Daniel C. Hyde[1]
Blake L. Jones[3]
Ross Flom[2]
Chris L. Porter[3]

[1]*Department of Psychology, Harvard University, 1118 WJH, 33 Kirkland Street, Cambridge, MA 02138*
Email: dchyde@fas.harvard.edu

[2]*Department of Psychology, Brigham Young University, Provo, UT*

[3]*School of Family Life, Brigham Young University, Provo, UT*

Neural Signatures of Face–Voice Synchrony in 5-Month-Old Human Infants

ABSTRACT: *Infants' unitary perception of their multisensory world, including learning from people (faces and speech), hinges on temporal synchrony. Despite its importance, relatively little work has investigated the brain processes involved in infants' perception of temporal synchrony. In two experiments, we examined event-related brain potentials (ERPs) to asynchronous and synchronous audio-visual speech in infants. Both experiments showed the early auditory P2 was greater for the synchronously presented pairings and later attentional processing (Nc) was greater for asynchronous pairings. In addition, dynamic stimuli used in Experiment 2 produced a greater early visual response (N1) to the asynchronous condition and an enhanced memory-related slow wave (PSW) later for the synchronous condition. These results suggest that, like adults, auditory–visual integration for young infants begins early during sensory processing rather than later during higher-level cognitive processing. However, unlike adults, infants' brain responses may be biased towards synchrony. Furthermore, effects of attentional and memory processing confirm interpretations of behavioral looking patterns suggesting infants find synchrony more familiar. © 2011 Wiley Periodicals, Inc.* Dev Psychobiol 53: 359–370, 2011.

Keywords: *event-related potentials; infancy; perception; intersensory; multisensory; sensory integration; attention; memory; auditory processing; visual processing*

INTRODUCTION

Infants' experiences with others, in particular the faces and voices of others, are fundamental for early cognitive, social, and linguistic development. In fact, faces and voices are so important to development that newborn infants look longer toward face-stimuli compared to non-face stimuli (Johnson, Dzurawaec, Ellis, & Morton, 1991) and prefer human speech sounds to nonspeech sounds (Butterfield & Siperstein, 1970; Vouloumanos & Werker, 2007). While infants' experience with faces and voices is essential for development, perhaps more important is the integration or ''binding'' of faces and voices into a unitary perception.

One cue that specifies whether a particular sight and sound go together, and other sights and sounds do not, is temporal synchrony. A substantial body of evidence has accrued showing that within the first 2.5- to 3.5-months of life infants are sensitive to changes in temporal synchrony in auditory–visual events (Bahrick & Lickliter, 2002; Lewkowicz & Kraebel, 2004, for reviews). For example, 10- to 16-week-old infants looked longer to a video clip of an adult speaking a nursery rhyme that maintained lip–voice synchrony compared to an event where the movements of the lips and occurrence of speech were asynchronous by 400 ms (Dodd, 1979). Likewise numerous studies demonstrate within the first 3–6 months of life infants detect differences in temporal audio-visual synchrony within the context of everyday objects and events (e.g., Bahrick, 1992; Bahrick, Netto, & Hernandez-Reif, 1998; Lewkowicz, 1992a,b, 1996; Scheier, Lewkowicz, & Shimojo, 2003).

Received 30 August 2010; Accepted 7 December 2010
Correspondence to: Daniel C. Hyde
E-mail: dchyde@fas.harvard.edu
Contract grant sponsor: Family Studies Center and the School of Family Life at Brigham Young University.
Published online 26 January 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/dev.20525

In addition to infants discriminating events on the basis of temporal synchrony, infants are also able to recognize and match sights and their associated sounds on the basis of temporal synchrony. For example, 4.5- to 5-month olds' matching of the visual lip movements and the corresponding auditory articulation of a vowel sound (i.e., vowel sounds /a/ and /i/) is partially dependent on temporal synchrony (Kuhl & Meltzoff, 1982). Likewise, in the domain of objects, others have shown that 4-month olds show a visual preference for an object when the auditory–visual relationship is temporally synchronous but no apparent preference when the auditory–visual relationship is asynchronous (Bahrick, Walker, & Neisser, 1981; Lewkowicz, 1992a; Spelke, Born, & Chu, 1983).

Perceiving multimodal synchronous events has been demonstrated to facilitate learning and memory in adults and infants (e.g., Erber, 1975; Frank, Slemmer, Marcus, & Johnson, 2009; Grant & Seitz, 2000; Sumby & Pollack, 1954; Summerfield, 1979). It has been argued that infants' perception of audio-visual synchrony is important for learning about language, people, objects, and events throughout development (Bahrick, 2000, 2001; Bahrick & Lickliter, 2002; Lewkowicz, 2000; Lewkowicz & Kraebel, 2004). An example of this developmental advantage for audio-visual synchrony is found in a study where Bahrick (2001) habituated 4-, 7-, and 11-week-old infants to object-sound pairs. One pair consisted of a large metal nut that made ''thud'' sound upon impacting a surface and the other pair was a cluster of smaller metal nuts that made a ''crash'' sound upon impact. Results revealed that infants at all ages discriminated a disruption in the temporal synchrony of the object hitting the surface and the occurrence of the impact sound. However, older, but not younger infants learned and noticed when the object made the ''wrong'' sound upon impact. Thus, temporal synchrony provides a foundation for infants' subsequent learning about multi-sensory events, including objects (e.g. Bahrick, 2001), people (Bahrick, 2000; Bahrick, Hernandez-Reif, & Flom, 2005), and language (Gogate, 2010; Gogate, Prince, & Matatyaho, 2009).

Developmental investigations of intersensory perception, therefore, provide evidence of infants' integration of information across sensory modalities, examine the conditions under which they are able to do so, and reveal the importance of synchrony to learning. Developmental research, however, provides little information regarding how intersensory perception of synchrony actually occurs. Developmental neurophysiologic investigations may provide key insights into the process of intersensory perception by documenting the spatial, temporal, and functional properties of neural activity that give rise intersensory perception. For example, several studies on the neural signatures of synchrony perception in adults suggest that the integration of information across the senses begins during early rather than later stages of sensory processing. For example, when adults are presented with synchronous audio-visual speech the early auditory event-related brain potential (ERP) components (N1–P2) are attenuated compared to the response when speech is presented alone. This suggests that at least some aspects of sensory integration are occurring before higher-level cognitive processing such as semantic categorization (Besle, Fort, Delpuech, & Giard, 2004; Näätänen & Winkler, 1999; van Wassenhove, Grant, & Poepel, 2005). Interestingly, the attenuation found in the auditory ERP for synchronous audio-visual pairings does not occur when audio-visual speech is presented asynchronously (Pilling, 2009), seemingly ruling out explanations of the auditory attenuation being based on attentional shifts between visual and auditory sensory information (Besle et al., 2004; van Wassenhove et al., 2005) or inhibition of auditory processing during audio-visual stimulation due to focusing on visual information (Shulman et al., 1997). Therefore, the evidence in adults supports ''early integration'' models of intersensory perception (e.g., Braida, 1991; Green, 1998) and runs contrary to other models that propose integration occurs during later stages of cognitive processing (e.g., Massaro, 1987, 1998). One question that arises is whether integration occurs during early sensory processing in infants in a manner similar to adults?

