DOI: 10.1111/infa.12533

INFANCY THE OFFICIAL JOURNAL OF THE INTERNATIONAL CONGRESS OF INFANT STUDIES WILEY

RESEARCH ARTICLE

Intersensory processing of faces and voices at 6 months predicts language outcomes at 18, 24, and 36 months of age

Elizabeth V. Edgar[1] | **James Torrence Todd**[2] | **Lorraine E. Bahrick**[2]

[1]Yale Child Study Center, Yale University School of Medicine, New Haven, Connecticut, USA

[2]Department of Psychology, Florida International University, Miami, Florida, USA

Correspondence
Elizabeth V. Edgar.
Email: elizabeth.edgar@yale.edu

Funding information
National Institute of Child Health and Human Development, Grant/Award Numbers: R01 HD053776, R01HD094803

Abstract

Intersensory processing of social events (e.g., matching sights and sounds of audiovisual speech) is a critical foundation for language development. Two recently developed protocols, the Multisensory Attention Assessment Protocol (MAAP) and the Intersensory Processing Efficiency Protocol (IPEP), assess individual differences in intersensory processing at a sufficiently fine-grained level for predicting developmental outcomes. Recent research using the MAAP demonstrates 12-month intersensory processing of face-voice synchrony predicts language outcomes at 18- and 24-months, holding traditional predictors (parent language input, SES) constant. Here, we build on these findings testing younger infants using the IPEP, a more comprehensive, fine-grained index of intersensory processing. Using a longitudinal sample of 103 infants, we tested whether intersensory processing (speed, accuracy) of faces and voices at 3- and 6-months predicts language outcomes at 12-, 18-, 24-, and 36-months, holding traditional predictors constant. Results demonstrate intersensory processing of faces and voices at 6-months (but not 3-months) accounted for significant unique variance in language outcomes at 18-, 24-, and 36-months, beyond that of traditional predictors. Findings highlight the importance of intersensory processing of face-voice synchrony as a foundation for language development as early as 6-months and reveal that individual differences assessed by the IPEP predict language outcomes even 2.5-years later.

1 | INTRODUCTION

Parent language input is a well-established predictor of individual differences in child language development (Hart & Risley, 1995; Hoff, 2003; Huttenlocher et al., 1991; Rowe, 2008). Greater quantity (amount) and quality (diversity) of parent language input are associated with better child language outcomes (Hart & Risley, 1992; Huttenlocher et al., 1991; Weisleder & Fernald, 2013; Weizman & Snow, 2001). Socioeconomic status (SES) is also a well-known predictor of child language, with higher SES predicting greater quality of parent language input (Hart & Risley, 1995; Rowe, 2018), which in turn predicts increases in child vocabulary across age (Hoff, 2003). In contrast, there has been little research characterizing the role of individual differences in intersensory processing of audiovisual events (e.g., audiovisual speech) as a foundation of child language outcomes, despite agreement that it is an important early foundation for language development (Bahrick et al., 2020; Bahrick, Todd, et al., 2018; Edgar et al., 2022). Intersensory processing involves detecting redundancy across the senses (e.g., face-voice synchrony during speech) and provides a basis for perceiving audible and visible stimulation as a unitary event (e.g., audiovisual speech) and, in turn, serves as a foundation for further processing the unified event (e.g., prosodic and affective information, and object labeling during audiovisual speech; Bahrick et al., 2020).

Two recently developed measures, the Multisensory Attention Assessment Protocol (MAAP; Bahrick, Todd et al., 2018) and the Intersensory Processing Efficiency Protocol (IPEP; Bahrick, Soska, et al., 2018) now allow researchers to assess fine-grained individual differences in multisensory attention and intersensory processing in young infants in a context highly relevant for language acquisition, that of dynamic faces and voices during audiovisual speech. The MAAP assesses individual differences in three “multisensory attention skills”—sustaining attention (attention maintenance or overall interest in sights and sounds), shifting/disengaging attention (speed of moving attention to sights and sounds), and intersensory processing (matching synchronous sights and sounds)—for both audiovisual social (speech) and nonsocial (object) events. Using this measure, we recently found that intersensory processing (but not sustaining or shifting/disengaging attention) for social (but not nonsocial) events at 12 months of age was a strong predictor of child language outcomes. Intersensory processing predicted child speech production (assessed by direct observation of the child) and expressive vocabulary (on the MB-CDI, a parent-report form; Fenson et al., 2007; Jackson-Maldonado et al., 2003) at 18 and 24 months of age, even after controlling for parent language input and SES (Edgar et al., 2022). These findings replicate and extend our prior research with toddlers and young children (2- to 5-year-olds) demonstrating that intersensory processing of faces and voices predicts receptive and expressive language outcomes (Bahrick, Todd, et al., 2018). Together, these findings indicate that intersensory processing of faces and voices in infancy and early childhood provides an important foundation for language development. Further, by assessing individual differences in intersensory processing we can predict which children will benefit most from the parent language input and other language learning opportunities provided by their environment.

The present study builds on the study by Edgar et al. (2022), extending it to younger infants (3- and 6-month-olds), earlier and later language outcomes (12, 18, 24, and 36 months of age), and a more fine-grained measure of just intersensory processing (IPEP). The IPEP is an ideal measure for this purpose given that our prior study demonstrated that intersensory processing (but not other multisensory attention skills: attention maintenance or shifting/disengaging speed) was a strong predicter of language outcomes. Further, the present study extends our prior findings by investigating individual differences in both speed and accuracy of intersensory matching. Given that intersensory processing skills develop rapidly across the first 6 months of life (for reviews, see Bahrick et al., 2020; Bremner et al., 2012), might they also predict child language outcomes at 12, 18, 24, and 36 months? Similar to our prior study (Edgar et al., 2022), we investigated to what extent individual differences in intersensory processing of faces and voices during natural, synchronous, audiovisual speech would predict language outcomes given comparable levels of parent language input and SES.

1.1 | Intersensory processing of audiovisual speech: A foundation for child language development

Intersensory processing provides a fundamental basis for guiding infant selective attention and perceptual development (Bahrick et al., 2020; Bahrick & Lickliter, 2012; E. J. Gibson, 1969). It fosters selective attention to unitary events during multimodal stimulation (e.g., faces and voices of a person speaking) and helps infants filter out irrelevant stimulation from co-occurring events (e.g., a nearby conversation or activity). A skill that develops in early infancy, intersensory processing involves detecting intersensory redundancy (e.g., the synchronous co-occurrence of stimulation across two or more senses). Intersensory redundancy is highly salient and recruits attention to properties of events that are common across sense modalities (i.e., amodal information) including temporal synchrony, rhythm, intensity, and tempo. Most events provide multiple forms of amodal information (Bahrick, 2010; Bahrick et al., 2020; Bahrick & Todd, 2012). One property of amodal events, temporal synchrony (simultaneous changes in patterns of visual and acoustic stimulation, including auditory and visual onset, offset, duration, and common temporal patterning), is proposed to be the "glue" that binds stimulation across the senses (Bahrick & Lickliter, 2002; Lewkowicz, 2000a). Temporal synchrony is considered to be a global amodal property that facilitates the detection of other (nested) amodal properties including duration, rhythm, and tempo (Bahrick, 1992, 1994, 2001). Here, we focus on intersensory processing across the auditory and visual modalities.

In face-to-face interactions, the perceiver can both hear what is said and see the corresponding articulatory gestures (Kuhl & Meltzoff, 1982; Rosenblum, 2008; Stevenson et al., 2014). When a person speaks, they provide highly salient amodal information—their vocalizations and mouth movements are spatially co-located, and share common rhythm, tempo, and intensity shifts (Gogate et al., 2001; Gogate & Hollich, 2010). Further, mapping a word onto an object is a multisensory activity involving linking a sound with a visual object or event. Parents intuitively use intersensory redundancy to help infants learn language, often labeling an object while holding and moving it in synchrony with its label (Gogate et al., 2000). The shared onset, offset, and duration of the simultaneous movement and naming recruits selective attention and provides salient amodal information that links the speech sounds with the object. Although the relation between an object and its name is arbitrary to a word-mapping novice, the intersensory redundancy provided by the simultaneous movement and labeling reduces the uncertainty about the word-referent relation (Gogate & Hollich, 2010). We have proposed that better intersensory processing skills promote more accurate and efficient processing of audiovisual speech events, allowing infants to take greater advantage of parent language input and language learning opportunities such as word mapping (Bahrick et al., 2020; Edgar et al., 2022).

1.2 | Group-level differences in intersensory processing of audiovisual speech

Infant intersensory processing has been studied extensively with methods appropriate for group-level analyses including the intermodal preference (Bahrick, 1983, 1988; Lewkowicz, 1992; Spelke, 1976) and habituation methods (Bahrick & Lickliter, 2004; Bahrick & Pickens, 1988; Caron et al., 1988; Lewkowicz, 2000b, 2003; Walker-Andrews & Grolnick, 1983). These methods have been used to assess intersensory processing skills for groups of infants at specific ages. Studies using these methods have revealed that infants can match faces and voices on the basis of a wide range of amodal properties. For example, newborns can detect face-voice synchrony in point-light displays of a woman speaking (Guellaï et al., 2016) and in nonhuman primate species on the basis of temporal synchrony (Lewkowicz, 2010). Infants from 2 to 4 months of age can match vowel sounds with the corresponding shape of lip movements on the basis of spectral information in the vowel sounds (Kuhl & Meltzoff, 1982; Patterson & Werker, 1999, 2003). By 4–7 months of age infants match faces and voices on the basis of the speaker's affect (Soken & Pick, 1992; Vaillant-Molina et al., 2013; Walker, 1982), gender (Richoz et al., 2017; Walker-Andrews et al., 1991), and age (Bahrick et al., 1998). Findings also demonstrate that infants can perceive amodal properties provided by nonsocial events such as objects striking a surface, including temporal synchrony, rhythm, tempo, as well as temporal microstructure, including object substance and composition (see Bahrick, 1983, 1987, 1988; Bahrick et al., 2002; Bahrick & Lickliter, 2000, 2004; Lewkowicz, 1992; Lewkowicz & Marcovitch, 2006). Thus, group-level studies assessing accuracy of intersensory processing reveal that infants can detect audiovisual synchrony and match faces and voices under a variety of conditions across the first half year of life. In contrast, speed of intersensory processing (how quickly infants find the synchronous audiovisual event) has received virtually no research focus in infants, although it has been assessed in adults (for a review, see Fiebelkorn et al., 2012).

