.

Areas of interest (AOIs) were defined as the grid square encompassing each event (i.e., each woman and the blue background; each object, surface, and the black background) and were 594 pixels wide and 453 pixels tall within a 1920 × 1080 pixel resolution. AOI parsing based on the media pixel space was done in SPSS 20 through syntax. The length of each fixation and whether it fell within each AOI or off-AOI was derived from the filtered fixation data and matched to the target and distractor locations based on the target AOI on each trial.

Three dependent measures within the audiovisual condition and the silent, visual control conditions were derived from the processed eye-tracking data. Accuracy in selecting the target across trials was calculated as the *proportion of trials on which the target was fixated* (PTTF): the number of trials on which the target event (or yoked "target" event) was fixated (i.e., a fixation of at least 50 ms that landed in the target event's AOI) divided by the total number of trials with useable gaze data. Accuracy in matching the target on each trial was calculated by totaling the duration of all fixations in the AOI of the target event (or yoked "target" event) and dividing that value by the total duration of all fixations in all AOIs. This produced a metric of the *proportion of total looking time* (PTLT) to the AOI of the target event on each trial. PTLT on each trial was then averaged across all trials within each condition (resulting in a measure similar to standard two-screen intersensory procedures (e.g., Kuhl & Meltzoff, 1982; Walker, 1982). Speed in selecting the target on each trial was calculated as the *latency* from trial onset to

produce a fixation (of at least 50 ms) within the AOI of the target event (or yoked "target" event). Latency on each trial was averaged across all trials within each condition, similar to measures of speed of shifting in infant eye-tracking studies (Amso & Johnson, 2006).

Results

First, we conducted preliminary analyses of differences in the amount of time looking at the screen, the spatial spread of gaze, and the length and number of fixations between the audiovisual and the silent, visual control conditions. Only two significant differences emerged. There was longer overall looking (proportion of available looking time to the screen, PALT), and broader spatial scanning (number of events [AOIs] fixated) in the audiovisual than the silent visual condition, ts > 2.8, ps < .007 (see Supporting Information). These factors were statistically controlled in subsequent analyses to ensure a more precise measure of intersensory processing.

Next, we investigated group differences by examining effects of condition (audiovisual and silent visual; within-participants), event type (social and nonsocial; between-participants), and interactions on speed and accuracy of intersensory processing. Differences between audiovisual and silent visual conditions (controlling for amount and spread of attention) reflect intersensory processing. To assess intersensory processing within the audiovisual condition alone, we compared accuracy in selecting and matching the sound-synchronous target events to attention to the same events when they were asynchronous distractors (controlling for chance looking to different stimuli and areas of the screen). Finally, we also characterized individual differences in the speed and accuracy of intersensory processing and investigated relations among these measures in individual children.

Audiovisual versus Visual Control Events: Speed and Accuracy of Intersensory Processing

Speed, accuracy in selection, and accuracy in matching did not differ consistently across the target event ordering or condition ordering, Fs < 1, ps > .4; number of events (AOIs) fixated on average did differ across condition ordering (see Supporting Information). We also found no differences in measures of speed and accuracy across or within conditions based on participants' sex, race, ethnicity, or home language. Subsequent analyses therefore assessed differences in speed and accuracy of intersensory processing during the audiovisual condition as compared with the silent visual control condition (where there is no intersensory redundancy) collapsing across the above variables. In each analysis and follow up test, we statistically controlled for the two potentially confounding behaviors that differed between audiovisual and silent visual conditions (i.e., duration of attention [PALT] and the number of events fixated).

Accuracy in selecting the sound-synchronous target: Frequency of looking (PTTF). Across the short 6-s trials in the IPEP, children displayed accurate selection of the source of redundant audiovisual stimulation. That is, their gaze landed on the sound-synchronous target event on a greater percentage of test trials in the audiovisual condition (M = 58.44%, SD = 16.09) compared to the percentage of trials on which they fixated on the yoked "target" events in the silent, visual control condition (M = 48.22%, SD = 17.40; see Figure 2A). A 2 × 2 mixed-design analysis of covariance (ANCOVA) confirmed a main effect of condition, $F(1, 58) = 15.99$, $p < .001$, partial $\eta 2 = .22$. There was no main effect of social or nonsocial event type, $F(1, 58) = 1.16$, $p = .29$, partial $\eta 2 = .02$, or interaction of condition with event type, $F(1, 58) = 0.08$, $p = .77$, partial $\eta 2 = .001$. The covariates in the analysis (PALT and number of events fixated) did not interact with condition or event type (see Supporting Information). Planned comparisons (controlling for the mean-centered covariates) revealed that children fixated the target on a greater percentage of trials in audiovisual compared to silent, visual stimulation for both social,

$F(1, 58) = 9.19$, $p = .004$, partial $\eta^2 = .14$, and nonsocial events, $F(1, 58) = 6.88$, $p = .011$, partial $\eta^2 = .11$. Overall, children fixated the source of intersensory redundancy more often than the visually-identical silent stimuli, for both social and nonsocial events, providing evidence of accuracy in intersensory target selection.

Accuracy in matching the sound-synchronous target: Duration of looking (PTLT). Children demonstrated greater accuracy in matching the sound-synchronized target in audiovisual compared to silent visual stimulation. Both the audiovisual target event and the yoked, control "target" event, appeared in the same location on the screen (same AOI) and contained identical visual stimulation, but the addition of synchronized auditory information in the audiovisual trials recruited a greater PTLT to the audiovisual target event than the visual-only control "target". Figure 2B displays the PTLT (in comparison with the chance PTLT of .167, indicated by the horizontal line in the figure) toward the target event as a function of condition and event type. A 2 × 2 mixed-design ANCOVA confirmed a main effect of condition on the PTLT to the target, $F(1, 58) = 19.81$, $p < .001$, partial $\eta 2 = .26$. Averaged across all trials within each condition, PTLT to the target was $M = .224$ ($SD = .084$) in the audiovisual condition and significantly lower, $M = .160$ ($SD = .075$), in the silent, visual condition. Planned comparisons (statistically controlling for the covariates at their mean-centered values) revealed greater PTLTs to the target in audiovisual compared to silent, visual stimulation for the social, $F(1, 58) = 12.07$, $p = .001$, partial $\eta^2 = .17$, and nonsocial events, $F(1, 58) = 7.95$, $p = .007$, partial $\eta^2 = .12$. The ANCOVA also revealed a main effect of event type, $F(1, 58) = 7.28$, $p = .009$, partial $\eta 2 = .11$, indicating greater PTLT to the social than the nonsocial events. However, we found no interaction of condition and event type, $F(1, 58) = 0.22$, $p = .6$, partial $\eta 2 = .004$. Neither

covariate (PALT; number of events fixated) interacted significantly with condition or event type (see Supporting Information).