Little if any neurophysiological work has been conducted on synchrony processing in infants. However, a few studies of the neurophysiologic correlates of cross-modal matching have been conducted and suggest that integration may in fact occur during earlier processing stages in infants, like in adults. For example, a recent study showed that infant ERPs revealed an early mismatch negativity (MMN), around 150–300 ms after stimulus onset, for an auditory syllable that was incongruent with the articulation movements of a previously seen face compared to the response to an auditory syllable that was congruent with the articulation movements of a previously seen face (Bristow et al., 2009). Based on these findings and given the influential role of synchrony in audio-visual integration, it is likely that manipulations of synchrony will also influence neural processes during these earlier stages of sensory processing.

There is also reason to believe that synchrony manipulations (i.e., manipulating whether visual and auditory stimuli are presented synchronously or asynchronously) will affect later neural processes associated with attentional orienting, familiarity, and/or recognition memory. Specifically work on cross-modal emotion perception in infants has shown that 7-month-old infants demonstrate an attenuated attentional or orienting response (Nc) and a greater positive slow wave (PSW) to the synchronous presentation of a face–voice pairing with congruent

emotions compared to the presentation of affectively incongruent face–voice pairings (Grossmann, Striano, & Friederici, 2006). The infant Nc component is thought of as an indicator of attentional orienting and has been localized to frontal and prefrontal regions (e.g., Ackles & Cook, 1998; de Haan & Nelson, 1997; Goldman, Shapiro, & Nelson, 2004; Nelson, 1994; Reynolds & Richards, 2005). The later occurring PSW is thought to reflect the updating of memory (Nelson, 1994, 1996, 1998; Nelson & Collins, 1991, 1992). Source analyses of infant ERPs suggest the PSW originates from memory regions of the temporal lobe (e.g., Reynolds, Courage, & Richards, 2010; Reynolds & Richards, 2005). Based on this evidence, the specific pattern of an attenuated Nc and a larger PSW to synchronous audio-visual speech is thought to reflect attentional orienting to the incongruent pairing and recognition of the congruent pairing (Grossmann et al., 2006). It is likely that an attenuated Nc response and a larger PSW response will be observed to synchronous compared to asynchronous presentation of faces and voices because presumably face–voice synchrony is more familiar or recognizable compared to face–voice asynchrony.

To investigate the effects of synchrony on early sensory and later attentional processing, we recorded the electrophysiological response in 5-month-old human infants to synchronous and asynchronous bimodal audio-visual speech. In Experiment 1, 5-month olds viewed both synchronous and asynchronous pairings of static pictures of female faces with a voice saying ''hi.'' In Experiment 2, a second group of 5-month olds again saw synchronous and asynchronous conditions except the events were presented dynamically as movies. Five-month olds were used in both experiments because infants of this age have shown reliable behavioral discrimination and recognition of temporal synchrony (Bahrick & Lickliter, 2002; Lewkowicz & Kraebel, 2004).

EXPERIMENT 1

Methods

Participants. A total of 36, 5-month-old infants (mean age = 150.44 days) participated in Experiment 1. Sixteen infants made up the final data set (8 females). Another 20 infants participated in the experiment, but were excluded from analysis for the following reasons: 13 for fussiness or crying causing the experiment to be terminated before completion and 7 for too few good segments after artifact rejection. This attrition rate is comparable to previous ERP studies of young infants (e.g., Hyde, Jones, Porter, & Flom, 2010; Hyde & Spelke, in press; Izard, Dehaene-Lambertz, & Dehaene, 2008; Quinn, Westerland, & Nelson, 2006). Informed consent was obtained from one parent or guardian before beginning. Families did not receive any financial compensation for participation, but were given a certificate of appreciation along with a color photo of their infant wearing the sensor cap.

Stimuli and Procedure. Stimuli consisted of face–voice pairs from two different actresses presented synchronously or asynchronously. In the synchronous condition, the face and the voice saying ''hi'' appeared in perfect temporal synchrony and lasted for 1 s. In the asynchronous condition the voice sounded 400 ms before the face appeared and then the face and voice overlapped for 600 ms. Faces portrayed a happy emotional expression and voices were made to be friendly in nature to maintain infants' interest. A blank-screen inter-stimulus interval was presented between stimuli that varied randomly in length from 1,000 to 2,000 ms. Each condition appeared 30 times for a total of 60 stimuli (30 of each actress). The order of presentation was pseudo-random with the constraint that each condition (synchrony and asynchrony) appeared once before re-randomization occurred. Two times during the experiment a short break was taken by showing the infant a picture of a farm animal with an accompanying animal sound. This was done in an effort to reduce boredom and recapture infants' attention to the video screen. Stimuli were presented from a 17-inch computer screen and computer speakers approximately 80–100 cm from the infant seated in a parent's lap. The parent was instructed to look at the top of the infant's head instead of the display in order to blind them from the experimental condition that was being presented. Parents were also instructed not to speak to the infant in an attempt to reduce environmental interference.

Data Acquisition. Infants' heads were first measured and then, while the vertex was being located and marked on the baby's scalp, the appropriate size sensor net was soaked in a potassium chloride solution. While the infant sat in a parent's lap, the sensor net was systematically placed on the head with reference to the identified vertex and left and right mastoid bones. Before data collection, impedances were checked and maintained below 50 Hz. The ongoing EEG was recorded from scalp locations using a 64 channel HydroCel Geodesic Sensor Net (Electrical Geodesics Inc., Eugene, OR) as infants were presented with stimuli in a dimly lit room. Data were recorded at 250 samples per second and digitally filtered online at 0.1–100 Hz, referenced to the Cz. Stimuli were paused when the infant looked away from the screen and resumed when the infant faced forward again. If the baby became fussy, began to cry, or continually looked away, data collection was terminated.

Data Reduction. Data from infants that completed the experiment were further processed in three steps. First, raw data were lowpass filtered at 30 Hz, segmented into epochs from 200 ms before to 1,900 ms after stimulus onset, and baseline corrected to the 200 ms before the stimulus was presented. Second, each epoch was visually examined for artifacts by two experienced staff members. Any epoch containing an eye blink, eye movement, excessive noise, or more than 10 bad channels was rejected from further analysis. Questionable epochs were discussed between staff members until consensus could be reached. Any subject retaining less than 10 good epochs in either experimental condition after artifact rejection was eliminated from the final analysis. Remaining acceptable epochs were further processed by running an automated bad channel replacement algorithm based on spherical spline interpolation for trials containing less than 10 bad channels, then creating averages for each experimental condition for each subject, re-referencing the data to the average reference, and again baseline correcting to 200 ms before stimulus onset to correct for any absolute amplitude differences created by processing. In addition to subject averages, a grand average for all conditions was created for visual inspection purposes.