Studies using these methods designed for group-level analyses have indicated that intersensory processing in infancy serves as a foundation for language development. For example, studies have demonstrated that synchronous, but not asynchronous, object movement and labeling promotes object-label matching in infants and toddlers (Gogate & Bahrick, 1998; Gogate et al., 2006; Jesse & Johnson, 2016). These methods, however, are not designed to provide scores for individual infants, have no established psychometric properties, and are thus not appropriate for predicting outcomes or assessing change across development.

1.3 | Individual differences in intersensory processing of audiovisual speech

In contrast, individual difference measures assessing the skills of individual infants relative to one another can address the extent to which intersensory processing in infancy predicts individual differences in outcomes such as language, social, or cognitive functioning. In this approach, unlike the group-level approach, demonstrating group-level evidence of intersensory processing is not a goal. Rather, a main strategy is to test participants at an age that will reveal meaningful variability across individual participants in intersensory processing skills and determine if this variability predicts variability in outcomes (e.g., language). Assessing infants at an age when they are just learning a skill can reveal meaningful variability between participants (where some participants are successful and others not, with no overall group-level success; for a statistical argument, see Jaccard & Becker, 2009) and may be optimal for predicting developmental outcomes (e.g., see Edgar et al., 2022; Marchman & Fernald, 2008; Rose & Feldman, 1987).

Research has found that greater preferences for a sound-matched speech event (e.g., woman articulating an /a/ lip movement paired with an /a/ soundtrack) over a sound-mismatched event (/a/ lip movement paired with /o/ soundtrack) at 6 months predicted larger vocabulary size (assessed via parent-report) at 12 months (Altvater-Mackensen & Grossmann, 2015; also see Altvater-Mackensen et al., 2016). However, it is unclear whether these findings of intersensory processing of spectral information in single-syllable speech stimuli would extend to intersensory processing of face-voice synchrony in fluid speech, or whether the findings would extend to language outcomes at older ages or to more fine-grained measures of language derived from direct observations of the child (rather than parent-report). Further, individual differences in infant attention to the mouth (relative to the eyes) of a person speaking single syllables predicts concurrent and prospective vocabulary size (Chawarska et al., 2022; Morin-Lessard et al., 2019; Tenenbaum et al., 2015; Tsang et al., 2018; Young et al., 2009; but see Hillairet de Boisferon et al., 2018). The mouth of a person speaking provides highly salient amodal information (presumably more than do the eyes), and thus attention to the mouth relative to the eyes of a person speaking may reflect intersensory processing of audiovisual speech (although this premise has yet to be tested empirically).

1.4 | The present study

Within the first 6 months of life, infants learn to efficiently locate the source of a sound, in both social and nonsocial events, while filtering out other concurrent auditory and visual stimulation (e.g., Bahrick, 1983; Kuhl & Meltzoff, 1982; Lewkowicz, 1992; Spelke, 1976). Given that assessing infants at an age when they are just learning a skill can reveal important variability between participants, which is optimal for predicting developmental outcomes, we focused on the earliest age (3 or 6 months) at which intersensory processing would predict language outcomes, while controlling for other well-established predictors. We addressed this question using the IPEP. It assesses speed and accuracy of intersensory processing in the context of six concurrent, dynamic social and nonsocial events. Infants must detect the sound-synchronous target event from among five competing visual distractor events that are asynchronous with the soundtrack, simulating the "noisiness" of picking out a speaker from a crowd. Individual scores are derived for each measure across a number of trials for each infant, making it fine-grained enough to reliably predict outcomes, with a relatively stable mean. Finally, the IPEP does not require verbal responses or language comprehension, making it appropriate for use with preverbal infants and children.

Here, we demonstrate the viability of the IPEP as an index of individual differences in infant intersensory processing for evaluating developmental relations with language outcomes. Unlike the MAAP, which indexes the accuracy of intersensory matching using a method similar to a traditional two-screen intermodal matching method, the IPEP indexes both accuracy and speed of intersensory matching of a single synchronous target event in the presence of five asynchronous distractor events. Like the MAAP, it assesses intersensory matching for both social and nonsocial events. However, given that audiovisual speech events provide the most relevant context for language learning and that both our prior studies (Bahrick, Todd et al., 2018; Edgar et al., 2022) found that attention to social, but not to nonsocial events, predicted language outcomes, in this study we again focused on intersensory processing of faces and voices rather than object events.

1.4.1 | Predictions

Using the IPEP, the present study examines the unique contribution of intersensory processing at 3 and 6 months in predicting child language outcomes at 12, 18, 24, and 36 months. We assessed the earliest age (3 or 6 months) intersensory processing would predict language outcomes and the earliest age (12, 18, 24, 36 months) language outcomes could be predicted by intersensory processing skills. By 6 months, we expected results to parallel and extend those of our previous study (Edgar et al., 2022) assessing intersensory processing of social events with the MAAP. That is, we hypothesized that greater intersensory processing (measures of speed and accuracy) at 6 months, greater parent language input (quantity and quality) at 6 months, and higher SES (maternal education) would be correlated with greater child speech production (quantity and quality) and vocabulary size (receptive and expressive) at 18, 24, and 36 months. Further, we hypothesized that in our regression models, intersensory processing of social events at 6 months of age would predict significant unique variance in child language outcomes at 18, 24, and 36 months, while controlling for parent language input and SES. In contrast, we did not have strong predictions about the ability of intersensory processing at 3 months to predict language outcomes, given limited attention skills and less experience interacting with caregivers. Nor did we have strong predictions about whether we would see evidence of intersensory processing at either age predicting language outcomes as early as 12 months, given limited language at this age and our limited number of language assessments (MB-CDI receptive and expressive vocabulary, but no measures of child speech production).

2 | METHOD

2.1 | Participants

One-hundred and four infants participated as part of a larger ongoing longitudinal study on the development of multisensory attention skills and language, cognitive, and social outcomes entitled "Development of Intermodal Perception of Social and Nonsocial Events". Written informed consent was obtained from a parent or guardian for each child before any assessment or data collection to be in accordance to guidelines laid down in the Declaration of Helsinki. All procedures involving human participants received IRB approval from the Social and Behavioral Review Board of Florida International University (IRB-13-0448-CR06). The final sample consisted of a total of $N = 103$ infants (one infant only participated at 3 and 6 months, had no language outcome measures at older ages, and thus was excluded from the analyses). All infants were born between 38 weeks, 1 day and 41 weeks, 7 days gestational age, weighed 5 lbs. or more at birth, had APGAR scores of 9 or 10, and had no vision or hearing problems per parental report. Infants were assessed at 3, 6, 12, 18, 24, and 36 months. Demographic information for the sample can be found in Table 1. For a summary of the assessments administered at each age and dependent variables, see Table 2.

2.2 | Child intersensory processing measures: IPEP

2.2.1 | Stimulus events

The IPEP consists of 48 8-s trials with 24 social and 24 nonsocial trials presented in four alternating blocks of 12 trials each. This updated version of the IPEP was modified and refined based on stimuli and procedures used in Bahrick, Soska et al. (2018) including filming new social events (see Figure 1), increasing trial length from 6- to 8-s to be more appropriate for younger infants, and making social/nonsocial trial blocks a within participants factor. As before, trials consisted of a 2 (rows) x 3 (columns) grid of 6 dynamic social or nonsocial events. The entire grid was 67.3 × 38.1 cm

TABLE 1 Demographic information for the sample ($N = 103$).

	N	**Percentage**
Gender		
Male	51	49.5%
Female	52	50.5%
Ethnicity		
Hispanic	66	64%
Non-Hispanic	34	33%
Did not disclose	3	2.9%
Race		
White/European-American	69	67%
Black/African-American	16	15.5%
Asian/Pacific Islander	2	2%
More than 1 race	9	8.7%
Did not disclose	7	6.8%
Maternal education		
High school or equivalent	14	13.6%
Some college	16	15.5%
Associate's degree	15	14.6%
Bachelor's degree	26	25.2%
Master's degree or higher	26	25.2%
Did not disclose	6	5.8%
Home language		
English	63	61.2%
Spanish	30	29.1%
Both English and Spanish	1	1%
Did not disclose	5	4.9%
Age	***M***	**SD**
3-month visit	3.03	0.18
6-month visit	5.97	0.20
12-month visit	12.05	0.25
18-month visit	18.05	0.42
24-month visit	24.19	0.37
36-month visit	36.13	0.64

(51.3° visual angle), and each square of the grid covered 20.3 × 16.5 cm (16.5° visual angle). The social events depict six women, each telling a different story using infant-directed speech. Nonsocial events depict six wooden objects (single objects or clusters of objects) being dropped on a surface in erratic temporal patterns. On each trial, the natural soundtrack is synchronized with the movements of one event while the movements of the other five events are asynchronous with the soundtrack. Thus, the infant's task is to visually fixate the sound-synchronous speaker (target event) amidst the five asynchronous distractors on each trial. For an example video, please visit https://nyu.databrary.org/volume/336. A smiley face is presented zooming in and out for two-seconds between each trial

TABLE 2 Protocols, assessments used to index each construct, ages administered, and dependent variables.