Moreover, children demonstrated a mean PTLT to the target events that was significantly greater than chance (1 out of 6 AOIs; 0.167) in the audiovisual condition for both social, $t(31) = 6.29$, $p < .001$, $d = 1.11$, and nonsocial events, $t(31) = 2.22$, $p = .034$, $d = 0.39$. As expected, PTLTs in the visual only condition were not greater than chance for social, $t(31) = 0.58$, $p = .57$, $d = 0.21$, or nonsocial events, $t(31) = -1.73$, $p = .09$, $d = -0.62$. In sum, in the audiovisual condition, children fixated the source of intersensory redundancy longer than expected by chance and longer than the same event when it was silent, for both social and nonsocial events, providing evidence of accurate intersensory matching.

Speed in selecting the sound-synchronous target (Latency). Children showed no differences between their speed in selecting the sound-synchronous target in audiovisual stimulation ($M = 2.35$ s, $SD = 0.74$) and their speed to fixate that same yoked "target" event in silent, visual stimulation ($M = 2.14$ s, $SD = 0.80$; see Figure 2C). A 2 × 2 mixed-design ANCOVA (with PALT and number of events fixated as mean-centered covariates) confirmed no effect of condition, $F(1, 58) = 2.10$, $p = .15$, partial $\eta 2 = .04$. The ANCOVA did reveal a main effect of event type, $F(1, 58) = 7.09$, $p = .01$, partial $\eta 2 = .11$, due to faster latency to fixate the target event for social compared to nonsocial events. However, there was no interaction of event type with condition, $F(1, 58) = 0.07$, $p = .8$, partial $\eta 2 = .001$, and follow-up comparisons indicated no reliable differences in speed across conditions within social, $F(1, 58) = 0.70$, $p = .41$, partial $\eta 2 = .01$, or nonsocial events, $F(1, 58) = 1.46$, $p = .23$, partial $\eta 2 = .03$. After controlling for PALT and number of events fixated, the speed in selecting the target event was still not reliably different between the two conditions, $F(1, 58) = 2.09$, $p = .15$, partial $\eta^2 = .04$, and

interactions with covariates did not qualify this finding (see Supporting Information). These findings indicate that children find the target more quickly for social than nonsocial events but show no difference in speed between audiovisual and silent visual stimulation.

In summary, accuracy in both selecting and matching the target event was greater in the audiovisual than the silent visual control condition, providing evidence of intersensory processing. Audiovisual events also broaden spatial scanning (more AOIs fixated) as compared with visual events and promote more individual looks within the target AOI (see Supporting Information). Even when accounting for these differences, children still show accurate intersensory processing, looking longer and on more trials to the sound-synchronous target events. With respect to speed of intersensory processing, children found the target after a mean latency of 2.33 s (*SD* = 0.72).

Audiovisual Stimulation: Intersensory Accuracy

To what extent do we find evidence of intersensory processing within the audiovisual condition alone? To answer this question, we assessed group-level accuracy relative to chance measures of looking behavior. These analyses were designed to test for convergent evidence with those comparing accuracy in audiovisual versus silent visual stimulation above. Because there is no measure of chance for latency, and latency did not differ between stimulation conditions, analyses of latency were not included.

Accuracy in selecting the sound synchronous target: Frequency of looking (PTTF). Children varied in the base number of events they fixated per trial (*M* = 3.15, *SD* = 0.72, *range* = 1.5-4.75). To assess whether children fixated the target on more trials than expected by chance, we compared each child's accuracy in selecting the target (PTTF) relative to the average number of events (out of the 6 possible events) the child fixated per trial (the chance value for the child).

For instance, a child who fixates an average of 2 events per trial has a chance rate of .33 (2 out of 6) of fixating the target event on each trial. Thus, for each child, we computed a difference score between the percentage of trials on which the target was fixated (PTTF) and the mean percentage of events fixated per trial—a "frequency difference" score. Scores with differences greater than 0 indicate that the target was fixated more often than expected if they had fixated just their "baseline" number of events in a random pattern.

Indeed, in the audiovisual condition, children showed a frequency difference score significantly greater than 0 (M = .059, SD = .10), $t(63) = 4.47$, $p < .001$, $d = 0.61$. This difference was also significant for both the social (M = .068, SD = .08), $t(31) = 5.08$, $p < .001$, $d = 0.94$, and nonsocial events (M = .050, SD = 0.13), $t(31) = 2.24$, $p = .033$, $d = 0.53$. In contrast, frequency difference scores were not significantly greater than 0 in the silent, visual control condition for social (M = .006, SD = .12), $t(31) = 0.29$, $p = .77$, $d = 0.10$, or nonsocial events (M = -.009, SD = .11), $t(31) = -0.48$, $p = .63$, $d = 0.17$. In sum, children fixated the audiovisual target event more frequently than expected by chance. Within the audiovisual condition, they showed selective attention to intersensory redundancy over and above their chance distribution of spatial looking.

Accuracy in matching the sound synchronous target: Duration of looking (PTLT). We assessed accuracy in matching using a cross-AOI control. Each AOI in the grid of six social or nonsocial events depicted a different face/object event, which might attract children's attention differentially, depending on its visual appearance or spatial location (AOI). We thus controlled for this by calculating a difference score for each child (see Walker-Andrews et al., 1991). We averaged the child's PTLT to each event's AOI across the 5 trials within a block when that event was silent and subtracted that average from the PTLT to that same event's AOI

when it was the synchronous target. On audiovisual trials, if children are attending to the source of audiovisual redundancy this "duration difference" should be significantly greater than 0.

Indeed, children demonstrated a significant, positive duration difference overall (M = .071, SD = .10), $t(63) = 5.55$, $p < .001$, $d = 0.69$, for social events (M = .095, SD = .08), $t(31) = 6.29$, $p < .001$, $d = 1.11$, and for nonsocial events (M = .046, SD = .11), $t(31) = 2.34$, $p = .026$, $d = 0.41$. In contrast, the duration differences in the silent, visual control condition were not different from 0 for social (M = .012, SD = .09), $t(31) = 0.73$, $p = .47$, $d = 0.26$, or nonsocial events (M = -.018, SD = .08), $t(31) = -1.25$, $p = .22$, $d = -0.45$. In sum, children showed accurate intersensory matching even when accounting for potential, differential interest in the events/AOIs. They looked to the sound-synchronous target event for a greater proportion of the total trial time relative to that same target event when it was not in sound.

Individual Differences in Audiovisual Accuracy and Speed

The IPEP revealed meaningful individual differences in our three measures of intersensory processing in the audiovisual condition. Accuracy in selection (PTTF) ranged from 30% to 100% (CV = .28) with 1/3 of children showing PTTFs below 53% and 1/3 above 66.67%, indicating a concentration of performance within 50-70% but clear groups of low and high performance. Accuracy in matching (PTLT) ranged from .04 to .44 (CV = .38) with 1/3 of children showing PTLTs below .2 and 1/3 above .26, indicating accuracy scores clustering around .2-.25 but groups performing below chance and others well above chance. Speed in selection (latency) ranged from 0.71 s to 3.83 s (CV = .31) with 1/3 of children showing latencies below 1.92 s and 1/3 showing latencies above 2.57 s, suggesting a relatively even spread of performance across the range of observed latencies. Moreover, there was a wide range in how

speed and accuracy in selection interacted for individual children and individual differences in scanning patterns (see Supporting Information).