Data Analysis. Data were examined during three times windows of interest based on visual inspection of the grand average waveforms and scalp topography (all experimental conditions averaged together) and guided by previous research on speech and face processing: auditory P2 (150–250 ms), attentional Nc (400–800 ms for synchronous and 800–1,200 ms for asynchronous), and PSW (1,100–1,300 ms for synchronous and 1,500–1,700 ms for asynchronous). The auditory P2 and the PSW was measured over left (22, 24, 25, 26) and right (46, 48, 49, 52) lateral sites and the attentional Nc was measured over fronto-central scalp sites (sites 4, 7, 15, 16, 20, 21, 41, 50, 51, 53, 54), as has been customary in previous studies of auditory, attentional, and memory processing in infants (de Hann, 2007 for a review). It should be noted that visual inspection of the grand average waveform (average of all experimental conditions) is the standard and accepted method of choosing time windows and electrode sites of interest in infants, as it conservatively estimates the average evoked response of all conditions without bias towards any particular pattern of modulation between experimental conditions (e.g., recent papers using similar method: Elsabbagh et al., 2009; Grossmann, Gliga, Johnson, & Mareschal, 2009; Kushnerenko, Teinonen, Volein, & Csibra, 2008; Scott & Nelson, 2006). Mean amplitudes of the time windows averaged over sites of interest were compared across conditions using paired samples *t*-tests.

Results

Auditory Response. The onset of the auditory stimulus was the same for both the asynchronous and the synchronous condition. A small negativity (N1) was followed by a more prominent positive component (auditory P2) over lateral scalp sites. Analysis of the auditory P2 yielded a significant difference over left lateral sites with the synchronous condition eliciting a greater amplitude positivity compared to the asynchronous condition, $t(15) = 2.30$, $p < 0.05$ (Fig. 1). Mean amplitude over right lateral sites for the P2 was not significantly different between conditions during this time frame, $t(15) = 0.53$, $p = 0.61$.

Attentional Response. A large slow negative component (Nc) was observed in both experimental conditions over fronto-central scalp locations. This negativity peaked around 600 ms for the synchronous condition and peaked around 1,000 ms for the asynchronous condition and thus seems to be associated with the onset of the visual stimulus (as the face appeared 400 ms later in the asynchronous condition). A comparison of the mean amplitude of the synchronous Nc (400–800 ms) to that of the asynchronous Nc (800–1200 ms) revealed a significant difference in the average ERP over fronto-central scalp sites, $t(15) = 2.94$, $p = .01$, with the asynchronous condition eliciting a greater magnitude negativity compared to the synchronous condition (Fig. 1).

Memory Processing. A large PSW was observed over temporal sites for both conditions. This positivity peaked around 1,200 ms for the synchronous condition and 1,600 ms for the asynchronous condition (accounted for by the 400 ms delay in visual stimulation). No significant differences, however, were observed in the mean amplitude of left, $t(15) = 1.65$, $p > .1$, or right, $t(15) = 0.15$, $p > .1$, lateral scalp sites between experimental conditions.

Discussion

Results suggest integration of information across the senses in infants, like adults (e.g., Besle et al., 2004; Näätänen & Winkler, 1999; Pilling, 2009; van Wassenhove et al., 2005), begins early during sensory processing but also extends into later attentional processing. Specifically, the magnitude of early auditory-evoked electrophysiological components was greater for the synchronous condition compared to the asynchronous condition. This pattern runs in contrast to the attenuated response for synchronously presented face/voice pairings seen in adults. Given the hypothesized importance of synchronous audio-visual events for learning, it is possible infants' pattern of responses, relative to adults,

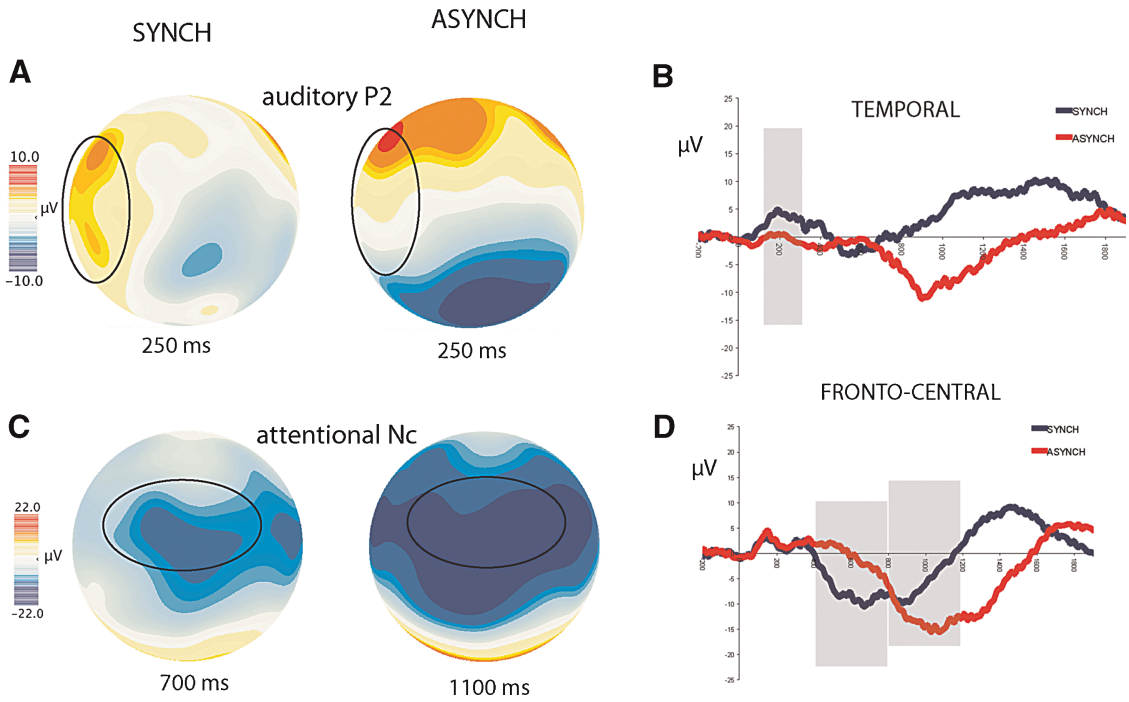

FIGURE 1 Summary of Experiment 1 results. (A) Grand average scalp topography at 250 ms over left temporal sites for each experimental condition. (B) Average waveform from −200 to 1,900 ms averaged over left temporal sites characterizing auditory processing. The shaded region represents the statistical comparison of experimental conditions for the auditory P2. (C) Grand average scalp topography at 700 and 1,100 ms characterizing the Nc for the synchronous and asynchronous conditions over fronto-central sites. (D) Average waveform from −200 to 1,900 ms averaged over fronto-central sites. The shaded region represents the comparison of experimental conditions for the Nc. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

may reflect a processing bias to detect synchrony when it is present (e.g., Bahrick, 2000, 2001; Bahrick & Lickliter, 2002; Lewkowicz & Kraebel, 2004).