Construct	Protocol/assessment	Ages	Dependent variables
Infant intersensory matching	Intersensory Processing Efficiency Protocol (IPEP)	3 and 6 months	Accuracy Speed
Parent language input	Parent-child interaction (PCI)	6 months	Quantity- tokens Quality- types
Child speech production	Parent-child interaction (PCI)	18, 24, 36 months	Quantity- tokens Quality- types
Child vocabulary size	Mac-Arthur Bates Communicative Development Inventory (MB-CDI)	12, 18, 24 months	Expressive vocabulary Receptive vocabulary
	Expressive Vocabulary Test (EVT)	36 months	Expressive vocabulary
	Peabody Picture Vocabulary Test (PPVT)	36 months	Receptive vocabulary

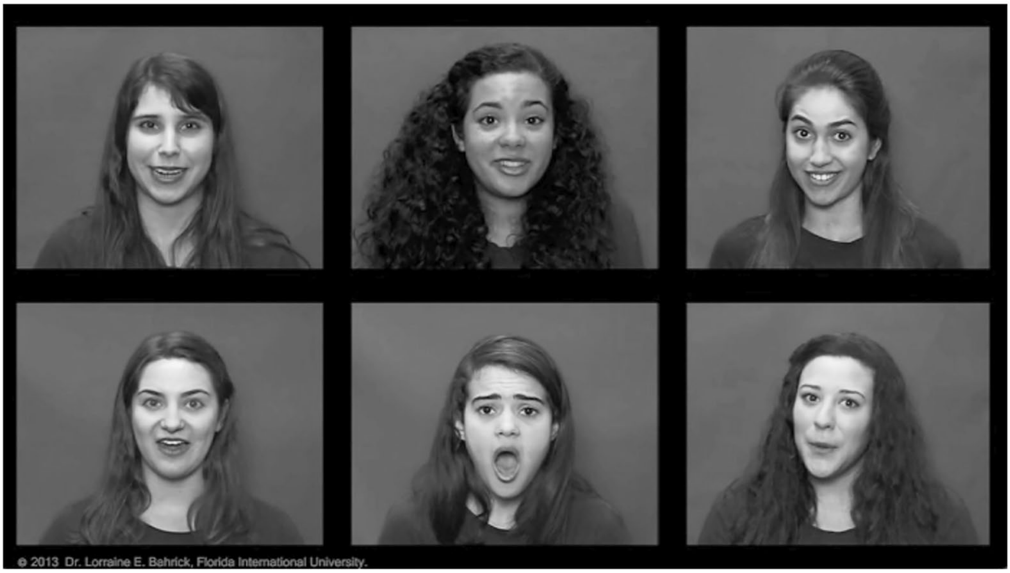
© 2013 Dr. Lorraine E. Bahrick, Florida International University.

FIGURE 1 Static image of the dynamic audiovisual social events from the Intersensory Processing Efficiency Protocol (IPEP). On each trial, all six women are shown speaking while the natural and synchronous soundtrack to only one of them is heard.

to attract the infant's attention to the center of the screen. Six different types of smiley faces, each of different primary colors, were presented in a pseudorandom order across trials.

2.2.2 | Procedure

A 119.4-centimeter widescreen monitor (NEC Multisync PV61) was used to present the IPEP and a Tobii X120 eye-tracker was used to record gaze fixations. Infants were seated on their caregiver's lap approximately 70 cm in front of the monitor, and 60 cm in front of the Tobii eye-tracker. The eye-tracker, located directly under the monitor, was tilted upward, 20°, towards the child's eyes. An experimenter, seated behind the child, presented the stimulus events to the monitor using Tobii Studio

(Version 3) from a computer (Mac Pro Computer with 16 GB of RAM, a 3.33-GHz processor, and a 400-MHz graphics card). Caregivers wore black-out glasses to ensure they were unaware of the location of the sound-synchronous target event.

The experimenter viewed a live recording from a video camera (SONY FDR-AX33) placed facing the infant to ensure the infant was seated in an optimal position for eye-tracking calibration and for viewing the stimuli. Tobii Studio's "Infant" 5-point calibration procedure was used to calibrate the infrared corneal reflection-to-pupil tracking system for each infant. The experimenter calibrated the infant's eye-gaze to five points on the widescreen monitor for accurate calculation of infant visual fixations during the procedure.

The 24 social and 24 nonsocial trials were arranged into four alternating blocks of 12 trials each (social, nonsocial, social, nonsocial, or vice versa, counterbalanced across participants). Social and nonsocial events in the IPEP are presented in separate blocks and designed to be analyzed separately depending on the research focus of the study. The present study focuses on social events, given the importance of audiovisual speech perception for predicting language outcomes, as well as results of our previous studies using the MAAP in which performance on social (but not nonsocial) trials predicted language outcomes (e.g., Edgar et al., 2022).

Infant eye gaze was sampled at 120 Hz by the Tobii X120 system. The number of usable trials at 3 months ranged from 2 to 48 trials with an average of 37.88 (*SD* = 11.89 trials) out of 48 trials, and at 6 months, it ranged from 15 to 48 trials with an average of 43.27 (*SD* = 7.27 trials) out of 48 trials. Trials in which infants were inattentive (less than 250 ms looking to the screen) were excluded from analyses. Further details regarding eye-tracking parameters and data processing are presented in the Supporting Information S1 (p. 1).

2.2.3 | IPEP measures

The IPEP provides three measures of intersensory processing: accuracy of intersensory matching, speed of intersensory matching, and frequency of target selection. In the present study, we focus on just two of these measures: accuracy of intersensory matching (duration of looking to the sound synchronous target) and speed of intersensory matching (reaction time to fixate the target). Frequency of target selection (proportion of total trials on which the infant fixated the sound synchronous target event) was not significantly correlated with any of our outcome variables and was thus excluded from subsequent analyses. Although this measure reflects fixation of the sound synchronous target event, it does not index intersensory matching because it does not indicate whether the target event was fixated more often than the other asynchronous events during the trial. Accuracy of intersensory matching (PTLT; proportion of total looking time to the sound-synchronous "target" event) is the traditional measure used in studies of intersensory processing and assesses how long the infant fixates the sound-synchronous visual event. Greater looking to the sound-synchronous event provides an opportunity for longer and deeper processing of the multimodal event. PTLT was calculated by dividing the looking time to the AOI depicting the sound-synchronous target event by the total looking time to all six AOIs. PTLTs greater than 0.167 (chance) indicate a preference for the sound-synchronous target event. Speed of intersensory matching (RT) assesses how quickly infants visually fixate the sound-synchronous event. Faster speeds in fixating the target event reflect faster intersensory matching and more time for processing the multimodal event. RT was calculated as the latency from trial onset to produce a fixation (of at least 50 ms) to the sound-synchronous event. PTLTs and latency scores were calculated for each trial and then averaged across all trials within each condition (social, nonsocial).

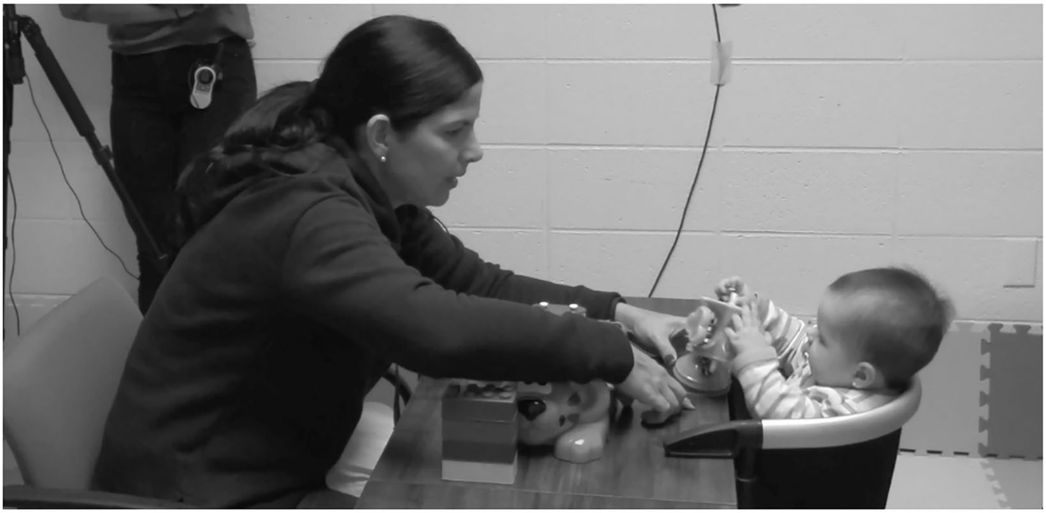

FIGURE 2 Parents and children received three age-appropriate toys during the Parent-Child Interaction (PCI). Each interaction was video recorded by three cameras placed in corners of the playroom (see Edgar et al., 2022, for details). Above is a side view of a parent seated across from a 6-month-old infant playing with one of the three toys provided.

2.3 | Parent language input and child language production measures

Parents and children participated in an 8-min ($M = 8.15$ min, range = 3.30 to 12.28) semi-structured Parent-Child Interaction (PCI) at 6, 12, 18, 24, and 36 months (a somewhat different task was used at 3 months and parent language input was not transcribed). In a lab playroom, the parent and child were seated facing each other at a table (40 × 28 in) in the center of the room (see Figure 2). At 6, 12, and 18 months, children sat in a seat attached to the edge of the table, and at 24 and 36 months, they sat in a booster seat attached to a chair.

At each age, parent and child speech during the PCI was transcribed by trained research assistants who watched the video recordings. Transcription units were words. A second trained research assistant checked the original transcriptions to establish reliability. Any disagreements between the primary transcriber and the secondary transcriber were decided by a third research assistant, who was not aware of the topic of disagreement. The Child Language Data Exchange Systems (CHILDES; MacWhinney, 2000) FREQ program was used to calculate the quantity (tokens; total number of words spoken) and quality (types; total number of different, or unique, words spoken) of parent language input and child language production. To equate across PCIs of different durations, a per-minute ratio was calculated by dividing the number of types (or tokens) by the duration of the interaction. Only speech directed to the child was included in type and token calculations (e.g., parents rarely spoke to the experimenter, but this speech was not transcribed).

2.4 | Child vocabulary measures

Parents completed the age-appropriate Mac-Arthur Bates Communicative Development Inventories (MB-CDI) at 12, 18, and 24 months in either English (Fenson et al., 2007) or Spanish (Jackson-Maldonado et al., 2003) or both, depending on parental report of the child's primary language (for details, see Supplement, pp. 1–2). At 12 and 18 months, they completed the *Words and Gestures* form (appropriate for ages 8–18 months), which assesses both receptive and expressive vocabulary. At 24 months, they completed the *Words and Sentences* form (appropriate for ages 16–30 months) which assesses just expressive vocabulary (and not receptive vocabulary). At 36 months, children received the Peabody Picture Vocabulary Test—4th Edition (PPVT; Dunn & Dunn, 2007) to assess the child's receptive vocabulary size and the Expressive Vocabulary Test—2nd Edition (EVT; Williams, 2007) to assess the child's expressive vocabulary size (there is no age-appropriate MB-CDI at 36 months of age). In contrast with the MB-CDI, the PPVT and EVT involve direct observations of the child.