Relations Between Measures of Speed and Accuracy

At the individual level, children showed several inter-relations among measures of speed, accuracy, and attention allocation measures in the audiovisual IPEP (see Table 1). Of note is the significant, positive correlation between the two measures of accuracy. However, there were also marginal correlations between our attention allocation covariates and intersensory speed and accuracy. Thus, we used a regression framework to reveal the unique relations among measures of speed and accuracy, independent of any effects of attention allocation.

In a multiple regression (Table 2), we found evidence of unique relations among the measures of intersensory processing efficiency (speed and accuracy) at the individual participant level. As seen in Table 2A, after controlling for attention allocation measures and holding speed constant, children's accuracy in selecting (PTTF) and matching (PTLT) the sound synchronous target were positively related, $p < .001$. Children who fixated the target more frequently across trials looked at that event longer within each trial. Moreover, after controlling for attention allocation and holding accuracy in matching constant, children's speed (latency) and accuracy in selecting the target (PTTF) were positively related, $p = .008$. Children who found the target more frequently, typically had fixated that target event later in the trial (i.e., longer latencies)—suggesting that some children continued to search for the target longer into each trial than other children, and as a result, found the target more often. As shown in Table 2B, after controlling for attention allocation measures and holding constant accuracy in selecting the target, speed in target selection (latency) and accuracy in matching the target (PTLT) were significantly negatively related, $p = .01$. Independent of how often they located the target, children who found

the target earlier in the trial (after a shorter latency) looked to the target longer (greater matching). See Supporting Information for additional variance decomposition analyses.

Discussion

Although sensitivity to intersensory redundancy has been shown to guide and organize early cognitive, social, and language development, developmental research has been limited by methods designed for group-level approaches. We lacked fine-grained measures appropriate for assessing individual differences in intersensory processing skills in nonverbal participants to characterize how these early skills are refined across development and lead to later outcomes. The Intersensory Processing Efficiency Protocol (IPEP) was developed to address this need.

The present study provides a foundation for the IPEP as a fine-grained individual difference measure of intersensory processing efficiency (speed and accuracy) appropriate for young children. We tested 3- to 5-year-olds in the IPEP using remote eye-tracking. Children saw a number of short trials depicting either six concurrent social events (women speaking) or nonsocial events (objects striking a surface), with a different target event in synchrony with its natural soundtrack on each trial. We assessed the speed and accuracy of intersensory processing (looking to the sound-synchronized target event) in this audiovisual condition and also compared it with the same children's looking patterns in a silent, visual control condition. The current study offers a proof of concept of the feasibility of the IPEP and also illuminates novel and previously unstudied aspects of children's intersensory processing abilities.

Intersensory Processing Accuracy

Children demonstrated convergent evidence for accurate intersensory processing in the IPEP across two types of analyses. *Audiovisual versus visual stimulation*: Children accurately selected the sound-synchronous target events. They fixated the audiovisual target more

frequently than the same event when it was a silent control "target" (in a visual-only yoked control condition). And once they fixated it, they attended to it longer than the silent control. However, they showed no difference in the speed of fixating the sound-synchronous target versus the silent control. *Audiovisual stimulation alone*: During the audiovisual trials, children found and fixated the synchronous target more often than expected by chance (i.e., compared to the child's own average number of events fixated per trial). And once found, children attended to the synchronous target longer than to the five asynchronous distractor events on each trial.

These findings replicate and extend prior research demonstrating intersensory processing of speech and nonsocial events in young children (Bebko et al., 2006; Lewkowicz & Flom, 2014; Mongillo et al., 2008; Stevenson et al., 2014). They do so with a more complex display (more concurrent visual sources of information) and under more attentionally demanding conditions (shorter trials, continuous events) than in prior studies.

Individual Relations Among Speed and Accuracy of Intersensory Processing

Children showed consistent correlations between speed in selecting, accuracy in selecting, and accuracy in matching the audiovisual-synchronous event in multiple regression analyses—even when statistically controlling for their overall attention to the display and their spatial spread of visual scanning. Children with higher accuracy in selecting the target (finding it on more trials) also showed higher accuracy in matching the sound-synchronous target event (looking to it longer). Although fixating the target is a requisite for continuing to attend to it, we found a meaningful and substantial range in accuracy of matching (PTLT) even among children who found a similar number of targets (PTTF). There was substantial unshared variability between these two measures (~66%). Children who found the target more quickly (shorter latency) also showed higher accuracy in matching the sound-synchronous target (looking to it

longer) and lower accuracy in selecting the target (finding it on fewer trials). This suggests that children who continued to search longer into the trial found the target on more trials but looked at it for shorter durations. Further, analyses of scanning patterns revealed novel information about looking strategies. Longer looking to the sound-synchronous target events (greater accuracy in matching) stemmed from a number of frequent shorter fixations rather than from finding the target and attending to it continuously (see Supporting Information). These findings provide new insight into the correlations between speed and accuracy of intersensory processing and the scanning patterns used during exploration of social and nonsocial audiovisual events.

Speed and accuracy in selecting and matching the source of intersensory redundancy, however, likely stem from distinct visual search strategies, given that they shared little variance (~9%) in the present study. On the other hand, the two measures of accuracy (selecting and matching) likely reflect a more closely-related underlying construct, since they shared a greater proportion of variance (34%). These findings are consistent with the view that the IPEP indexes a combination of speed and accuracy, a construct we call intersensory processing efficiency. A similar coupling of speed and accuracy (language processing efficiency) is found by Fernald and colleagues (2006).

Differences in Scanning Audiovisual Versus Visual Events

The IPEP also provides opportunities for micro level analyses of attention allocation. Children demonstrated a number of differences in both spatial and temporal scanning patterns in stimulation from audiovisual versus visual only control events. Spatially, during audiovisual stimulation, children broadened their scanning to fixate more events, whereas during silent, visual stimulation, scanning was more constrained. The proportion of available time that children spent looking at the display and number of events fixated reliably increased in the presence of

audiovisual stimulation. Temporally, in the audiovisual condition, children produced more frequent fixations to the target throughout each trial, returning their gaze most often to the sound synchronous target event, in comparison with the silent, visual events (see Supporting Information). Rather than capturing attention immediately and for extended periods of time, intersensory redundancy recruits broader and more frequent visual foraging.

Although prior research has investigated differences in looking to audiovisual versus visual stimulation, particularly in infants (Bahrick, Todd, Castellanos, & Sorondo, 2016; Reynolds, Zhang, & Guy, 2013), this research has focused primarily on measures of look duration. Little research has characterized differences in speed, scanning patterns, or strategies for exploring audiovisual and visual stimulation. This relatively novel level of analysis for assessing intersensory processing is also valuable for characterizing individual differences in strategies of attention allocation (see Figure S2 in Supporting Information).