The later Nc response, indicative of attentional processing or orienting, showed greater amplitude potentials for the asynchronous condition. In accordance with previous studies, this response suggests that infants found the asynchronous condition more interesting, unfamiliar, or novel compared to the synchronous condition (Ackles & Cook, 1998; de Haan & Nelson, 1997; Goldman et al., 2004; Nelson, 1994; Reynolds & Richards, 2005). This response may reflect the tendency of young infants to focus on the temporal relations among the sense information when encountering unfamiliar, asynchronous audio-visual events (Lewkowicz, 2010).

However, both conclusions are tentative primarily because the static face–voice pairings introduced several limitations. For example, to achieve asynchrony, a delay in the presentation of the visual face and the auditory voice was required. Because of this delay, we cannot rule out the possibility that the sensory and attentional effects observed are due to actual sensory differences between experimental conditions resulting from the presentation method rather than the effects of synchrony. In addition, because of the delay, a valid comparison of visual processing was not possible. Another limitation of the static stimuli is that they are not ecologically valid; rarely do we see voices paired with faces that are not moving. Furthermore, it is also important to note that because we used a static face, these results are limited to conclusions regarding synchrony of onset between face and voice, rather than synchrony of dynamic movement.

EXPERIMENT 2

In an attempt to better understand the results of Experiment 1, we conducted a second experiment in

which we presented infants with dynamic face–speech pairings. The use of dynamic events in Experiment 2 allowed us to examine infants' response to ongoing rhythmic face–speech synchrony and asynchrony compared to Experiment 1 where we examined infants' response to the synchronous and asynchronous onset of a face and voice.

Methods

Similar methods as those used in Experiment 1 were used in Experiment 2. Differences are described below.

Participants. A total of 39, 5-month-old infants (mean age = 155.85 days) participated in Experiment 2. All infants completed the experiment, but data from four infants were unusable because of improper EEG net placement and data from 14 infants were rejected for too few trials after artifact rejection (<12 per experimental condition). The final data set for Experiment 2 consisted of 21 infants (9 females).

Stimuli. Stimuli were dynamic faces paired with an audio track of a voice saying, ''Oh, hi baby.'' Two different actresses' face–speech pairings were used and were either presented synchronously or asynchronously. More specifically, the audio track started at the same time for both synchronous and asynchronous conditions; the synchronous condition contained a segment of video where facial movements matched the words being said but the asynchronous condition contained a different segment of the video that did not correspond to the audio track (faces mouthed the words, ''you're such a beautiful baby'').

Data Reduction and Analysis. The change in stimuli evoked components of slightly different latencies and, in some cases, different scalp topography compared to Experiment 1. As a result, we analyzed defined the components of interest with slightly different time windows and electrode sites to characterize neural processing in Experiment 2. Sites again were chosen through visual inspection of the grand average waveform (all conditions averaged together into one waveform). Specifically, we analyzed the auditory P2 (200–300 ms) and the PSW (1,000–1,300 ms) over the same lateral scalp sites as Experiment 1, and the Nc (500–700 ms) over fronto-central scalp sites (sites: 2, 11, 12, 13, 14, 57, 59, 60). In addition, we analyzed the effects of synchrony on early visual processing by comparing conditions on mean amplitude of the visual N1 (100–200 ms) and later visual P400 (400–600 ms) over occipital sites (sites: 35, 37, 39). Statistical comparisons were made using paired-samples *t*-test for the mean amplitude averaged over all scalp sites within each group over the entire time window of interest for each experimental condition.[1]

Results

Auditory Response. Auditory processing was characterized by a very small negativity (N1) followed by a more prominent positive component peaking around 250 ms (auditory P2) over lateral scalp sites. A comparison of the mean amplitude of the auditory P2 revealed a significant difference over left, $t\,(20) = 2.30$, $p < .05$, and right, $t\,(20) = 3.09$, $p < .01$, lateral sites with the synchronous condition eliciting the greater amplitude positivity compared to the asynchronous condition (Fig. 2). This positivity was sustained over lateral sites for the duration of the trial.

Visual Response. Both synchronous and asynchronous trials elicited a prominent visual N1 shortly after visual stimulus onset characteristic of face processing in infancy (de Haan, 2007). An analysis of early visual processing for N1 revealed a significant difference between conditions, where the asynchronous condition produced more negative amplitude potentials compared to the synchronous condition, $t\,(20) = 2.65$, $p = .015$ (Fig. 2). Early processing was followed by a larger evoked positive component (P400) which peaked around 490 ms. A comparison of the mean amplitude of the P400 between the synchronous and the asynchronous conditions over occipital sites yielded no significant difference ($p = 0.83$).

Attentional Response. A comparison of the mean amplitude of Nc (500–700 ms) in response to synchronous and asynchronous conditions revealed a significant difference over fronto-central scalp sites, $t\,(20) = 2.26$, $p < .05$, where the asynchronous condition elicited a greater magnitude negativity (more negative) compared to the synchronous condition (Fig. 3).

Memory Processing. A comparison of the mean amplitude of the PSW revealed a significant difference over left, $t\,(20) = 2.93, p < .01$, and right, $t\,(20) = 3.31, p < .01$, scalp sites, with the synchronous condition eliciting a greater amplitude positivity during this time frame compared to the asynchronous condition (Fig. 3).

Discussion

The use of dynamic face/speech stimuli in Experiment 2 allowed us to control for some of the limitations of Experiment 1, like low level audio/visual sensory

[1]No significant peak latency differences were observed between experimental conditions for any of the components of interest in Experiment 2 (all *p*'s > .196).