3 | RESULTS

3.1 | Data analysis overview

The present study examined the extent to which intersensory matching (both speed and accuracy) of social events at 3 and 6 months predicted child language outcomes at 12, 18, 24, and 36 months, while holding constant other well-known predictors of child language, including parent language input at 6 months (both quantity and quality) and SES. We first examined intersensory matching accuracy at both the group- and participant-level to determine whether infants showed significant face-voice matching at 3 and 6 months of age. We then conducted correlations for the 3- and 6-month predictors (speed and accuracy of intersensory matching, quantity and quality of parent language input, and maternal education) with child language outcomes, including child quantity and quality of speech at 18, 24, and 36 months, expressive vocabulary at 12, 18, 24, and 36 months, and receptive vocabulary at 12, 18, and 36 months (measures of receptive vocabulary are not included in the *Words and Sentences* form MB-CDI appropriate for 24 months). Last, multiple regression models were conducted to address the main research questions. They each included five predictors of language outcomes: accuracy of intersensory matching, speed of intersensory matching, quality (types) of parent language input, quantity (tokens) of parent language input, and maternal education. With a sample size of $N = 103$, there is sufficient power for multiple regression analyses to detect a non-zero path coefficient that accounts for 6% unique variance (assuming a β of 0.80, a two-tailed p-value of 0.05, five predictors, and an R^2 of 0.30). Multicollinearity among predictors was not an issue (see Table S1 for variance inflation factors).

3.1.1 | Missing data

Robust Full Information Maximum Likelihood (FIML) estimation was used for all analyses in MPlus (Version 1.8.6). Missing data ranged from 6.7% (maternal education) to 51% (MB-CDI receptive and expressive vocabulary; see Table 3). To ensure that data were not systematically missing or missing not at random (MNAR or non-ignorable missingness; Rubin, 1976), we conducted missing value analyses using various techniques (e.g., t-tests, logistic regression, Little's MCAR test). T-tests and logistic regressions revealed that missingness was not related to any of the main predictors or outcomes. From these analyses, we concluded that the data were missing at random (MAR; Rubin, 1976), supporting the use of FIML.

3.1.2 | Secondary analyses

Secondary analyses were conducted to assess the influence of language spoken at home, gender, race, and ethnicity as covariates in predicting child language outcomes. Overall, their inclusion did not

TABLE 3 Means (M), standard deviations (SD), sample sizes (N), and percentages of missing data for 3- and 6-month intersensory matching (both speed and accuracy) for social events, and parent language input (both quantity and quality), as well as 12-, 18-, 24-, and 36-month child language outcomes.

	M	SD	*N*	Missing
3-month intersensory matching				
Accuracy	0.175	0.08	89	13.6%
Speed	2.72	1.03	89	13.6%
6-month intersensory matching				
Accuracy	0.165	0.04	90	13.5%
Speed	2.60	0.75	90	13.5%
6-month parent language input				
Quality (types)	12.26	5.32	84	19.2%
Quantity (tokens)	40.56	19.79	84	19.2%
12-month child language outcomes				
Receptive vocabulary (MB-CDI)	83.36	74.13	66	36%
Expressive vocabulary (MB-CDI)	9.65	12.76	66	36%
18-month child language outcomes				
Child speech quality (types)	0.65	0.75	76	26.9%
Child speech quantity (tokens)	1.50	1.80	76	26.9%
Receptive vocabulary (MB-CDI)	231.67	148.90	51	51%
Expressive vocabulary (MB-CDI)	61.75	77.93	51	51%
24-month child language outcomes				
Child speech quality (types)	2.70	2.09	70	32.7%
Child speech quantity (tokens)	6.13	5.25	70	32.7%
Expressive vocabulary (MB-CDI)	275.37	179.99	51	51%
36-month child language outcomes				
Child speech quality (types)	6.18	3.60	76	26.9%
Child speech quantity (tokens)	16.08	12.81	76	26.9%
Receptive vocabulary (PPVT)	108.85	15.68	71	31.7%
Expressive vocabulary (EVT)	106.85	15.34	67	35.6%

change the strength of the main predictors in predicting the child language outcome measures (for details, see Supplement, pp. 2–5). Thus, the present study did not include home language, gender, race, or ethnicity as covariates in the main analyses. Secondary analyses were also performed and included in the manuscript to assess the significance of intersensory matching at both the group and individual participant levels (see Intersensory Matching of Social Events: Group-Level and Individual-Participant Analyses: Secondary Analyses).

3.2 | Intersensory matching of social events: Correlational analyses

Descriptive statistics for 3- and 6-month intersensory matching (speed and accuracy) of social events, 6-month parent language input (quantity and quality), and child language outcomes at 12, 18, 24, and 36 months are displayed in Table 3 and correlations among these variables are displayed in Table 4.

TABLE 4 Correlations among predictors (accuracy and speed of intersensory matching of social events at 3 and 6 months, quantity and quality of parent language input at 6 months, and maternal education) and child language outcomes at 12, 18, 24, and 36 months.

	12-Month child language outcomes	
Predictors	**Receptive**	**Expressive**
3-Month intersensory matching		
Accuracy	0.09	−0.02
Speed	−0.02	0.07
6-Month intersensory matching		
Accuracy	0.05	0.38***
Speed	0.07	0.13
6-Month parent language input		
Quality (types)	−0.09	0.08
Quantity (tokens)	0.18	−0.09
Maternal education	0.21*^{f}	−0.26*

	18-Month child language outcomes			
Predictors	**Types**	**Tokens**	**Receptive**	**Expressive**
3-Month intersensory matching				
Accuracy	0.24*^{f}	0.27**	0.21*^{f}	0.23*^{f}
Speed	−0.13	−0.07	0.21*^{f}	−0.08
6-Month intersensory matching				
Accuracy	0.37***	0.33***	−0.01	0.33***
Speed	0.16	0.12	−0.11	0.10
6-Month parent language input				
Quality (types)	0.12	0.02	0.04	0.07
Quantity (tokens)	0.19*^{f}	0.02	0.03	0.12
Maternal education	0.27**	0.21*^{f}	0.07	0.14

	24-Month child language outcomes		
Predictors	**Types**	**Tokens**	**Expressive**
3-Month intersensory matching			
Accuracy	−0.11	−0.08	0.07
Speed	−0.24*	−0.27**	−0.13
6-Month intersensory matching			
Accuracy	0.35***	0.40***	0.25**
Speed	0.10	0.07	−0.01
6-Month parent language input			
Quality (types)	0.28**	0.21*^{f}	0.22*^{f}
Quantity (tokens)	0.24*	0.21*^{f}	0.23*^{f}
Maternal education	0.46***	0.34***	0.26**

(Continues)

TABLE 4 (Continued)

	36-Month child language outcomes			
Predictors	Types	Tokens	Receptive	Expressive
3-Month intersensory matching				
Accuracy	−0.13	−0.18	0.08	0.14
Speed	0.11	0.09	−0.18	−0.12
6-Month intersensory matching				
Accuracy	0.15	0.05	0.37***	0.25**
Speed	0.02	0.03	0.03	0.09
6-Month parent language input				
Quality (types)	0.15	0.09	0.22*^{f}	0.24*^{f}
Quantity (tokens)	0.12	0.10	0.18	0.10
Maternal education	0.23*^{f}	0.09	0.40***	0.46***

Note: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$, and **f* $p < 0.05$ but did not meet significance cut off ($p = 0.025$ for 12 months, $p = 0.0125$ for 18 and 36 months, $p = 0.0167$ for 24 months) when controlling for familywise error.

We first calculated Pearson correlation coefficients using FIML[1] and correcting for familywise error rate.[2] Correlations were conducted between our main predictor variables—accuracy of intersensory matching for social events at 3 and 6 months, speed of intersensory matching at 3 and 6 months, quality of parent language input at 6 months, quantity of parent language input at 6 months, and maternal education—as well as our language outcome variables—quality of child speech production, quantity of child speech production, receptive vocabulary, and expressive vocabulary—at 12, 18, 24, and 36 months of age (there was no measure of receptive vocabulary size at 24 months). Several novel findings emerged.

Results of our correlational analyses (see Table 4) revealed that accuracy of intersensory matching of social events at 6 months predicted a wide variety of language outcomes across 12, 18, 24, and 36 months of age (r-range: 0.25-0.40, $ps < 0.01$), including quality and quantity of child speech production at 18 and 24 months and expressive vocabulary size at all ages. In total, intersensory matching at 6 months predicted 9 of 13 language outcomes. In contrast, accuracy of intersensory matching of social events at 3 months was a weak predictor of language outcomes, predicting only 1 of 13 language outcomes. Overall, greater quality and quantity of parent language input at 6 months (r-range: 0.21–0.28, $ps < 0.01$), as well as greater maternal education (r-range: 0.23–0.46, $ps < 0.01$), predicted better child language outcomes, particularly at 24 and 36 months, consistent with prior findings. These correlational analyses informed the construction of our multiple regression models (see Supplement, pp. 7–9, for more details). All models were constructed using intersensory matching at 6 months of age given that it was a moderately strong predictor of 9 of the 13 language outcomes.

[1] All correlations conducted using FIML were compared to traditional bivariate pairwise Pearson's r correlations (excluding participants with missing data) to ensure that findings derived from FIML were similar to the general pattern of findings from participants with complete data. All findings using FIML paralleled those of the bivariate pairwise correlations, with similar magnitudes and directions.

[2] At 18 and 36 months, there were four child language outcomes (child speech production: quantity and quality; child vocabulary size: receptive and expressive) and thus we used a familywise significance level of $p < 0.0125$ (0.05/4; two-tailed) to evaluate results. At 24 months, there were three child language outcomes (child speech production: quantity and quality; child expressive vocabulary size) and thus we used a familywise significance level of $p < 0.0167$ (0.05/3; two-tailed) to evaluate results.