Further, children engage in what appears to be a more sequential acoustically-driven visual search in the IPEP. We found no difference in children's speed of fixating the sound-synchronous target event in audiovisual stimulation as compared with fixating the yoked control "target" event in silent, visual stimulation. Thus, there is no evidence that the acoustically-synchronous target event "pops out" from the background of asynchronous events. These findings provide evidence of serial visual search for the source of intersensory redundancy earlier in development than previously documented, paralleling findings of adults (Fujisaki, Koene, Arnold, Johnston, & Nishida, 2006).

The IPEP is a Promising Measure of Individual Differences in Intersensory Processing

A number of factors make the IPEP promising for use as an individual difference measure. 1) By assessing three complementary indices of intersensory processing, the IPEP

provides a more comprehensive picture of intersensory skills than prior methods. 2) It provides a fine-grained assessment of individual variation in intersensory processing (with appropriate variability across children) by averaging performance across a number of short trials (rather than fewer trials of longer duration). 3) The protocol relies on visual attention and requires no verbal skills, providing a common metric for assessing change across a wide age range. 4) It simulates the complexity of the multimodal environment—with multiple concurrent sources of visual stimulation—and thus can be generalized to natural, complex, learning environments.

Refined Audiovisual Social/Nonsocial IPEP Protocol for Use Across the Lifespan

We have recently developed and refined the IPEP to provide both social and nonsocial events in a single protocol and to be used from infancy through adulthood. After establishing convergent evidence of intersensory skills in the current study—in the audiovisual versus visual control condition and within the audiovisual condition alone—we were able to eliminate the visual-only trials blocks, provide blocks of both social and nonsocial events to each participant, while at the same time doubling the number of trials for each event type. This refined protocol has the potential to enhance reliability of the measure and allows for a within-participant comparison of performance on social and nonsocial events, important for characterizing atypical development, such as autism. Further, given that infants process information more slowly than children, we lengthened the trials from 6 to 8 s. Thus, the refined audiovisual social/nonsocial protocol has 48 8-s trials, with 24 social and 24 nonsocial trials presented in four alternating blocks of 12 social and 12 nonsocial trials. With these modifications, the IPEP can now provide a fine-grained index of intersensory processing for both social and nonsocial events for infants, children, and adults alike.

Preliminary findings using the refined protocol demonstrate that infants also show reliable evidence of intersensory processing on the IPEP (Bahrick, Soska, Todd, Saunders, & Bein, 2014), intersensory accuracy predicts receptive language in the first year of life (Soska, Todd, & Bahrick, 2016), and IPEP performance predicts pre-literacy skills in preschoolers (Bahrick, McNew, Todd, Martinez, Cheatham-Johnson, & Hart, 2017). The IPEP captures meaningful individual differences in intersensory processing of social and nonsocial events, which likely translate to differences in the language and social skills that rely on this foundation.

Why might better intersensory processing skills assessed by the IPEP translate to enhanced language and literacy skills? Children with better intersensory processing (faster and more accurate audiovisual matching) are likely to have faster, more accurate word mapping (linking speech sounds to objects; see Gogate & Hollich, 2010; Gogate & Maganti, 2016). Children with better intersensory processing skills are also likely to learn more readily from the social contexts (which provide complex and rapidly changing intersensory information) that support language and social development. These gains in language skills earlier in development likely cascade into improved cognitive, language, and other academic skills later in childhood (e.g., Marchman & Fernald, 2008).

The IPEP offers researchers the potential to characterize the development of intersensory processing across the entire lifespan using a single, common protocol. The display is sufficiently complex to challenge adult attentional selectivity, and measures of speed of intersensory processing can vary or improve even when accuracy is at ceiling. At the same time, the protocol is simple enough for infants and children to show reliable intersensory matching. Further, despite using a common protocol, participants of different ages will likely show selective attention to different properties of the events and use different processing strategies (see Bahrick, 2001;

Bahrick, Todd, Castellanos, & Sorondo, 2016; Franchak, Heeger, Hasson, & Adolph, 2016; Frank, Vul, & Saxe, 2012). For older children and adults, the linguistic content of the social events may be more salient as their attention and language skills mature and affect speech perception and audiovisual processing (for related ideas see Bowerman & Levinson, 2001; Frank, Vul, & Saxe, 2012). Thus, given the richness of the audiovisual events, the IPEP provides opportunities for detecting and processing different kinds information, making the protocol engaging and appropriate for characterizing intersensory processing skills across the lifespan.

Limitations and Future Research Directions

The present study lays a foundation for using the IPEP as a measure of fine-grained individual differences in intersensory processing, appropriate for exploring longitudinal change and assessing relations with developmental outcomes using a nonverbal procedure. However, a number of areas require further development.

Reliability. Measures of test-retest reliability are needed to provide an accurate estimate of reliability for both speed and accuracy of intersensory processing. Reliability of the measure will delimit its ability to predict outcomes. This research is currently in progress in our adaptation of this measure for infants.

Validity. The validity of the IPEP will need to be more thoroughly established. We demonstrate meaningful relations within our protocol between speed and two measures of accuracy. Given that detecting audiovisual redundancy is critical for the typical development of social and language skills (Bahrick & Lickliter, 2012; Suanda, Smith, & Yu, 2017; Vaillant-Molina & Bahrick, 2012), we also expect that individual differences in intersensory speed and accuracy to be linked to downstream developments these domains. This research is currently in progress as described above (see Refined Audiovisual Social/Nonsocial IPEP Protocol).

Characterizing underlying attention constructs. Accurately characterizing the constructs underlying each of the IPEP measures is critical for evaluating individual differences in intersensory processing, characterizing deficits, and developing interventions to enhance intersensory skills. To further explore relations between IPEP measures and the underlying constructs they reflect, future research can evaluate structural models depicting interrelations among the measures and outcomes. They can also assess whether developmental gains and/or deficits in one facet of intersensory processing (e.g., accuracy in selecting) show parallel development gains/deficits in another facet (e.g., accuracy in matching). Convergent findings would suggest a common construct.

Speed of intersensory processing. Although children did not look to the sound-synchronous target event more quickly in the audiovisual than in the visual control condition, there was ample variability across children to use latency as an individual difference variable for predicting outcomes. Further, once children found the synchronous target, they looked to it longer than in silent visual stimulation. Thus, differences in speed provide different opportunities for further processing an event. At what ages and under what conditions children show faster shifting to audiovisual than visual-only stimulation is thus an important topic for future research.