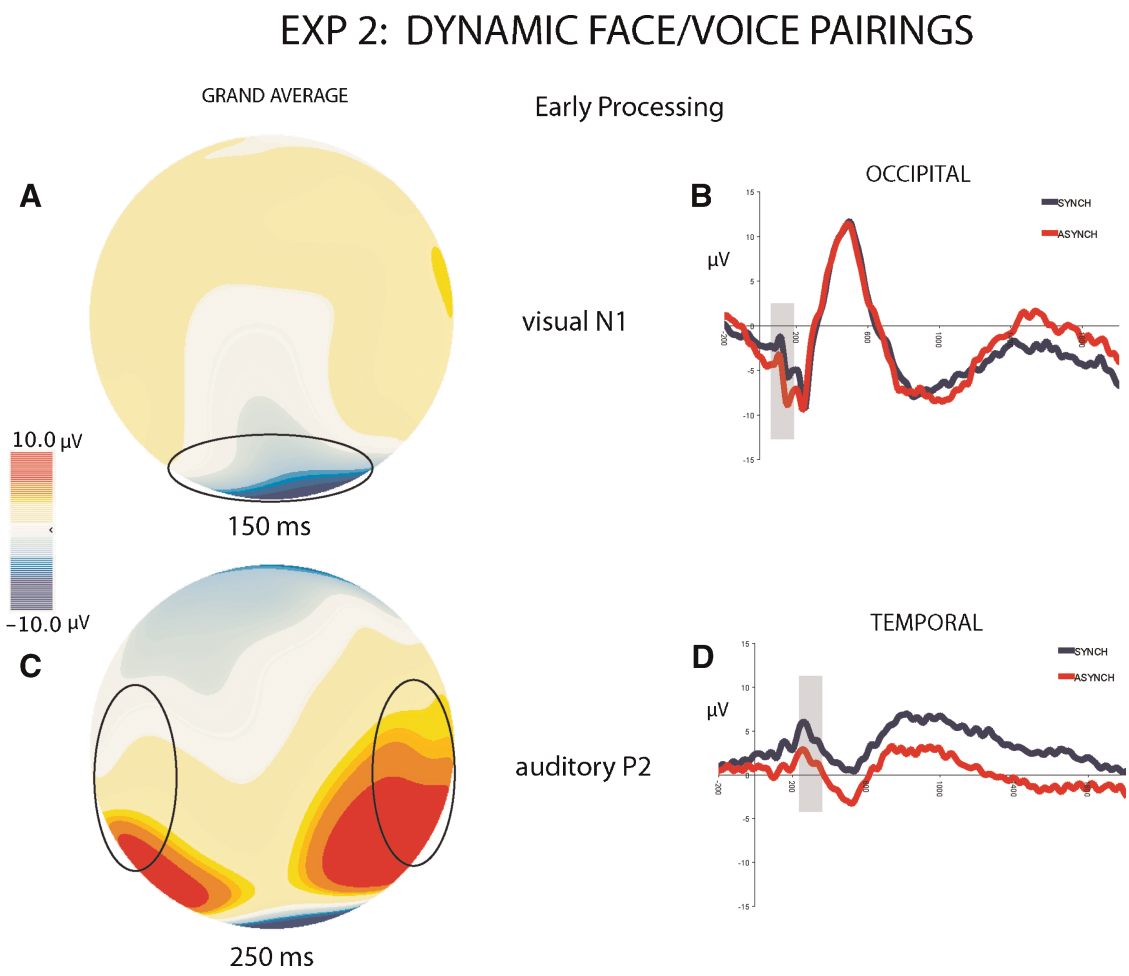

FIGURE 2 Summary of Experiment 2 early processing results. (A) Grand average scalp topography at 150 ms over occipital sites (average of all experimental conditions). (B) Average waveform from −200 to 2,000 ms averaged over occipital sites. The shaded region represents the comparison of experimental conditions for the visual N1 (100–200 ms). (C) Grand average scalp topography at 250 ms characterizing the auditory P2 over left and right temporal sites. (D) Average waveform from −200 to 2,000 ms averaged over left and right temporal sites. The shaded region represents the comparison of experimental conditions for the auditory P2. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

differences and stimulus timing across experimental conditions. Even after doing so, several of the results of Experiment 1 were replicated in Experiment 2. First, in both experiments the early auditory P2 response was greater for the synchronous compared to the asynchronous condition. Second, the later Nc response was greater for the asynchronous compared to the synchronous condition across both experiments. In addition, to replicating the findings of Experiment 1, Experiment 2 extended the results of Experiment 1 by allowing an analysis of visual processing. The analysis of visual processing revealed greater amplitude visually evoked potentials (N1) for the asynchronous condition during early processing and no difference in later visual processing between conditions. Finally, Experiment 2 was different from Experiment 1 in that Experiment 2 tested for the effects of rhythmical synchrony between dynamic faces and speech, whereas Experiment 1 tested for the effects of synchrony on the onset of a static face and voice. Consequently, a significant difference in the late PSW, with a greater positivity observed for the synchronous condition, was observed in Experiment 2 but not in Experiment 1. A comparison of the results from Experiments 1 and 2 are summarized in Table 1.

GENERAL DISCUSSION

Two experiments tested infants' neurophysiological response to face–voice synchrony. Experiment 1 tested