3.3 | Intersensory matching of social events: Multiple regression analyses

Our primary analyses consisted of multiple regressions assessing the role of intersensory processing of social events on language outcomes in the context of other predictors. Findings from our correlational analyses revealed that accuracy of intersensory matching of social events at 6 months, but not at 3 months, predicted a variety of child language outcomes at 12, 18, 24, and 36 months. Thus, we only included intersensory matching at 6 months, but not at 3 months, in our primary regression models. In contrast with regressions, findings from bivariate correlations do not reveal relative importance or aggregate effects of multiple predictors, including SES, parent language input, and intersensory matching on language outcomes. Our regression analyses assessed whether accuracy of intersensory matching of social events at 6 months would remain a significant predictor of child language outcomes when holding constant parent language input and SES. They assessed the relative predictive power (unique variance) of each of these variables in predicting language outcomes when the others were controlled, as well as the aggregate effects (total variance accounted for) of all of these variables together in predicting language outcomes. To address these key research questions, our main analyses consisted of multiple regressions conducted using FIML to assess the role of 6-month accuracy of intersensory matching of social events, SES, and parent language input as predictors of child language outcomes. We conducted regression analyses for each of the child language outcomes including quality of child speech (18, 24, and 36 months), quantity of child speech (18, 24, and 36 months), receptive vocabulary (12, 18, and 36 months; there was no measure of receptive vocabulary at 24 months), and expressive vocabulary (12, 18, 24, and 36 months). We also included speed of intersensory matching in all of our models given that it has typically not been assessed along with accuracy of matching nor have studies previously assessed it as a predictor of language outcomes. We controlled for both quantity and quality of parent language input at 6 months and maternal education to examine the extent to which intersensory matching of social events predicted language outcomes at 12, 18, 24, and 36 months, over and above that of these well-established predictors at 6 months of age.

For each of the outcome variables, we conducted five multiple regression models to assess the amount of unique variance (ΔR^2) attributable to each predictor in predicting the outcome variable. The unique variance attributable to a given predictor (and not attributable to any other predictor in the model) is the change in R^2 when that predictor is entered last in the regression model (i.e., holding all other predictors constant; Cohen et al., 2003). In other words, the unique variance of 0.08 is the proportion of variance attributable to 6-month accuracy of intersensory matching when predicting 18-month expressive vocabulary holding 6-month speed of intersensory matching, quantity of parent language input, quality of parent language input, and maternal education constant (see Table 5). To derive the unique variance for each predictor, each of the five predictors at 6 months (accuracy of intersensory matching, speed of intersensory matching, parent speech quality, parent speech quantity, and maternal education) was entered into the regression model in a different order (1st, 2nd, 3rd, 4th, 5th). For example, in Model 1 we derived the unique variance attributable to accuracy of intersensory matching in predicting an outcome by entering it last (i.e., holding constant all other predictors entered earlier: maternal education, quality of parent language input, quantity of parent language input, speed of intersensory matching, and so forth for Models 2 through 5; for details, see Tables S5–S8). The amount of total variance explained by all 5 predictors, as well as the unique variance explained by each predictor in predicting each of the language outcomes at each age are summarized in Table 5.

Remarkably, 6-month accuracy of intersensory matching of social events was a significant predictor and accounted for unique variance in eight child language outcomes at 18, 24, and 36 (but not 12) months

TABLE 5 Amount of unique variance accounted for by each predictor variable at 6 months (accuracy and speed of intersensory matching for social events, quantity and quality of parent language input, and maternal education) in predicting child language outcomes at 18, 24, and 36 months ($N = 103$).

	12-Month language outcomes	
Predictors	**Expressive**	**Receptive**
Total variance	0.15	0.09
Unique variance		
6-Month intersensory matching		
Accuracy	0.04	0.01
Speed	0.07*	0.01
6-Month parent language input		
Quantity	0.06*	0.02
Quality	0.06*	0.02
Maternal education	0.12**	0.06*

	18-Month language outcomes			
Predictors	**Quantity**	**Quality**	**Expressive**	**Receptive**
Total variance	0.18†	0.27**	0.16	0.02
Unique variance				
6-Month intersensory matching				
Accuracy	0.11**	0.13**	0.08*	0.00
Speed	0.03*	0.04*	0.04	0.02
6-Month parent language input				
Quantity	0.00	0.02	0.00	0.00
Quality	0.01	0.02	0.00	0.00
Maternal education	0.04	0.06*	0.01	0.01

	24-Month language outcomes		
Predictors	**Quantity**	**Quality**	**Expressive**
Total variance	0.29**	0.35***	0.15
Unique variance			
6-Month intersensory matching			
Accuracy	0.14***	0.10**	0.06*
Speed	0.02	0.00	0.00
6-Month parent language input			
Quantity	0.01	0.00	0.01
Quality	0.01	0.00	0.00
Maternal education	0.08*	0.14****	0.03

	36-Month language outcomes			
Predictors	**Quantity**	**Quality**	**Expressive**	**Receptive**
Total variance	0.02	0.08	0.32***	0.30**
Unique variance				

TABLE 5 (Continued)

Predictors	36-Month language outcomes Quantity	Quality	Expressive	Receptive
6-Month intersensory matching				
Accuracy	0.00	0.03	0.08*	0.15***
Speed	0.00	0.00	0.03	0.00
6-Month parent language input				
Quantity	0.00	0.00	0.03	0.00
Quality	0.00	0.00	0.04*	0.00
Maternal education	0.00	0.04*	0.17**	0.13**

Note: ***$p < 0.001$, **$p < 0.01$, *$p < 0.05$.

of age (range: 6%–15%, *p*s < 0.05; see Table 5).[3] In contrast, 6-month speed of intersensory matching accounted for a smaller but significant amount of unique variance (range: 3%–4%, *p*s < 0.05) in child speech quality and quantity at just 18 months (but not at 24 or 36 months). Further, maternal education accounted for a significant amount of unique variance in 8 of the 13 child language outcomes, child speech quality at 18, 24, and 36 months, child speech quantity at 24 months, and expressive and receptive vocabulary at 12 and 36 months (range: 6%–17%, *p*s < 0.05). Six-month parent language input (both quantity and quality) accounted for a non-significant amount of unique variance in most child language outcomes (range: 0%–4%, *p*s > 0.05) with just three exceptions: 6-month parent language quality and quantity significantly predicted receptive vocabulary at 12 months, and 6-month parent language quality significantly predicted receptive vocabulary at 18 months (*p*s < 0.05). Details regarding the amount of unique variance attributed to each child language outcome by each predictor, as well as details quantifying relations between 6-month accuracy and speed of intersensory matching of social events and child language outcomes can be found in the Supplement, pp. 9–13 and Tables S5–S8.

3.4 | Intersensory matching of social events: Group-level and individual-participant analyses: Secondary analyses

Although not central to our research questions, we examined intersensory matching of social events to assess if infants showed significant intersensory matching as a group and individually at 3 and 6 months of age. At the group-level, *t*-tests against the chance value of target fixation (0.167; the likelihood of fixating one of 6 events longer by chance) revealed that infants showed no evidence of intersensory matching at 3 (t (88) = 0.89, $p = 0.38$) or 6 (t (89) = −0.43, $p = 0.67$) months of age.[4] At the individual-level, we chose to use cumulative binomial probability tests for each participant because of the small sample size (less than or equal to 24 trials for each participant). Given that

[3] In all cases, intersensory matching of social events at 6 months of age remained a significant predictor even when controlling for the average number of AOIs fixated on the IPEP, along with the other main predictors (speed of intersensory matching, quantity and quality of parent language input, maternal education).

[4] Note that from a statistical perspective, group-level means are irrelevant to correlational analyses. For example, it is well known that the correlation between two variables, X1 and X2, is unaffected by adding a constant to either or both of the variables. If we add a constant of *k* to each child's intersensory matching scores at 6 months, none of our correlations will change between it and language nor will the significance tests associated with those correlations (Jaccard & Becker, 2009).

infants completed up to 24 test trials, power analyses indicated that there was insufficient power to detect a significant effect with a single-sample *t*-test. Using cumulative binomial probability tests, we were able to assess the likelihood that an infant scored above the chance value of 0.167 on a significant number of trials according to tests of cumulative binomial probability (e.g., a score of 8 or more trials out of 24 is associated with a cumulative binomial probability of $p < 0.04$). Cumulative binomial probability was adjusted depending on the number of trials the child completed, (3 months: $M = 18.98$, $SD = 6.16$; 6 months: $M = 20.99$, $SD = 4.72$). At 3 months of age, out of the 89 infants who completed the IPEP, 21 (24%) showed intersensory matching greater than 0.167. At 6 months of age, out of the 90 infants who completed the IPEP, 32 (36%) showed intersensory matching greater than 0.167. Thus, the number of infants who show evidence of significant intersensory matching appears to increase with age from 3 to 6 months, and by 6 months, approximately a third of the infants showed evidence of significant intersensory matching at the individual participant level. Overall, there was a good amount of individual variability, providing a solid basis for predicting outcomes.

3.5 | Intersensory matching of nonsocial events: Supplemental analyses

To assess whether our finding that intersensory processing of social events predicts language outcomes is specific to social events or is also evident for nonsocial events, we conducted the same correlational analyses with nonsocial events and found few correlations with language. We found no evidence that accuracy or speed of intersensory matching of nonsocial events at 3 months was a significant predictor of child language outcomes (see Table S2). For accuracy of intersensory matching of nonsocial events at 6 months, there was only 1 significant correlation (out of 13): intersensory matching of nonsocial events predicted child expressive vocabulary at 18 months ($p < 0.05$). For speed of intersensory matching of nonsocial events at 6 months, there were 3 significant correlations (out of 13): infants who were slower to fixate the sound-synchronous event had better quantity of child speech production, receptive vocabulary, and expressive vocabulary at 36 months (*p*s < 0.001; similar to results from social trials).

We followed up on these significant correlations with regression analyses, controlling for parent language input (quality and quantity) and SES (maternal education). Neither accuracy nor speed of intersensory matching of nonsocial events at 6 months was a significant predictor of any child language outcome when holding constant parent language input and SES. Thus, replicating our prior findings (Bahrick, Todd et al., 2018; Edgar et al., 2022), accuracy of intersensory processing for social, but not nonsocial events, contributes to child language outcomes at 18, 24, and 36 months, holding parent language input and SES constant. For details on nonsocial analyses, see Supplement, pp. 5-7.

3.6 | Parent language input at older ages predicts language outcomes: Supplemental analyses

Given that parent language input (quantity and quality) at 6 months was a weak predictor of child language outcomes, might parent language input at older ages be a stronger predictor of child language outcomes when controlling for intersensory matching at 6 months? Correlations revealed that parent language input at 18, 24, and 36 (but not 12) months was a moderately strong predictor of child language outcomes (for details on correlational analyses, see Table S4). Therefore, we modeled our regression analyses after the prior analyses using 6-month intersensory processing speed and accuracy as predictors, but using parent language input at 18, 24, or 36 months (rather than at 6 months). When parent language input (quality and quantity) at older ages was substituted for 6-month parent language input in our multiple regression models, results indicated that parent language input at 24 months significantly predicted child language at 24 months, and parent language input at 36 months significantly predicted child language at 36 months (*p*s < 0.05) holding all other predictors constant. In contrast, parent language input at 18 months was not a predictor of child language at 18 months. For details on these supplemental regression analyses, see Supplement, pp. 14–16 and Tables S9–S12.