Atypical populations. The viability of using the IPEP with nonverbal children and those of atypical development must be explored. If successful, this would provide the first measure of intersensory processing that could be utilized across a variety of populations and ages without changes to the protocol. It could be used to characterize intersensory processing in children who show impairments, including children with ASD (Bebko et al., 2006; Stevenson et al., 2014), children born preterm (Gogate et al., 2014; Pickens et al., 1994), and children with dyslexia (Hairston, Burdette, Flowers, Wood, & Wallace, 2005). Because the IPEP provides both social

and nonsocial events, it can identify processing impairments specific to social events. Social events present both dynamic faces and linguistic information, both areas of processing impairments in atypical development. Thus, differential effects across social and nonsocial events could reflect impairments in processing faces, language, or both (see Patten, Labban, Casenhiser, & Cotton, 2016). Early identification of intersensory processing impairments has potential to aid in identifying those at high risk for language, social, and cognitive impairments, as well as inform early interventions focused on improving intersensory processing skills.

Summary

The IPEP is a novel and unique tool for assessing fine-grained individual differences in attention to dynamic audiovisual social and nonsocial events across development. We illustrate its effectiveness and the range of new information it can provide in a sample of 3- to 5-year-old children. Because it requires no verbal skills, the IPEP is appropriate for studying intersensory processing across a wide age range and in typical and atypical populations. The availability of a fine-grained individual difference measure opens the door to exploring relations between basic intersensory processing skills and more complex, later developing social, cognitive, and language skills that rely on this foundation.

References

Altvater-Mackensen, N., & Grossmann, T. (2015). Learning to match auditory and visual speech cues: Social influences on acquisition of phonological categories. *Child Development*, *86*(2), 362–378. http://doi.org/10.1111/cdev.12320

Altvater-Mackensen, N., Mani, N., & Grossmann, T. (2016). Audiovisual speech perception in infancy: The influence of vowel identity and infants' productive abilities on sensitivity to (mis)matches Between auditory and visual speech cues. *Developmental Psychology*, *52*(2), 191–204. http://doi.org/10.1037/a0039964

Amso, D., & Johnson, S. P. (2006). Learning by selection: Visual search and object perception in young infants. *Developmental Psychology*, *42*(6), 1236–1245. http://doi.org/10.1037/0012-1649.42.6.1236

Bahrick, L. E. (1983). Infants' perception of substance and temporal synchrony in multimodal events. *Infant Behavior & Development*, *6*(4), 429–451. http://doi.org/10.1016/S0163-6383(83)90241-2

Bahrick, L. E. (1987). Infant's intermodal perception of two levels of temporal structure in natural events. *Infant Behavior & Development*, *10*, 387–416.

Bahrick, L. E. (1988). Intermodal learning in infancy: Learning on the basis of two kinds of invariant relations in audible and visible events. *Child Development*, *59*(1), 197–209.

Bahrick, L. E. (1994). The development of infants' sensitivity to arbitrary intermodal relations. *Ecological Psychology*, *6*(2), 111–123. http://doi.org/10.1207/s15326969eco0602_2

Bahrick, L. E. (2000). Increasing specificity in the development of intermodal perception. In D. Muir & A. Slater (Eds.), *Infant development: The essential readings* (pp.117-136). Oxford, England: Blackwell.

Bahrick, L. E. (2001). Increasing specificity in perceptual development: Infants' detection of nested levels of multimodal stimulation. *Journal of Experimental Child Psychology, 79,* 253-270. http://doi.org/10.1006/jecp.2000.2588

Bahrick, L. E. (2010). Intermodal perception and selective attention to intersensory redundancy: Implications for typical social development and autism. In G. Bremner & T. D. Wachs (Eds.), *Blackwell handbook of infant development* (2nd ed., pp. 120–166). Oxford, England: Blackwell.

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36*(2), 190–201. http://doi.org/10.1037//0012-1649.36.2.190

Bahrick, L. E., & Lickliter, R. (2002). Intersensory redundancy guides early perceptual and cognitive development. In R. Kail (Ed.), *Advances in child development and behavior* (Vol. 30, pp. 153–187). New York: Academic Press.

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 183–205). Oxford, England: Oxford University.

Bahrick, L. E., & Lickliter, R. (2014). Learning to attend selectively: The dual role of intersensory redundancy. *Current Directions in Psychological Science, 23*(6), 414–420. http://doi.org/10.1177/0963721414549187

Bahrick, L. E., & Todd, J. T. (2012). Multisensory processing in Autism Spectrum Disorders: Intersensory processing disturbance as a basis for atypical development. In B. E. Stein (Ed.), *The new handbook of multisensory processing* (pp. 657–674). Cambridge, MA: MIT.

Bahrick, L. E., Flom, R., & Lickliter, R. (2002). Intersensory redundancy facilitates discrimination of tempo in 3-month-old infants. *Developmental Psychobiology*, *41*(4), 352–363. http://doi.org/10.1002/dev.10049

Bahrick, L. E., McNew, M. E., Todd, J. T., Martinez, J., Cheatham-Johnson, R., & Hart, K. C. (2017, April). *Individual differences in intersensory processing predict pre-literacy skills in young children*. Poster presented at the Society for Research in Child Development. Austin, TX.

Bahrick, L. E., Soska, K. C., Todd, J. T., Saunders, J. F., & Bein, V. (2014, July). *Assessing individual differences and age-related changes in intersensory processing across infancy: A new method*. Poster presented at the International Conference on Infant Studies. Berlin, Germany.

Bahrick, L. E., Todd, J. T., Castellanos, I., & Sorondo, B. M. (2016). Enhanced attention to speaking faces versus other event types emerges gradually across infancy. *Developmental Psychology*, *52*(11), 1705–1720. http://doi.org/10.1037/dev0000157

Bebko, J. M., Weiss, J. A., Demark, J. L., & Gomez, P. (2006). Discrimination of temporal synchrony in intermodal events by children with autism and children with developmental disabilities without autism. *Journal of Child Psychology and Psychiatry*, *47*(1), 88–98. http://doi.org/10.1111/j.1469-7610.2005.01443.x

Bowerman, M., & Levinson, S. (Eds.). (2001). *Language acquisition and conceptual development*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620669

Caron, A. J., Caron, R. F., & MacLean, D. J. (1988). Infant discrimination of naturalistic emotional expressions: The role of face and voice. *Child Development*, *59*(3), 604. http://doi.org/10.2307/1130560

Curtin, S., Campbell, J., & Hufnagle, D. (2012). Mapping novel labels to actions: How the rhythm of words guides infants' learning. *Journal of Experimental Child Psychology*, *112*(2), 127–140. http://doi.org/10.1016/j.jecp.2012.02.007

Dawson, G., Meltzoff, A. N., Osterling, J., Rinaldi, J., & Brown, E. (1998). Children with autism fail to orient to naturally occurring social stimuli. *Journal of Autism and Developmental Disorders*, *28*(6), 479–485.

Eppler, M. A. (1995). Development of manipulatory skills and the deployment of attention. *Infant Behavior & Development*, *18*(4), 391–405.