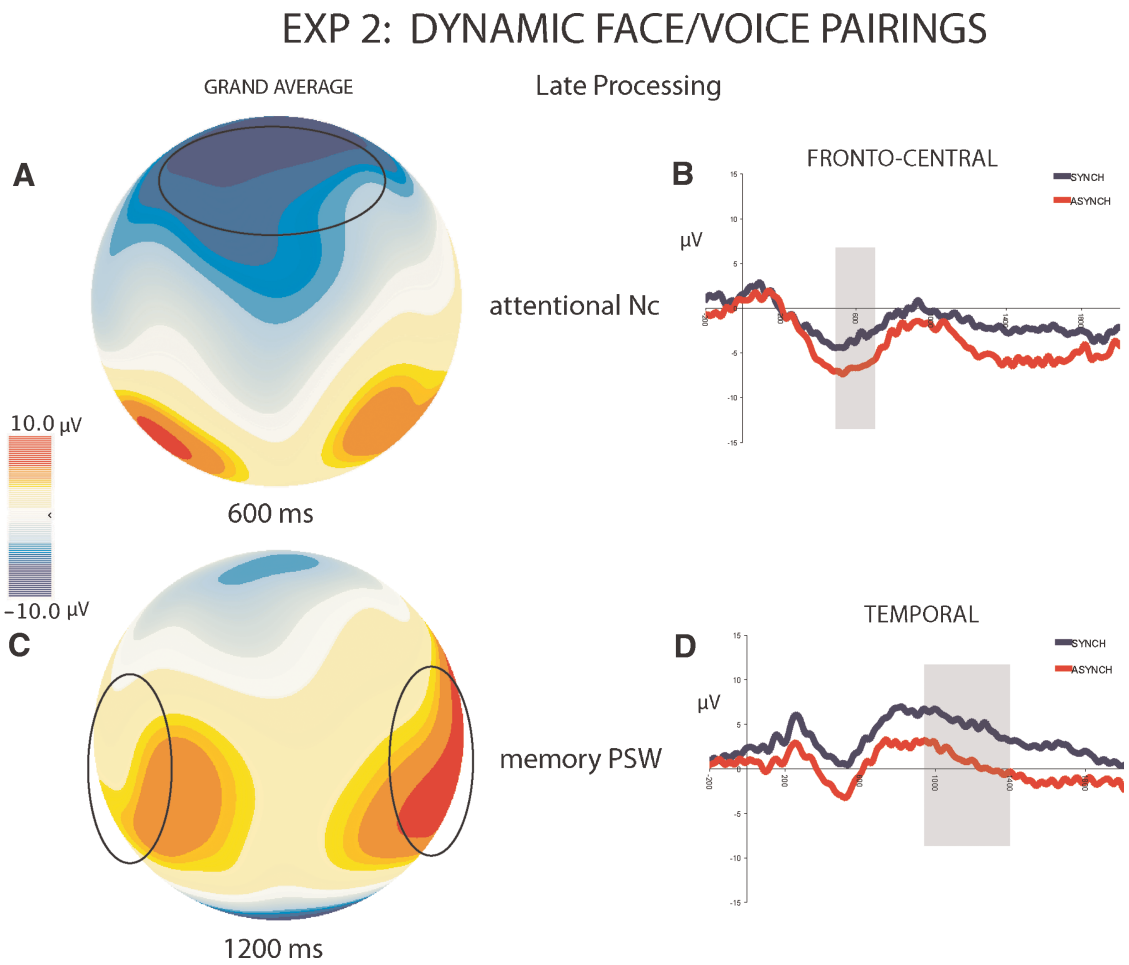

FIGURE 3 Summary of Experiment 2 late processing results. (A) Grand average scalp topography at 600 ms over fronto-central sites (average of all experimental conditions). (B) Average waveform from −200 to 2,000 ms averaged over fronto-central sites. The shaded region represents the comparison of experimental conditions for the Nc. (C) Grand average scalp topography at 1,200 ms characterizing the positive slow wave (PSW) over temporal sites. (D) Average waveform from −200 to 2,000 ms averaged over left and right temporal sites. The shaded region represents the comparison of experimental conditions for the PSW. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 1. Comparison of Results

	Experiment 1		Experiment 2	
Component	Synchronous	Asynchronous	Synchronous	Asynchronous
Auditory *P2*	left, $p = .036$ right, n.s.		left, $p = .033$ right, $p = .006$	
Visual *N1*	—	—		$p = .015$
Visual *P400*	—	—	n.s.	n.s.
Nc		$p = .010$		$p = .035$
PSW	n.s.	n.s.	left, $p = .008$ right, $p = .003$	

Significant effects are listed under the condition column that showed a greater brain response (synchronous or asynchronous). Cells containing "—" designate comparisons that were not able to be tested. Cells containing "n.s." indicate comparisons where no significant difference was observed.

the effects of onset synchrony and asynchrony between a static face and a voice and Experiment 2 tested the effects of synchrony and asynchrony between a moving face and speech. Results revealed both commonalities and differences in auditory, visual, attentional, and memory processing.

Early Auditory Processing of Synchrony

In adults it has been observed that synchronous AV speech produces attenuated early auditory-evoked potentials (N1/P2) compared to asynchronous AV speech or unimodal auditory speech (e.g., Pilling, 2009). In contrast, infants in our study showed the opposite pattern of results. Specifically, synchronous AV speech produced greater amplitude early auditory-evoked components compared to asynchronous speech. While the difference between asynchronous and synchronous AV speech suggests integration of information across the senses is occurring early in the processing stream for infants, the functional profile of the brain response (i.e., greater response for synchronous compared to asynchronous) suggests that the infant brain may respond differently to synchrony manipulations compared to what has been observed in adults. This is not unlikely given the previous developmental work showing substantial changes in intersensory perception over development (e.g., Flom, Whipple, & Hyde, 2009; Lewkowicz & Ghazanfar, 2009; Lewkowicz, Sowinski, & Place, 2008). For example, Lewkowicz (2010) argued that exposure to asynchronous events causes young infants to focus more on the temporal relationship between the audio and visual input, where similar exposure in adulthood begins to bias the observer to see the event as synchronous. In general, it seems the difference between the results of our study and studies with adults likely reflects an initial bias towards detection of synchronous audio-visual information in infancy.

Early Visual Processing of Synchrony

Unfortunately, the effects of audio-visual synchrony on visual processing are not well understood in infants or adults. In our study, we found that early visually evoked potentials were greater in amplitude for asynchronously presented stimuli compared to synchronously presented stimuli. One possibility that we cannot rule out is that the components we label as ''auditory'' and ''visual'' processing actually reflect early multisensory processing and thus originate from a common multisensory region rather than separate primary sensory cortical regions. This is not entirely implausible given that the similarity in timing and reverse polarity of the N1 and P2. Regardless of their origin, however, these results further support the notion that sensory integration begins early in the processing stream during preliminary sensory processing rather than later conceptual processing.

Attentional Response to Synchronous and Asynchronous Events

In both studies, we found a greater magnitude Nc for the asynchronous condition compared to the synchronous condition. We interpret this finding as potential evidence that infants found the asynchronous condition more interesting or unfamiliar, causing them to attend or orient to it more than the synchronous condition. This interpretation follows from extensive work on the attentional response of infants to familiar and unfamiliar stimuli, showing unfamiliar stimuli evoke a larger Nc response compared to familiar stimuli (de Haan, 2007; Nelson & Collins, 1991, 1992; Reynolds & Richards, 2005) and parallels behavioral work showing asynchronous events elicit increased attention in young infants (Lewkowicz et al., 2008).

Processing of Synchrony in Memory

Dynamic, but not static, stimuli elicited a PSW that was greater for the synchronous condition compared to the asynchronous condition. We interpret this difference as the synchronous condition being retrieved and updated in long-term memory where the asynchronous condition was not or was to a lesser degree. The difference between experiments may have resulted from the ecological validity of the stimuli; infants may be more familiar and able to recognize from memory dynamic face/voice pairings, where static face/voice pairings may not be recognized. If this interpretation is correct, our findings on the PSW confirm the interpretation of the Nc that synchronous stimuli is more familiar to infants than asynchronous stimuli.

Neural Basis of Audio-Visual Synchrony Perception

Two distinct networks of brain regions have been associated with the temporal integration audio-visual speech in previous literature: one network for the fused perception of speech and another network for the detection of temporal discrepancies between audio and visual information. The fusion of audio and visual information to perceive speech has been associated with activation in superior temporal sulcus (STS), superior temporal gyrus (STG), the intraparietal sulcus (IPS), Heschl's gyrus, and the inferior frontal gyrus (Calvert, 2001; Miller & D'Esposito, 2005; Wright, Pelphrey, Allison, McKeown, & McCarthy, 2003). Furthermore, some have proposed that of all these regions that seem to respond to the integration of information, only the superior temporal

regions, in fact, selectively respond to synchronous speech (Calvert, 2001; Doesburg, Emberson, Rahi, Cameron, & Ward, 2008; Wright et al., 2003). In contrast, the detection of asynchrony between audio and visual information has been associated with activity in distinct posterior parietal, insular, prefrontal, and cerebellar areas (Jones & Callan, 2003; Kaiser, Hetrich, Ackermann, Mathiak, & Lutzenberger, 2005; Miller & D'Esposito, 2005). Furthermore, a recent EEG study with adults showed contrasting signatures for synchronous and asynchronous bimodal speech, further suggesting distinct networks underlie the integration of information across the senses to perceive multisensory speech (Doesburg et al., 2008).