Importantly, accuracy of intersensory matching of social events at 6 months remained a moderately strong predictor of a variety of language outcomes, even after holding quantity and quality of parent language input at 18, 24, or 36 months constant. Thus, when children receive equal amounts of parent language input at 6, 18, 24, or 36 months, accuracy of intersensory matching of faces and voices at 6 months continues to explain a significant proportion of leftover variability in child language outcomes.

4 | DISCUSSION

In the present study, we examined the contribution of accuracy and speed of intersensory processing of social events (faces and voices) at 3 and 6 months of age as a predictor of child language outcomes at 12, 18, 24, and 36 months, along with well-established predictors including parent language input (quantity and quality) at 6 months and SES (maternal education). Results revealed that the accuracy of intersensory matching for social events at 6 months, but not at 3 months, predicted a variety of child language outcomes (including both receptive and expressive vocabulary as well as quantity and quality of child speech production) at 18, 24, and 36 (but not 12) months, over and above the contribution of parent language input and SES. Infants with higher levels of intersensory matching of faces and voices at 6 months had larger vocabularies and produced more speech at 18, 24, and 36 months. Also, the earliest we find evidence of intersensory matching assessed by the IPEP predicting language outcomes is 6 months of age. Further, although accuracy of intersensory matching at 6 months correlated with one language outcome at 12 months, it did not predict significant unique variance in child language outcomes until 18 months of age. Together, these results indicate that the accuracy of intersensory processing skills at 6 months (maintaining attention to the face of a person speaking amidst other dynamic faces) is a unique and important predictor of child language development across the second and third years of life. These findings are elaborated below.

4.1 | Intersensory matching of social events at 6 months, but not 3 months, predicts multiple child language outcomes at 12, 18, 24, and 36 months

Accuracy of intersensory matching of faces and voices predicted two types of language outcomes across the first 3 years of life, child vocabulary (using the MB-CDI, a parent-report measure) and child speech production (direct observation). Infants who show greater accuracy of intersensory processing at 6 months of age go on to show greater language outcomes at 18, 24, and 36 months (two and a half years later). Overall, these findings complement and extend our original findings with 12-month-olds (Edgar et al., 2022) in important ways. Intersensory matching of faces and voices at an earlier age, 6 months, not only predicted child language outcomes at 18 and 24 months (similar to our prior study) but also two and a half years later, at 36 months. These novel findings indicate that at 6 months, given equal amounts of parent language input (quantity and quality) and comparable SES, the accuracy of intersensory processing of faces and voices can predict which children will benefit most from language learning opportunities provided by parent language input. Infants who are more efficient at detecting face-voice synchrony likely have greater attentional resources available, enabling them to further process audiovisual speech events (Bahrick & Lickliter, 2014), and attend to other behaviors that take place in the context of language learning opportunities. This may include following eye-gaze direction and detecting gesture, facial and vocal affect, and prosody signaling communicative intent, all skills that are built on intersensory processing (Bahrick & Lickliter, 2012; Gogate & Hollich, 2010). Therefore, efficient processing of intersensory information may allow infants to abstract more relevant information from available input, cascading to other language learning skills such as joint attention (Morales et al., 1998), word-mapping (see Gogate & Bahrick, 1998; Gogate & Hollich, 2010), and statistical learning (Smith et al., 2014), and ultimately enhancing language development.

In contrast with accuracy of intersensory matching, speed of intersensory matching has rarely been investigated as a predictor of developmental outcomes. The present findings are among the first to demonstrate the role of speed of intersensory matching in predicting developmental outcomes. Speed of matching faces and voices at 6 months of age predicted unique variance in expressive vocabulary at 12 months and child speech production (quantity and quality) at 18 months over and above other predictors. These findings highlight the importance of speed of intersensory matching for predicting later language outcomes. Although speed of intersensory matching was a weaker predictor of language outcomes than accuracy of intersensory matching, it may be that speed of matching becomes a stronger predictor later in development when improvements in accuracy have plateaued. Future research will explore this topic.

In contrast with data from 6-month-olds, speed and accuracy of intersensory matching of faces and voices at 3 months of age were only weakly related to child language outcomes at 18, 24, and 36 months. Accuracy of intersensory matching at 3 months predicted only the quantity of child speech production at 18 months of age. Speed of intersensory matching at 3 months did not predict child language outcomes at any age. Although intersensory processing skills develop rapidly across the first 6 months of life (e.g., Bahrick, 1983; Kuhl & Meltzoff, 1982; Lewkowicz, 1992; Spelke, 1976), relations between intersensory processing of faces and voices at 3 months and language outcomes are not yet evident (above and beyond the contribution of parent language input at 6 months, and SES) as assessed by the IPEP in the present study. The lack of significant findings at 3 months may be due to task difficulty and task demands. Detecting face-voice synchrony across six different speakers within an 8 s trial may be too difficult for many (but not some infants, for details, see below "Secondary Analyses: Intersensory Matching of Social Events at the Group- vs. Individual-Level") infants of this age. Younger infants scan more slowly and likely detect fewer areas of the screen than older infants. Individual differences predict outcomes best when task demands are optimally matched to skills of the perceiver (Bahrick et al., 2010; Edgar et al., 2022).

Overall, convergent findings across two different protocols, the MAAP and IPEP, demonstrate the important role of infant intersensory processing of faces and voices in later language outcomes (Bahrick, Todd et al., 2018; Edgar et al., 2022). Both protocols assess accuracy of intersensory processing of social and nonsocial events. The MAAP assesses intersensory matching (along with sustained attention and shifting/disengaging attention) to two side-by side audiovisual events, similar to a traditional intermodal matching paradigm, in the presence and absence of competing stimulation from an irrelevant central distractor event. In contrast, the IPEP is a more fine-grained measure of just intersensory processing that assesses both speed and accuracy of intersensory matching in the presence of five other competing visual events. Convergent findings across the MAAP and IPEP and across age (6 and 12 months) demonstrate that individual differences in intersensory processing of faces and voices during audiovisual speech in the first year of life are important predictors of later language outcomes above and beyond well-established predictors of parent language input and SES.

4.2 | Intersensory matching of social events at 6 months predicts child language outcomes, controlling for parent language input at older ages (18, 24, 36 months).

Accuracy of intersensory matching of faces and voices at 6 months even predicted child language outcomes when later parent language input, at 18, 24, or 36 months, was held constant. After controlling for parent language input at these older ages, accuracy of intersensory matching of faces and voices at 6 months predicted expressive vocabulary size at 18, 24, and 36 months, as well as child speech production (quality and quantity) at 18 and 24 months. Thus, given equal amounts of parent language input at 18, 24, and 36 months, accuracy of intersensory matching of faces and voices at 6 months still predicts which children will benefit the most from parent language input later in development.

Although a host of studies have demonstrated the powerful role of parent language input for promoting child language development (Hart & Risley, 1995; Huttenlocher et al., 1991; Rowe, 2008), the contribution of intersensory processing of faces and voices—a skill that underlies the child's ability to effectively utilize language input—remains significantly understudied. The present study highlights the central role of intersensory processing of audiovisual speech in this developmental process. Both the accuracy and speed of face-voice matching play significant roles in determining which children will benefit most from the language environment provided by caregivers. Further, at the younger ages tested here (6 months) and in our prior study using the MAAP (12 months; Edgar et al., 2022), intersensory processing skills were shown to be more robust predictors of later language development than parent language input. They predicted a wider variety of measures (including quality and quantity of speech production and receptive and expressive vocabulary) across a greater range of ages (at 18, 24, and 36 months).

4.3 | Secondary analyses: Intersensory matching of social events at the group- versus individual-level

The present study was designed to assess how variability across individual differences in intersensory matching contribute to variability of individual differences in language outcomes. Although our main research question focused on using individual variability to predict later outcomes, one can also assess mean performance across individuals within a group as a way of comparing intersensory matching on the IPEP with those of prior studies using group-level approaches to assessing intersensory matching. This is relevant for our measure of accuracy of intersensory matching, but not for speed of matching. In contrast with accuracy, there is no “mean value” that can serve as a useful index of “success” for speed of intersensory matching.

As a group, infants showed no evidence of intersensory matching of faces and voices at 3 or 6 months of age on the IPEP. Given that the IPEP assesses face-voice matching under high levels of competing stimulation (i.e., a single sound-synchronous target in the presence of five asynchronous distractor events), matching is more difficult than in traditional tasks previously used to assess intersensory processing (e.g., two-screen intermodal matching task with just one asynchronous distractor event). Thus, we expect evidence of group-level matching to emerge later in development. Our prior research indicates significant intersensory matching at the group level in 2- to 5-year-old children (Bahrick, Todd et al., 2018). Therefore, significant group-level matching emerges sometime between 6 and 24 months as assessed using the IPEP.

Using the IPEP, we can also assess mean performance across trials for individual participants. In contrast to the group-level approach, at the individual participant level a sizeable number of infants showed evidence of significant intersensory matching of faces and voices. Approximately one-quarter of the infants showed significant intersensory matching at 3 months of age, and one-third of the infants showed significant intersensory matching at 6 months of age. Thus, even under conditions where there is a high degree of competing stimulation (from five concurrent faces of women speaking out of synchrony with the soundtrack heard), a portion of infants at both 3 and 6 months of age are able to pick out the face that is synchronized with the voice.