Fernald, A., & Marchman, V. A. (2012). Individual differences in lexical processing at 18 months predict vocabulary growth in typically developing and late-talking toddlers. *Child Development*, *83*(1), 203–222. http://doi.org/10.1111/j.1467-8624.2011.01692.x

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98–116. http://doi.org/10.1037/0012-1649.42.1.98

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A. (2008). Looking while listening: Using eye movements to monitor spoken language. In I. A. Sekerina, E. M. Fernandez, & H. Clahsen (Eds.), *Developmental psycholinguistics* (Vol. 44, pp. 97–135). Philadelphia, PA: John Benjamins.

Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, *43*(1), 238–252. http://doi.org/10.1037/0012-1649.43.1.238

Foss-Feig, J. H., Kwakye, L. D., Cascio, C. J., Burnette, C. P., Kadivar, H., Stone, W. L., & Wallace, M. T. (2010). An extended multisensory temporal binding window in autism

spectrum disorders. *Experimental Brain Research*, *203*(2), 381–389. http://doi.org/10.1007/s00221-010-2240-4

Foxe, J. J., Molholm, S., Del Bene, V. A., Frey, H.-P., Russo, N. N., Blanco, D., et al. (2015). Severe multisensory speech integration deficits in high-functioning school-aged children with Autism Spectrum Disorder (ASD) and their resolution during early adolescence. *Cerebral Cortex*, *25*(2), 298–312. http://doi.org/10.1093/cercor/bht213

Franchak, J. M., Heeger, D. J., Hasson, U., & Adolph, K. E. (2016). Free viewing gaze behavior in infants and adults. *Infancy*, *21*(3), 262–287. http://doi.org/10.1111/infa.12119

Frank, M. C., Vul, E., & Saxe, R. (2012). Measuring the development of social attention using free-viewing. *Infancy*, *17*(4), 355–375. http://doi.org/DOI: 10.1111/j.1532-7078.2011.00086.x

Fujisaki, W., Koene, A., Arnold, D., Johnston, A., & Nishida, S. (2006). Visual search for a target changing in synchrony with an auditory signal. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1588), 865–874. http://doi.org/10.1098/rspb.2005.3327

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gogate, L. J. (2010). Learning of syllable-object relations by preverbal infants: the role of temporal synchrony and syllable distinctiveness. *Journal of Experimental Child Psychology*, *105*(3), 178–197. http://doi.org/10.1016/j.jecp.2009.10.007

Gogate, L. J., & Bahrick, L. E. (1998). Intersensory redundancy facilitates learning of arbitrary relations between vowel sounds and objects in seven-month-old infants. *Journal of Experimental Child Psychology*, *69*(2), 133–149. http://doi.org/10.1006/jecp.1998.2438

Gogate, L. J., Bolzani, L., & Betancourt, E. A. (2006). Attention to maternal multimodal aming

by 6- to 8-month-old infants and learning of word–object relations. *Infancy, 9*(3), 259–288. http://doi.org/10.1207/s15327078in0903_1

Gogate, L. J., & Hollich, G. (2010). Invariance detection within an interactive system: A perceptual gateway to language development. *Psychological Review, 117*(2), 496–516. http://doi.org/10.1037/a0019049

Gogate, L. J., & Maganti, M. (2016). The dynamics of infant attention: Implications for crossmodal perception and word-mapping research. *Child Development, 87*(2), 345–364. http://doi.org/10.1111/cdev.12509

Gogate, L. J., Maganti, M., & Perenyi, A. (2014). Preterm and term infants' perception of temporally coordinated syllable–object pairings: Implications for lexical development. *Journal of Speech, Language, and Hearing Research, 57*(1), 187–198. http://doi.org/10.1044/1092-4388(2013/12-0403)

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). Intersensory origins of word comprehension: An ecological-dynamic systems view. *Developmental Science, 4*, 1-37. http://doi.org/10.1111/1467-7687.00143

Grossman, R. B., Schneps, M. H., & Tager Flusberg, H. (2009). Slipped lips: Onset asynchrony detection of auditory-visual language in autism. *Journal of Child Psychology and Psychiatry, 50*(4), 491–497. http://doi.org/10.1111/j.1469-7610.2008.02002.x

Hairston, W. D., Burdette, J. H., Flowers, D. L., Wood, F. B., & Wallace, M. T. (2005). Altered temporal profile of visual-auditory multisensory interactions in dyslexia. *Experimental Brain Research, 166*(3-4), 474–480. http://doi.org/10.1007/s00221-005-2387-6

Jesse, A., & Johnson, E. K. (2016). Audiovisual alignment of co-speech gestures to speech supports word learning in 2-year-olds. *Journal of Experimental Child Psychology, 145*,

1–10. http://doi.org/10.1016/j.jecp.2015.12.002

Kaganovich, N. (2016). Development of sensitivity to audiovisual temporal asynchrony during midchildhood. *Developmental Psychology*, *52*(2), 232–241. http://doi.org/10.1037/dev0000073

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*(4577), 1138–1141. http://doi.org/10.1126/science.7146899

Lewkowicz, D. J. (1992). Infants' response to temporally based intersensory equivalence: The effect of synchronous sounds on visual preferences for moving stimuli. *Infant Behavior & Development*, *15*(3), 297–324. http://doi.org/10.1016/0163-6383(92)80002-C

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126*(2), 281–308.

Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: the hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology*, *39*(5), 795–804.

Lewkowicz, D. J. (2010). Infant perception of audio-visual speech synchrony. *Developmental Psychology*, *46*(1), 66–77. http://doi.org/10.1037/a0015579

Lewkowicz, D. J. (2014). Early experience and multisensory perceptual narrowing. *Developmental Psychobiology*, *56*(2), 292–315. http://doi.org/10.1002/dev.21197

Lewkowicz, D. J., & Flom, R. (2014). The audiovisual temporal binding window narrows in early childhood. *Child Development*, *85*(2), 685-694. http://doi.org/10.1111/cdev.12142

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences*, *109*(5), 1431–1436. http://doi.org/10.1073/pnas.1114783109

Lewkowicz, D. J., Leo, I., & Simion, F. (2010). Intersensory perception at birth: Newborns match nonhuman primate faces and voices. *Infancy*, *15*(1), 46–60. http://doi.org/10.1111/j.1532-7078.2009.00005.x

Marchman, V. A., Adams, K. A., Loi, E. C., Fernald, A., & Feldman, H. M. (2015). Early language processing efficiency predicts later receptive vocabulary outcomes in children born preterm. *Child Neuropsychology*, 1–17. http://doi.org/10.1080/09297049.2015.1038987

Marchman, V. A., & Fernald, A. (2008). Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental Science*, *11*(3), F9–16. http://doi.org/10.1111/j.1467-7687.2008.00671.x

Matatyaho-Bullaro, D. J., Gogate, L. J., Mason, Z., Cadavid, S., & Abdel-Mottaleb, M. (2014). Type of object motion facilitates word mapping by preverbal infants. *Journal of Experimental Child Psychology*, *118*, 27–40. http://doi.org/10.1016/j.jecp.2013.09.010

Montague, D. P. F., & Walker-Andrews, A. S. (2002). Mothers, fathers, and infants: The role of person familiarity and parental involvement in infants' perception of emotion expressions. *Child Development*, *73*(5), 1339–1352. http://doi.org/10.1111/1467-8624.00475

Nomikou, I., Koke, M., & Rohlfing, K. J. (2017). Verbs in mothers' input to six-month-olds: Synchrony between presentation, meaning, and actions is related to later verb acquisition. *Brain Sciences*, *7*(5), 1–19. http://doi.org/10.3390/brainsci7050052

Olsen, A. (2012). *The Tobii I-VT fixation filter*. *Tobii Technology*.