We observed greater evoked responses to synchrony during early auditory processing and greater evoked responses to asynchrony during visual, attentional, and memory processing. One possibility is that the components we label as early auditory and visual processing may, in fact, arise from multisensory cortical regions. Although purely speculative, it may be the case that our early auditory evoked potentials showing greater processing for synchrony originate from the temporal system that detects synchrony, whereas the visual and later slow wave activity showing greater amplitude potentials for the asynchronous condition originate from the brain system or structures sensitive to asynchronous audio-visual relations. Additional work would be needed to confirm this speculation. Nonetheless, early auditory processing in adults seems to be different, as the asynchronous condition elicits greater amplitude electrophysiological responses than the synchronous condition and may be engaging the asynchrony detection system rather than the synchrony detection system (Pilling, 2009; van Wassenhove et al., 2005). One important direction for future work would be to investigate how the interaction of these brain systems changes over development, given the developmental changes and biases toward processing certain features of audio-visual event over development.

Conclusion

Over the past few decades behavioral work has made substantial progress in elucidating the intersensory perceptual capacities of infants, the conditions under which infants are able to integrate information across the senses, the properties important to integration, and changes in the perceptual basis of integration over development. The purpose of this study was to build on this foundation to understand how this process is carried out in the brain. In doing so, we were able to provide a functional link between behavioral theory on looking behavior and neural processes. In particular, we observed neural responses indicative of early biases towards the detection of synchrony in young infants, later increased attention to asynchronously presented stimuli, and later recognition memory for the synchronous condition. All of these interpretations fit nicely with proposed processing, attentional, and memory biases in 5-month-old infants. The limitation of behavioral data is that all of these processes are combined into one measure, infant looking. The promise of our findings is that researchers can now devise experimental manipulations to examine each of these processes in isolation to discover the contribution each makes to intersensory perception and the changes that occur over development.

NOTES

This research was supported by the Family Studies Center and the School of Family Life at Brigham Young University. We also thank Joan Leishman, Sarah Ahlander, Rebecca Lawson, Jacob Christiansen, Ross Mangum, Clark Van Den Berghe, and Holly Montgomery, for their valuable assistance with the collection and preparation of the data used in this study.

REFERENCES

Ackles, P. K., & Cook, K. G. (1998). Stimulus probability and event-related potentials of the brain in 6-month-old human infants: A parametric study. International Journal of Psychophysiology, 29(2), 115–143.

Bahrick, L. E. (1992). Infants' perceptual differentiation of amodal and modality-specific audio-visual relations. Journal of Experimental Child Psychology, 53, 180–199.

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

Bahrick, L. E. (2001). Increasing specificity in perceptual development: Infants' detection of nested levels of multimodal stimulation. Journal of Experimental Child Psychology, 79, 253–270.

Bahrick, L. E., Hernandez-Reif, M., & Flom, R. (2005). The development of infant learning about specific face–voice relations. Developmental Psychology, 41(3), 541–552.

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

Bahrick, L. E., Netto, D., & Hernandez-Reif, M. (1998). Intermodal perception of adult and child faces and voices by infants. Child Development, 69(5), 1263–1275.

Bahrick, L. E., Walker, A., & Neisser, U. (1981). Selective looking by infants. Cognitive Psychology, 13(3), 377–390.

Besle, J., Fort, A., Delpuech, C., & Giard, M. (2004). Bimodal speech: Early suppressive visual effects in human auditory cortex. European Journal of Neuroscience, 20(8), 2225–2234.

Braida, L. D. (1991). Crossmodal integration in the identification of consonant segments. The Quarterly Journal of Experimental Psychology, 43(3), 647–677.

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J. (2009). Hearing faces: How the infant brain matches the face it sees with the speech it hears. Journal of Cognitive Neuroscience, 21(5), 905–921.

Butterfield, E. C., & Siperstein, G. N. (1970). Influence of contingent auditory stimulation upon non-nutritional suckle. In: J. F. Bosma (Ed.), Third Symposium on Oral Sensation and Perception: The mouth of the infant (pp. 313–334). Springfield, IL: Charles C. Thomas.

Calvert, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. Cerebral Cortex, 11(12), 1110–1123.

M. de Haan (Ed.). (2007). Infant EEG and event-related potentials. London: Psychology Press.

de Haan, M., & Nelson, C. A. (1997). Recognition of the mother's face by six-month-old infants: A neurobehavioral study. Child Development, 68, 187–210.

Dodd, B. (1979). Lip reading in infants: Attention to speech presented in-and out-of-synchrony. Cognitive Psychology, 11(4), 478–484.

Doesburg, S. M., Emberson, L., Rahi, A., Cameron, D., & Ward, L. M. (2008). Asynchrony from synchrony: Gamma-band neural synchrony and perception of audiovisual speech asynchrony. Experimental Brain Research, 185(1), 11–20.

Elsabbagh, M., Volein, A., Csibra, G., Holmboe, K., Garwood, H., Tucker, L., Krljes, S., Baron-Cohen, S., Bolton, P., Charman, T., Baird, G., & Johnson, M. H. (2009). Neural correlates of eye gaze processing in the infant broader Autism phenotype. Biological Psychiatry, 65, 31–38.

Erber, N. P. (1975). Auditory–visual perception of speech. Journal of Speech and Hearing Disorders, 40, 481–492.

Flom, R., Whipple, H., & Hyde, D. (2009). Infants' intermodal perception of canine (canis familairis) facial expressions and vocalizations. Developmental Psychology, 45, 1143–1151.

Frank, M. C., Slemmer, J. A., Marcus, G. F., & Johnson, S. P. (2009). Information from multiple modalities helps 5-month-olds learn abstract rules. Developmental Science, 12(4), 504–509.

Gogate, L. J. (2010). Learning of syllable-object relations by preverbal infants: The role of temporal synchrony and syllable distinctiveness. Journal of Experimental Child Psychology, 105(3), 178–197.

Gogate, L. J., Prince, C. G., & Matatyaho, D. J. (2009). Two-month-old infants' sensitivity to changes in syllable-object pairings: The role of temporal synchrony. Journal of Experimental Psychology: Human Perception and Performance, 35, 508–519.