4.4 | Intersensory processing of social, but not nonsocial, events predicts child language outcomes

Similar to our prior study with 12-month-olds (Edgar et al., 2022), individual differences in the accuracy of infant intersensory processing for social, but not nonsocial, events predicted child language outcomes. At 3 months of age, there was no evidence that accuracy of intersensory processing for nonsocial events predicts child language outcomes. Further, at 6 months, there was no evidence that accuracy of intersensory processing for nonsocial events predicts child language outcomes after holding parent language input and SES constant. Thus, we find no evidence that intersensory processing of nonsocial events at 3 (IPEP), 6 (IPEP), and 12 months (MAAP) of age predicts child language outcomes. Intersensory processing of social, but not nonsocial, events may predict language outcomes because social interactions provide the context in which language learning opportunities most often occur. Further, compared to nonsocial events, social events provide an extraordinary amount of intersensory redundancy across face, voice, and gesture (Bahrick et al., 2016; Bahrick & Todd, 2012) and are typically more complex and variable than nonsocial events (Adolphs, 2001; Dawson et al., 2004), thus demanding greater attentional resources from infants. It may be that the challenge of processing social events on the IPEP is more optimally matched to the processing capabilities of 6-month-olds, resulting in meaningful variability across infants that is related to language outcomes. Future research will assess the role of intersensory processing of nonsocial events in predicting outcomes other than language (e.g., overall cognitive functioning, spatial awareness, and/or working memory).

4.5 | SES (maternal education) also predicts multiple child language outcomes at 18, 24, and 36 months

Maternal education, an index of SES, was also a significant predictor of multiple child language outcomes, holding constant accuracy and speed of intersensory matching and parent language input (quality and quantity). It predicted child expressive and receptive vocabulary at both 12 and 36 months. It also predicted measures of child speech production: quality (but not quantity) of child speech at 18, 24, and 36 months. Thus, consistent with prior findings, maternal education plays an increasingly important role in fostering child language development across the first 3 years of life (Hart & Risley, 1995; Hoff, 2003; Rowe, 2018). Critically, accuracy of intersensory matching of faces and voices at 6 months predicted child language outcomes at 18, 24, and 36 months, holding maternal education constant.

4.6 | Parent language input at older ages (but not 6 months) predicts child language outcomes

Although it is well-established that parent language input plays an important role in promoting child language development (Hoff & Naigles, 2002; Weisleder & Fernald, 2013), these studies have focused on older children than those of the present study (e.g., 18 and 24 months). Overall, parent language input (quality and quantity) at 6 months was not a significant predictor of child language outcomes, holding accuracy and speed of intersensory matching and maternal education (SES) constant. The only exception was that quality (but not quantity) of parent language input at 6 months predicted expressive vocabulary at 36 months. In contrast, our supplemental analyses revealed that parent language input at older ages (24 and 36 months, but not 18 months) was a significant predictor of child language outcomes, holding other predictors constant. Findings are consistent with previous literature indicating that parent language input at older ages is a strong predictor of child language (Edgar et al., 2022; Gilkerson et al., 2018; Hoff & Naigles, 2002; Jones & Rowland, 2017; Pan et al., 2005), and that quality of parent language is a stronger predictor of language outcomes than quantity of parent language at older ages (Hsu et al., 2017; Huttenlocher et al., 1991, 2010; Jones & Rowland, 2017; Rowe, 2012). The current findings also replicate and extend our prior findings which also indicated that parent language input at older ages predicts child language outcomes (Edgar et al., 2022). Thus, given similar levels of intersensory matching skills at 6 months and maternal education, parents who provided greater language input at older ages have children with larger vocabulary size.

4.7 | Limitations and future research directions

The present study has several limitations that can be addressed by future research. First, it is possible that there are interactions among intersensory processing, maternal education, and parent language input. Our future research will examine non-linear relations among these variables, as well as potential mediating and moderating relations among our predictors. Second, only child expressive (and not receptive) vocabulary was available at 24 months, given how the MB-CDI forms differ for children of 12 and 18 versus 24 months. Thus, future studies should assess receptive and expressive vocabulary at all ages to better understand relations between infant intersensory processing and these child language outcomes. Third, the present study used different methods of assessing receptive vocabulary at 12 and 18 months (parent-report form; MB-CDI) as compared with 36 months (behavioral assessment of the child; PPVT). It is possible that the use of a behavioral assessment may provide a more fine-grained index of language at 36 months and thus might explain why intersensory processing at 6 months predicted receptive language at 36, but not at 12 or 18, months.

Other fruitful directions for future research include assessing other variables that predict intersensory processing abilities early in development to reveal potentially earlier parts of a developmental cascade to language outcomes. Currently, data from two different protocols (MAAP and IPEP) at two different ages (the MAAP at 12 months and the IPEP at 6 months) have demonstrated that individual differences in intersensory processing of faces and voices in the context of visual distractors are meaningful predictors of later language outcomes. Further, a number of other predictors not examined here including early language perception and production (e.g., speech processing efficiency, word mapping, early infant vocal production) and cognitive abilities (working memory, processing speed, early visual reception skills) have been shown to play an important role in later child language outcomes. Understanding these developmental cascades from earlier predictors of intersensory processing of faces and voices to language outcomes through other predictors will reveal more about underlying mechanisms of development and inform interventions targeted at improving child language development. Future research will also explore possible differences in intersensory processing strategies elicited by each protocol at different ages (e.g., serial search; pop out effects) as well as the impact of distractors using micro-level measures of looking time including eye tracking to reveal more about developmental change and the underlying mechanisms of development.

4.8 | Implications for the study of language development

Findings from the present study have a number of implications for the study of child language development. First, the present study adds to a growing body of literature highlighting the importance of assessing individual differences in intersensory processing for understanding relations between this basic, foundational skill and more complex developmental outcomes. It replicates and extends prior findings demonstrating that intersensory processing of social events (faces and voices) predicts concurrent and future language outcomes in typically developing children (Bahrick, Todd et al., 2018; Edgar et al., 2022) and children with ASD (Righi et al., 2018; Todd & Bahrick, in press). Second, it highlights the importance of infancy (the first 12 months) as a foundational period for the development of intersensory processing of social events. Though intersensory processing continues to improve with age, our findings suggest that more efficient selective attention to audiovisual speech in infancy may allow infants to take better advantage of early word learning opportunities (e.g., object labelling), which occur in the context of early social-communicative interactions with caregivers. Third, findings highlight the importance of characterizing developmental pathways and cascades from basic intersensory processing to later, more complex language skills that rely on this foundation in typical and atypical development. Our findings suggest that impairments of intersensory processing of faces and voices in infancy may be an indicator of risk for language delays. A goal of future research should be to characterize whether early individual differences in intersensory processing of faces and voices can identify children who go on to develop impaired language outcomes. If such links are established, interventions to train and improve early intersensory processing skills in infancy may be designed and lead to subsequent improvements in later language outcomes.

ACKNOWLEDGMENTS

We thank the children and families who participated in this research. We also thank the team of research assistants in the Infant Development Lab who contributed to transcription of the parent-child interactions. This research was supported by NICHD grants R01HD094803 and R01 HD053776 awarded to the second and third authors.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest with regard to the funding source for this study.

ORCID

Elizabeth V. Edgar https://orcid.org/0000-0003-4419-1876

REFERENCES

Adolphs, R. (2001). The neurobiology of social cognition. *Current Opinion in Neurobiology*, *11*(2), 231–239. https://doi.org/10.1016/s0959-4388(00)00202-6

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. https://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. https://doi.org/10.1037/a0039964

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

Bahrick, L. E. (1987). Infants' intermodal perception of two levels of temporal structure in natural events. *Infant Behavior and Development*, *10*(4), 387–416. https://doi.org/10.1016/0163-6383(87)90039-7

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. https://doi.org/10.1111/j.1467-8624.1988.tb03208.x

Bahrick, L. E. (1992). Infants' perceptual differentiation of amodal and modality-specific audio-visual relations. *Journal of Experimental Psychology*, *53*(2), 180–199. https://doi.org/10.1016/0022-0965(92)90048-b

Bahrick, L. E. (1994). The development of infants' sensitivity to arbitrary intermodal relations. *Ecological Psychology*, *6*(2), 111–123. https://doi.org/10.1207/s15326969eco0602

Bahrick, L. E. (2001). Increasing specificity in perceptual development: Infants' detection of nested levels of multimodal stimulation. *Journal of Experimental Child Psychology*, *79*(3), 253–270. https://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., Vol. 1, pp. 120–166). Blackwell Publishing. https://doi.org/10.1002/9781444327564.ch4

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. https://doi.org/10.1002/dev.10049

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology*, *36*(2), 190–201. https://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. V. Kail (Ed.), *Advances in child development and behavior* (pp. 153–187). Academic Press. https://doi.org/10.1016/S0065-2407(02)80041-6

Bahrick, L. E., & Lickliter, R. (2004). Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: A developmental test of the intersensory redundancy hypothesis. *Cognitive, Affective, & Behavioral Neuroscience*, *4*(2), 137–147. https://doi.org/10.3758/cabn.4.2.137

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 University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0008

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. https://doi.org/10.1177/0963721414549187

Bahrick, L. E., Lickliter, R., Castellanos, I., & Vaillant-Molina, M. (2010). Increasing task difficulty enhances effects of intersensory redundancy: Testing a new prediction of the Intersensory Redundancy Hypothesis. *Developmental Science*, *13*(5), 731–737. https://doi.org/10.1111/j.1467-7687.2009.00928.x

Bahrick, L. E., Lickliter, R., & Todd, J. T. (2020). The development of multisensory attention skills: Individual differences, developmental outcomes, and applications. In J. J. Lockman & C. S. Tamis-LeMonda (Eds.), *The Cambridge handbook of infant development* (pp. 303–338). Cambridge University Press.

Bahrick, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. *Child Development*, *69*(5), 1263–1275. https://doi.org/10.1111/j.1467-8624.1998.tb06210.x

Bahrick, L. E., & Pickens, J. N. (1988). Classification of bimodal English and Spanish language passages by infants. *Infant and Child Development*, *11*(3), 277–296. https://doi.org/10.1016/0163-6383(88)90014-8

Bahrick, L. E., Soska, K. C., & Todd, J. T. (2018a). Assessing individual differences in the speed and accuracy of intersensory processing in young children: The Intersensory Processing Efficiency Protocol. *Developmental Psychology*, *54*(12), 2226–2239. https://doi.org/10.1037/dev0000575

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 processes* (pp. 657–674). MIT Press.

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. https://doi.org/10.1037/dev0000157

Bahrick, L. E., Todd, J. T., & Soska, K. C. (2018b). The Multisensory Attention Assessment Protocol (MAAP): Characterizing individual differences in multisensory attention skills in infants and children and relations with language and cognition. *Developmental Psychology*, *54*(12), 2207–2225. https://doi.org/10.1037/dev0000594

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (2012). *Multisensory development*. Oxford University Press.