Patten, E., Labban, J. D., Casenhiser, D. M., & Cotton, C. L. (2016). Synchrony detection of linguistic stimuli in the presence of faces: Neuropsychological implications for language

development in ASD. *Developmental Neuropsychology*, *41*(5-8), 362–374. http://doi.org/10.1080/87565641.2016.1243113

Patten, E., Watson, L. R., & Baranek, G. T. (2014). Temporal synchrony detection and associations with language in young children with ASD. *Autism Research and Treatment*, *2014*, 1–8. http://doi.org/10.1155/2014/678346

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior & Development*, *22*(2), 237–247. http://doi.org/10.1016/S0163-6383(99)00003-X

Patterson, M. L., & Werker, J. F. (2003). Two-month-old infants match phonetic information in lips and voice. *Developmental Science*, *6*(2), 191–196. http://doi.org/10.1111/1467-7687.00271

Pereira, A. F., Smith, L. B., & Yu, C. (2014). A bottom-up view of toddler word learning. *Psychonomic Bulletin & Review*, *21*(1), 178–185. http://doi.org/10.3758/s13423-013-0466-4

Pickens, J., Field, T., Nawrocki, T., Martinez, A., Soutullo, D., & Gonzalez, J. (1994). Full-term and preterm infants' perception of face-voice synchrony. *Infant Behavior & Development*, *17*(4), 447–455. http://doi.org/10.1016/0163-6383(94)90036-1

Provasi, J., Lemoine-Lardennois, C., Orriols, E., & Morange-Majoux, F. (2017). Do preterm infants perceive temporal synchrony? An analysis with the eye-tracking system. *Timing & Time Perception*, *5*(2), 190–209. http://doi.org/10.1163/22134468-00002089

Reynolds, G. D., Zhang, D., & Guy, M. W. (2013). Infant attention to dynamic audiovisual stimuli: Look duration from 3 to 9 months of age. *Infancy*, *18*(4), 554–577. http://doi.org/DOI: 10.1111/j.1532-7078.2012.00134.x

Rose, S. A., Feldman, J. F., & Jankowski, J. J. (2012a). Implications of infant cognition for executive functions at age 11. *Psychological Science*, *23*(11), 1345–1355. http://doi.org/10.1177/0956797612444902

Rose, S. A., Feldman, J. F., Jankowski, J. J., & Van Rossem, R. (2012b). Information processing from infancy to 11 years: Continuities and prediction of IQ. *Intelligence*, *40*(5), 445–457. http://doi.org/10.1016/j.intell.2012.05.007

Ross, L. A., Molholm, S., Blanco, D., Gomez-Ramirez, M., Saint-Amour, D., & Foxe, J. J. (2011). The development of multisensory speech perception continues into the late childhood years. *The European Journal of Neuroscience*, *33*(12), 2329–2337. http://doi.org/10.1111/j.1460-9568.2011.07685.x

Samuelson, L. K., Smith, L. B., Perry, L. K., & Spencer, J. P. (2011). Grounding Word Learning in Space. *PloS One*, *6*(12), e28095–13. http://doi.org/10.1371/journal.pone.0028095

Seibold, D. R., & McPhee, R. D. (1979). Commonality analysis: A method for decomposing explained variance in multiple regression analyses. *Human Communication Research*, *5*(4), 355–365. http://doi.org/10.1111/j.1468-2958.1979.tb00649.x

Smith, E. G., & Bennetto, L. (2007). Audiovisual speech integration and lipreading in autism. *Journal of Child Psychology and Psychiatry*, *48*(8), 813–821. http://doi.org/10.1111/j.1469-7610.2007.01766.x

Soken, N. H., & Pick, A. D. (1992). Intermodal perception of happy and angry expressive behaviors by seven-month-old infants. *Child Development*, *63*(4), 787–795.

Soska, K. C., Todd, J. T., & Bahrick, L. E. (2016, May). *Individual differences in growth rate of intersensory processing are related to early language skills*. Poster presented at the International Congress on Infant Studies. New Orleans, LA.

Spelke, E. (1976). Infants' intermodal perception of events. *Cognitive Psychology*, *8*(4), 553–560. http://doi.org/10.1016/0010-0285(76)90018-9

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., Camarata, S., & Wallace, M. T. (2016). Keeping time in the brain: Autism spectrum disorder and audiovisual temporal processing. *Autism Research*, *9*(7), 720–738. http://doi.org/10.1002/aur.1566

Stevenson, R. A., Segers, M., Ncube, B. L., Black, K. R., Bebko, J. M., Ferber, S., & Barense, M. D. (2017). The cascading influence of multisensory processing on speech perception in autism. *Autism*, *98*(2), 1362361317704413. http://doi.org/10.1177/1362361317704413

Stevenson, R. A., Siemann, J. K., Schneider, B. C., Eberly, H. E., Woynaroski, T. G., Camarata, S. M., & Wallace, M. T. (2014). Multisensory temporal integration in autism spectrum disorders. *Journal of Neuroscience*, *34*(3), 691–697. http://doi.org/10.1523/JNEUROSCI.3615-13.2014

Suanda, S. H., Smith, L. B., & Yu, C. (2017). The multisensory nature of verbal discourse in parent–toddler interactions. *Developmental Neuropsychology*, *41*(5-8), 324–341. http://doi.org/10.1080/87565641.2016.1256403

Swettenham, J., Baron-Cohen, S., Charman, T., Cox, A., Baird, G., Drew, A., et al. (1998). The frequency and distribution of spontaneous attention shifts between social and nonsocial stimuli in autistic, typically developing, and nonautistic developmentally delayed infants. *Journal of Child Psychology and Psychiatry*, *39*(5), 747–753.

Tenenbaum, E. J., Shah, R. J., Sobel, D. M., Malle, B. F., & Morgan, J. L. (2012). Increased focus on the mouth among infants in the first year of life: A longitudinal eye-tracking study. *Infancy*, *18*(4), 534–553. http://doi.org/10.1111/j.1532-7078.2012.00135.x

Tomalski, P., Ribeiro, H., Ballieux, H., Axelsson, E. L., Murphy, E., Moore, D. G., &

Kushnerenko, E. (2013). Exploring early developmental changes in face scanning patterns during the perception of audiovisual mismatch of speech cues. *European Journal of Developmental Psychology*, *10*(5), 611–624. http://doi.org/10.1080/17405629.2012.728076

Vaillant-Molina, M., & Bahrick, L. E. (2012). The role of intersensory redundancy in the emergence of social referencing in 5½-month-old infants. *Developmental Psychology*, *48*(1), 1–9. http://doi.org/10.1037/a0025263

Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology*, *33*(3), 514–535. http://doi.org/10.1016/0022-0965(82)90063-7

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, *121*(3), 437–456.

Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology*, *3*(2), 55–75. http://doi.org/10.1207/s15326969eco0302_1

Woynaroski, T. G., Kwakye, L. D., Foss-Feig, J. H., Stevenson, R. A., Stone, W. L., & Wallace, M. T. (2013). Multisensory speech perception in children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *43*(12), 2891–2902. http://doi.org/10.1007/s10803-013-1836-5

Yu, C., & Smith, L. B. (2012). Embodied attention and word learning by toddlers. *Cognition*, *125*(2), 244–262. http://doi.org/10.1016/j.cognition.2012.06.016

Table 1

Correlations among measures of speed and accuracy on the IPEP (PTTF, PTLT, Latency) and attention allocation measures (PALT, #Events)

	PTTF	PTLT	PALT	#Events
Speed-Selecting (Latency)	r = .14	r = -.16	r = -.26	r = .06
	p = .26	p = .22	p = .037	p = .66
Accuracy-Selecting (PTTF)		r = .52	r = .25	r = .76
		p < .001	p = .047	p < .001
Accuracy-Matching (PTLT)			r = .19	r = .10
			p = .13	p = .46
Attention to Display (PALT)				r = .35
				p = .005

Note. PTTF, proportion of trials on which the target was fixated; PTLT, proportion of total looking time to the target; Latency to fixate target; PALT, proportion of available looking time to the display; #Events, number of events fixated per trial.

Table 2

Multiple regression analyses on measures of intersensory accuracy controlling for the other measures of speed and accuracy in the IPEP (PTTF, PTLT, Latency) and attention allocation measures (PALT, #Events)

A. Dependent Variable: Accuracy in Selecting (PTTF)

	Unstandardized *B* (SE)	β	*p*
PALT	-0.06 (.07)	-0.05	.4
#Events	0.16 (.01)	0.73	< .001
Latency	0.04 (.01)	0.16	.008
PTLT	0.94 (.11)	0.49	< .001
Final Model: $F(4, 59) = 65.61$, $p < .001$			

B. Dependent Variable: Accuracy in Matching (PTLT)

	Unstandardized *B* (SE)	β	*p*
PALT	0.07 (.05)	0.12	.2
#Events	-0.09 (.02)	-0.80	< .001
Latency	-0.03 (.01)	-0.24	.01
PTTF	0.59 (.07)	1.13	< .001
Final Model: $F(4, 59) = 19.95$, $p < .001$			

Note. PTTF, proportion of trials on which the target was fixated; PTLT, proportion of total looking time to the target; Latency to fixate target; PALT, proportion of available looking time to the display; #Events, number of events fixated per trial.

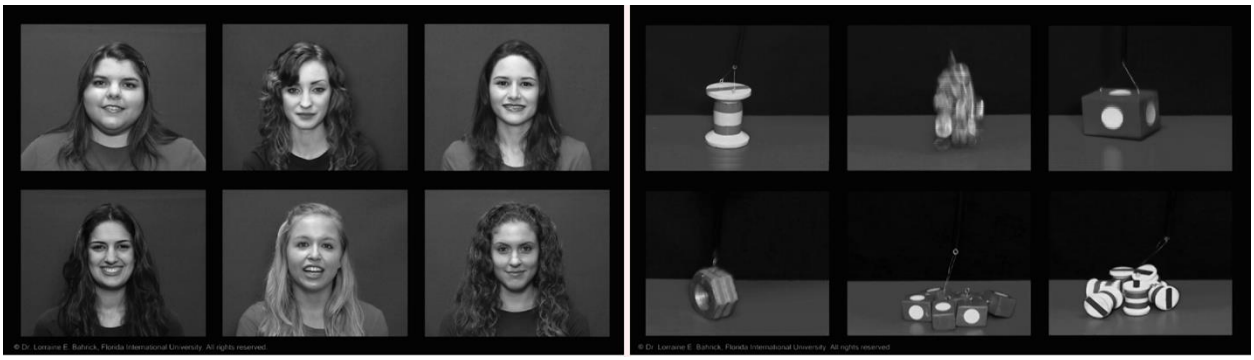

Figure 1. Still image of the dynamic social (left) and nonsocial (right) events presented to children in the IPEP. All six events moved on every trial, but on each trial a different woman or object was synchronized with the accompanying, natural soundtrack. The actresses appearing in the social events provided signed consent for their likenesses to be published.

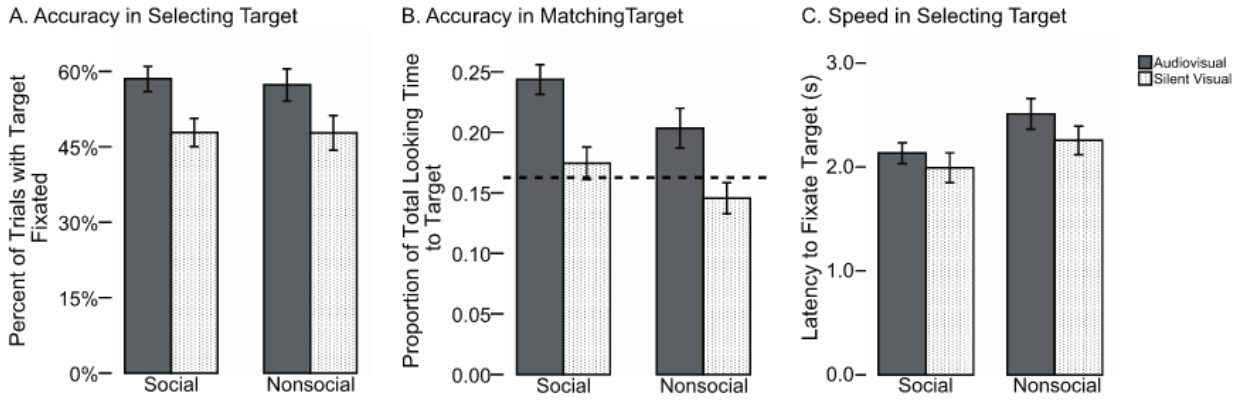

Figure 2. Bar graphs representing (A) accuracy in selecting the target (mean percentage of trials on which the target was fixated: PTTF), (B) accuracy in matching the target (mean proportion of total looking time to the target: PTLT), and (C) speed in selecting the target (mean latency to fixate target) across audiovisual and silent, visual trials and for social and nonsocial events. Dashed line in 2B represents chance value (0.167) for accuracy in matching. Bars represent the standard errors of the mean.