Goldman, D. Z., Shapiro, E. G., & Nelson, C. A. (2004). Measurement of vigilance in 2-year-old children. Developmental Neuropsychology, 25(3), 227–250.

Grant, K. W., & Seitz, P. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. The Journal of the Acoustical Society of America, 108(3), 1197–1208.

Green, K. P. (1998). The use of auditory and visual information during phonetic processing: Implications for theories of speech perception. In: R. Campbell, B. Dodd, & D. Burnham (Eds.), Hearing by Eye II: Advances in the Psychology of Speechreading and Auditory-visual Speech. Psychology Press Ltd., East Sussex, UK, 3–25.

Grossmann, T., Gliga, T., Johnson, M. H., & Mareschal, D. (2009). The neural basis of perceptual category learning in human infants. Journal of Cognitive Neuroscience, 21(12), 2276–2286.

Grossmann, T., Striano, T., & Friederici, A. D. (2006). Crossmodal integration of emotional information from face and voice in the infant brain. Developmental Science, 9(3), 309–315.

Hyde, D. C., Jones, B. L., Porter, C. L., & Flom, R. (2010). Visual stimulation enhances auditory processing in 3-month-old infants and adults. Developmental Psychobiology, 52(2), 181–189.

Hyde, D. C., & Spelke, E. S. Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition. To appear in *Developmental Science,* no. doi:10.1111/j.1467-7687.2010.00987.x

Izard, V., Dehaene-Lambertz, G., & Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. PLos Biology, 6(2), 0275–0285.

Johnson, M. H., Dzurawaec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. Cognition, 40(1–2), 1–19.

Jones, J. A., & Callan, D. E. (2003). Brain activity during audiovisual speech perception: An fMRI study of the McGurk effect. NeuroReport, 14, 1129–1133.

Kaiser, J., Hertrich, I., Ackermann, H., Mathiak, K., & Lutzenberger, W. (2005). Hearing lips: Gamma-band activity during audio-visual speech perception. Cerebral Cortex, 15, 646–653.

Kuhl, P. K., & Meltzoff, A. N. (1982). The bimodal perception of speech in infancy. Science, 218(4577), 1138–1141.

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory speech percept in human infants. PNAS, 105, 11442–11445.

Lewkowicz, D. J. (1992a). 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.

Lewkowicz, D. J. (1992b). Infants' responsiveness to auditory and visual components of a sounding/moving compound stimulus in human infants. Perception and Psychophysics, 52(5), 519–528.

Lewkowicz, D. J. (1996). Perception of auditory–visual temporal synchrony in human infants. Journal of Experimental Psychology: Human Perception and Performance, 22, 1094–1106.

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

Lewkowicz, D. (2010). Infant perception of audio-visual speech synchrony. Developmental Psychology, 46(1), 66–77.

Lewkowicz, D. J., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. Trends in Cognitive Sciences, 13, 470–478.

Lewkowicz, D. J., & Kraebel, K. S. (2004). The value of multisensory redundancy in the development of intersensory perception. In: G. A. Calvert, C. Spence, & B. E. Stein (Eds.), The handbook of multisensory processes (pp. 655–678). Cambridge, MA: MIT Press.

Lewkowicz, D., Sowinski, R., & Place, S. (2008). The decline of cross-species intersensory perception in human infants: Underlying mechanisms and its developmental persistence. Brain Research, 1242, 291–302.

Massaro, D. W. (1987). Speech perception by ear and eye: A paradigm for psychological inquiry. Hillsdale, NJ: Erlbaum.

Massaro, D. W. (1998). Perceiving talking faces: From speech perception to a behavioral principle. Cambridge, MA: MIT Press.

Miller, L. M., & D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. Journal of Neuroscience, 22, 5884–5893.

Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. Psychological Bulletin, 12, 826–859.

Nelson, C. A. (1994). Neural correlates of recognition memory in the first postnatal year of life. In: G. Dawson, & K. Fischer (Eds.), Human development and the developing brain (pp. 269–313). New York: Guilford Press.

Nelson, C. A. (1996). Electrophysiological correlates of early memory development. In: H. W. Reese, & M. D. Franzen (Eds.), Thirteenth West Virginia University Conference on Life Span Development Psychology: Biological and neurological mechanisms (pp. 95–131). Hillsdale, NJ: Erlbaum.

Nelson, C. A. (1998). The nature of early memory. Preventive Medicine, 27, 172–179.

Nelson, C. A., & Collins, P. F. (1991). Event-related potential and looking-time analysis of infants' responses to familiar and novel events: Implications for visual-recognition memory. Developmental Psychology, 27, 50–58.

Nelson, C. A., & Collins, P. F. (1992). Neural and behavioral correlates of recognition memory in 4- and 8-month-old infants. Brain and Cognition, 19, 105–121.

Quinn, P. C., Westerlund, A., & Nelson, C. A. (2006). Neural markers of categorization in 6 month old infants. Psychological Science, 17, 59–66.

Pilling, M. (2009). Auditory event-related potentials (ERPs) in audiovisual speech perception. Journal of Speech, Language, and Hearing Research, 52(4), 1073–1081.

Reynolds, G. D., Courage, M. L., & Richards, J. E. (2010). Infant attention and visual preferences: Converging evidence from behavior, event-related potentials and cortical source localization. Developmental Psychology, 46, 886–904.

Reynolds, G. D., & Richards, J. E. (2005). Familiarization, attention, and recognition memory in infancy: An event-related potential and cortical source localization study. Developmental Psychology, 41, 598–615.

Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. Developmental Science, 6(3), 233–241.

Scott, L. S., & Nelson, C. A. (2006). Featural and configural face processing in adults and infants: A behavioral and electrophysiological investigation. Perception, 35(8), 1107–1128.

Shulman, G. L., Corbetta, M., Buckner, R. L., Raichle, M. E., Fiez, J. A., Miezin, F.M., & Peterson, S. E. (1997). Top-down modulation of early sensory cortex. Cerebral Cortex, 7, 193–206.

Spelke, E. S., Born, W. S., & Chu, F. (1983). Perception of moving, sounding objects by four-month-old infants. Perception, 12(6), 719–732.

Sumby, W. H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. Journal of the Acoustical Society of America, 26, 212–215.

Summerfield, A. Q. (1979). Use of visual information in phonetic perception. Phonetica, 36, 314–331.

van Wassenhove, V., Grant, K. W., & Poeppel, D. (2005). Visual speech speeds up the neural processing of auditory speech. Proceeding of the National Academy of Sciences of the United States of America, 102, 1181–1186.

Vouloumanos, A., & Werker, J. F. (2007). Listening to language at birth: Evidence for a bias for speech in neonates. Developmental Science, 10(2), 159–164.

Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J., & McCarthy, G. (2003). Polysensory interaction along temporal regions evoked by audiovisual speech. Cerebral Cortex, 13, 1034–1043.