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), 604–616. https://doi.org/10.2307/1130560

Chawarska, K., Lewkowicz, D., Feiner, H., Macari, S., & Vernetti, A. (2022). Attention to audiovisual speech does not facilitate language acquisition in infants with familial history of autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *63*(12), 1466–1476. https://doi.org/10.1111/jcpp.13595

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analyses for the behavioral sciences* (3rd ed.). Lawrence Erlbaum Associates.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: Social orienting, joint attention, and attention to distress. *Developmental Psychology*, *40*(2), 271–283. https://doi.org/10.1037/0012-1649.40.2.271

Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary picture* (4th ed.). NCS Pearson.

Edgar, E. V., Todd, J. T., & Bahrick, L. E. (2022). Intersensory matching of faces and voices in infancy predicts language outcomes in young children. *Developmental Psychology*, *58*(8), 1413–1428. https://doi.org/10.1037/dev0001375

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). MacArthur-Bates Communicative Development Inventories: User's guide and technical manual (2nd ed.).

Fiebelkorn, I. C., Foxe, J. J., & Molholm, S. (2012). Attention and mulitsensory feature integration. In B. E. Stein (Ed.), *The new handbook of multisensory processing* (pp. 383–394). MIT Press.

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

Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018). Language experience in the second year of life and language outcomes in late childhood. *Pediatrics*, *142*(4). https://doi.org/10.1542/peds.2017-4276

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. https://doi.org/10.1006/jecp.1998.2438

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*(4), 878–894. https://doi.org/10.1111/1467-8624.00197

Gogate, L. J., Bolzani, L. H., & Betancourt, E. A. (2006). Attention to maternal multimodal naming by 6- to 8-month-old infants and learning of word-object relations. *Infancy*, *9*(3), 259–288. https://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. https://doi.org/10.1037/a0019049

Gogate, L. J., Walker-Andrews, A. S., & Bahrick, L. E. (2001). The intersensory origins of word comprehension: An ecological-dynamic systems view. *Developmental Science*, *4*(1), 1–18. https://doi.org/10.1111/1467-7687.00143

Guellaï, B., Streri, A., Chopin, A., Rider, D., & Kitamura, C. (2016). Newborns' sensitivity to the visual aspects of infant-directed speech: Evidence from point-line displays of talking faces. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(9), 1275–1281. https://doi.org/10.1037/xhp0000208

Hart, B., & Risley, T. R. (1992). American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental Psychology*, *28*(6), 1096–1105. https://doi.org/10.1037/0012-1649.28.6.1096

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H. Brookes Publishing.

Hillairet de Boisferon, A., Tift, A. H., Minar, N. J., & Lewkowicz, D. J. (2018). The redeployment of attention to the mouth of a talking face during the second year of life. *Journal of Experimental Child Psychology*, *172*, 189–200. https://doi.org/10.1016/j.jecp.2018.03.009

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. *Child Development*, *74*(5), 1368–1378. https://doi.org/10.1111/1467-8624.00612

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Developmen*t, *73*(2), 418–433. https://doi.org/10.1111/1467-8624.00415

Hsu, N., Hadley, P. A., & Rispoli, M. (2017). Diversity matters: Parent input predicts toddler verb production. *Journal of Child Language*, *44*(1), 63–86. https://doi.org/10.1017/S0305000915000690

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology*, *27*(2), 236–248. https://doi.org/10.1037/0012-1649.27.2.236

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive Psychology*, *61*(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Jaccard, J., & Becker, M. A. (2009). *Statistics for the behavioral sciences* (5th ed.). Cengage Learning.

Jackson-Maldonado, D., Thal, D., Marchman, V. A., Newton, T., Fenson, L., & Conboy, B. (2003). MacArthur Inventorios del Desarollo de Habilidades Communicativas: User's guide and technical manual.

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. https://doi.org/10.1016/j.jecp.2015.12.002

Jones, G., & Rowland, C. F. (2017). Diversity not quantity in caregiver speech: Using computational modeling to isolate the effects of the quantity and the diversity of the input on vocabulary growth. *Cognitive Psychology*, *98*, 1–21. https://doi.org/10.1016/j.cogpsych.2017.07.002

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. *Science*, *218*(4577), 1138–1141. https://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 and Development*, *15*(3), 297–324. https://doi.org/10.1016/0163-6383(92)80002-C

Lewkowicz, D. J. (2000a). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin*, *126*(2), 281–308. https://doi.org/10.1037/0033-2909.126.2.281

Lewkowicz, D. J. (2000b). Infants' perception of the audible, visible, and bimodal attributes of multimodal syllables. *Child Development*, *71*(5), 1241–1257. https://doi.org/10.1111/1467-8624.00226

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. https://doi.org/10.1037/0012-1649.39.5.795

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

Lewkowicz, D. J., & Marcovitch, S. (2006). Perception of audiovisual rhythm and its invariance in 4- to 10-month-old infants. *Developmental Psychobiology*, *48*(4), 288–300. https://doi.org/10.1002/dev.20140

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk (third edition): Transcription format and programs (3rd ed.)*. Lawrence Erlbaum Associates Publishers. https://doi.org/10.1162/coli.2000.26.4.657

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), 9–16. https://doi.org/10.1111/j.1467-7687.2008.00671.x.Speed

Morales, M., Mundy, P., & Rojas, J. (1998). Following the direction of gaze and language development in 6-month-olds. *Infant Behavior and Development*, *21*(2), 373–377. https://doi.org/10.1016/S0163-6383(98)90014-5

Morin-Lessard, E., Poulin-Dubois, D., Segalowitz, N., & Byers-Heinlein, K. (2019). Selective attention to the mouth of talking faces in monolinguals and bilinguals aged 5 months to 5 years. *Developmental Psychology*, *55*(8), 1640–1655. https://doi.org/10.1037/dev0000750

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, *76*(4), 763–782. https://doi.org/10.1111/1467-8624.00498-i1

Patterson, M. L., & Werker, J. F. (1999). Matching phonetic information in lips and voice is robust in 4.5-month-old infants. *Infant Behavior and Development*, *22*(2), 237–247. https://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. https://doi.org/10.1111/1467-7687.00271

Richoz, A. R., Quinn, P. C., De Boisferon, A. H., Berger, C., Loevenbruck, H., Lewkowicz, D. J., Lee, K., Dole, M., Caldara, R., & Pascalis, O. (2017). Audio-visual perception of gender by infants emerges earlier for adult-directed speech. *PLoS One*, *12*(1), 1–15. https://doi.org/10.1371/journal.pone.0169325

Righi, G., Tenenbaum, E. J., McCormick, C., Blossom, M., Amso, D., & Sheinkopf, S. J. (2018). Sensitivity to audio-visual synchrony and its relation to language abilities in children with and without ASD. *Autism Research*, *11*(4), 645–653. https://doi.org/10.1002/aur.1918

Rose, S. A., & Feldman, J. F. (1987). Infant visual attention: Stability of individual differences from 6 to 8 months. *Developmental Psychology*, *23*(4), 490–498. https://doi.org/10.1037/0012-1649.23.4.490

Rosenblum, L. D. (2008). Speech perception as a multimodal phenomenon. *Current Directions in Psychological Science*, *17*(6), 405–409. https://doi.org/10.1111/j.1467-8721.2008.00615.x

Rowe, M. L. (2008). Child-directed speech: Relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language*, *35*(1), 185–205. https://doi.org/10.1017/S0305000907008343

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech vocabulary development. *Child Development*, *83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Rowe, M. L. (2018). Understanding socioeconomic differences in parents' speech to children. *Child Development Perspectives*, *12*(2), 122–127. https://doi.org/10.1111/cdep.12271

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Smith, L. B., Suanda, S. H., & Yu, C. (2014). The unrealized promise of infant statistical word–referent learning. *Trends in Cognitive Sciences*, *18*(5), 251–258. https://doi.org/10.1016/j.tics.2014.02.007

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. https://doi.org/10.1111/j.1467-8624.1992.tb01661.x

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

Stevenson, R. A., Segers, M., Ferber, S., Barense, M. D., & Wallace, M. T. (2014). The impact of multisensory integration deficits on speech perception in children with autism spectrum disorders. *Frontiers in Psychology*, *5*, 1–4. https://doi.org/10.3389/fpsyg.2014.00379

Tenenbaum, E. J., Sobel, D. M., Sheinkopf, S. J., Malle, B. F., & Morgan, J. L. (2015). Attention to the mouth and gaze following in infancy predict language development. *Journal of Child Language*, *42*(6), 1173–1190. https://doi.org/10.1017/S0305000914000725

Todd, J. T., & Bahrick, L. E. (2022). Individual differences in multisensory attention skills in children with autism spectrum disorder predict language and symptom severity: Evidence from the Multisensory Attention Assessment Protocol (MAAP). (in press). *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-022-05752-3

Tsang, T., Atagi, N., & Johnson, S. P. (2018). Selective attention to the mouth is associated with expressive language skills in monolingual and bilingual infants. *Journal of Experimental Child Psychology*, *169*, 93–109. https://doi.org/10.1016/j.jecp.2018.01.002

Vaillant-Molina, M., Bahrick, L. E., & Flom, R. (2013). Young infants match facial and vocal emotional expressions of other infants. *Infancy*, *18*, 97–111. https://doi.org/10.1111/infa.12017

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

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

Walker-Andrews, A. S., & Grolnick, W. (1983). Discrimination of vocal expressions by young infants. *Infant Behavior and Development*, *6*(4), 491–498. https://doi.org/10.1016/S0163-6383(83)90331-4

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.1177/0956797613488145

Weizman, Z. O., & Snow, C. E. (2001). Lexical input as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology*, *37*(2), 265–279. https://doi.org/10.1037/0012-1649.37.2.265

Williams, K. T. (2007). *Expressive vocabulary test* (2nd ed.). NCS Pearson.

Young, G. S., Merin, N., Rogers, S. J., & Ozonoff, S. (2009). Gaze behavior and affect at 6 months: Predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Developmental Science*, *12*(5), 798–814. https://doi.org/10.1111/j.1467-7687.2009.00833.x

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Edgar, E. V., Todd, J. T., & Bahrick, L. E. (2023). Intersensory processing of faces and voices at 6 months predicts language outcomes at 18, 24, and 36 months of age. *Infancy*, *28*(3), 569–596. https://doi.org/10.1111/infa.12533