Received: 7 February 2023 | Revised: 23 June 2023 | Accepted: 21 September 2023

DOI: 10.1002/dev.22431

RESEARCH ARTICLE

Developmental Psychobiology WILEY

The effect of face orientation on audiovisual speech integration in infancy: An electrophysiological study

Magdalena Szmytke[1] | Dianna Ilyka[2] | Joanna Duda-Goławska[3] | Zuzanna Laudańska[3] | Anna Malinowska-Korczak[3] | Przemysław Tomalski[3]

[1]Faculty of Psychology, University of Warsaw, Warsaw, Poland

[2]Department of Psychology, University of Cambridge, Cambridge, UK

[3]Neurocognitive Development Lab, Institute of Psychology, Polish Academy of Sciences, Warsaw, Poland

Correspondence
Magdalena Szmytke, Faculty of Psychology, University of Warsaw, Stawki 5/7, 00-183 Warszawa, Poland.
Email: szmytkemagdalena@gmail.com

Przemysław Tomalski, Neurocognitive Development Lab, Institute of Psychology, Polish Academy of Sciences, Stefana Jaracza 1, 00-378 Warszawa, Poland.
Email: tomalski@mac.com

Funding information
Polish National Science Centre, Grant/Award Number: 2016/23/B/HS6/03860; Institute of Psychology, Polish Academy of Sciences

Abstract

Humans pay special attention to faces and speech from birth, but the interplay of developmental processes leading to specialization is poorly understood. We investigated the effects of face orientation on audiovisual (AV) speech perception in two age groups of infants (younger: 5- to 6.5-month-olds; older: 9- to 10.5-month-olds) and adults. We recorded event-related potentials (ERP) in response to videos of upright and inverted faces producing /ba/ articulation dubbed with auditory syllables that were either matching /ba/ or mismatching /ga/ the mouth movement. We observed an increase in the amplitude of audiovisual mismatch response (AVMMR) to incongruent visual /ba/-auditory /ga/ syllable in comparison to other stimuli in younger infants, while the older group of infants did not show a similar response. AV mismatch response to inverted visual /ba/-auditory /ga/ stimulus relative to congruent stimuli was also detected in the right frontal areas in the younger group and the left and right frontal areas in adults. We show that face configuration affects the neural response to AV mismatch differently across all age groups. The novel finding of the AVMMR in response to inverted incongruent AV speech may potentially imply the featural face processing in younger infants and adults when processing inverted faces articulating incongruent speech. The lack of visible differential responses to upright and inverted incongruent stimuli obtained in the older group of infants suggests a likely functional cortical reorganization in the processing of AV speech.

KEYWORDS

audiovisual mismatch response (AVMMR), audiovisual speech processing, EEG/ERP, face inversion effect, face processing, McGurk effect

1 | INTRODUCTION

Humans are exposed to speech sounds already in the womb. However, when they are born, their language experience is not purely auditory given that speech sounds are normally accompanied by faces and mouth movements (Weikum et al., 2007). Visual speech not only facilitates language acquisition but also activates the auditory cortex (Sams et al., 1991); thus, it is possible that speech may have bi- or multimodal neural representations early in development (Guellai et al., 2014; Kushnerenko et al., 2008; Teinonen et al., 2008).

One of the examples of bimodal speech representation is the McGurk effect, in which an auditory syllable is dubbed onto a visually mismatching syllable (speaking lips do not match the speech sound), for example, auditory (A) /ba/ onto visual (V) /ga/ (McGurk & MacDonald, 1976). Although auditory and visual cues are conflicting, the brain assimilates them, which results in the perception of the closest legal

Abbreviations: AV, audiovisual; AVMMR, audiovisual mismatch response; AVSI, audiovisual speech integration; EEG, electroencephalography; ERP, event-related potential.

phoneme—in this case, /da/. Different pairs of syllables can result in other types of illusions, not necessarily producing fused and coherent percepts. For example, the same syllables in the reversed compound (auditory /ga/ and visual /ba/) typically result in the illusory perception of combination /bga/, which in many languages, including Polish and English, is considered illegal.

The audiovisual speech integration (AVSI) in infants has been widely tested using the McGurk effect (Burnham & Dodd, 2004; for a review, see Tomalski, 2015), and studies showed that preverbal infants could detect the incongruence between auditory and visual stimuli (Lewkowicz, 2010). The visual cues from the articulation affect speech perception, support phonological learning and comprehension (Hazan et al., 2005), and enhance it under noisy conditions (Grant & Seitz, 2000; Teinonen et al., 2008). However, facial information facilitates speech perception only in an upright configuration. Previous adult studies reported that the McGurk illusion is mitigated by face inversion when configural facial information is disrupted (Hietanen et al., 2001; Massaro & Cohen, 1996). In addition, the interdependence between these two processes has been shown in an electrophysiological study using McGurk stimuli and face inversion. It showed that face configuration affects speech perception, resulting in a change detection only when the face is in the upright orientation (Eskelund et al., 2015). Thus, the configural aspect of face processing may influence the processing of audiovisual (AV) speech information, especially under incongruent conditions (Jordan & Bevan, 1997). These findings raise questions about how this overlapping interdependence between AVSI and face processing unfolds throughout the first year of life (Rosenblum et al., 2000).

Newborns show a basic preference for upright faces and face-like patterns (Farroni et al., 2005; Johnson et al., 1991); however, face processing undergoes significant changes in infancy (for a review, see Johnson et al., 2015). With age and experience, infants start to prefer faces of their own race (Kelly et al., 2007) as well as faces in the upright rather than inverted orientation. Changes in face processing occur also as a shift from featural to configural processing. A habituation study indicated that 4-month-olds process eyes and mouth featurally, whereas 10-month-olds integrate eyes and mouth into a whole face (Schwarzer et al., 2007). Extensive electrophysiological research on face processing shows that from the age of 3 months, infants differentiate between upright and inverted faces (de Haan et al., 2002; Halit et al., 2003; Peykarjou & Hoehl, 2013). From the age of 6 months, their neural responses to upright facial configuration become more pronounced (see de Haan et al., 2002), and it is shown that by the age of 7–9 months, they process faces configurally (Cohen & Cashon, 2001). These processes are likely the result of experience-dependent specialization in the processing of information that is present in an infant's environment. This specialization (also known as perceptual narrowing) is not limited to auditory (e.g., native phonemes; Werker & Tees, 1984) or visual processing (Kelly et al., 2007; Krasotkina et al., 2021; Weikum et al., 2007), but may occur to some extent in all modalities. Findings that speech and face processing share neural (Belin et al., 2011; Pascalis et al., 2014) and developmental mechanisms (Krasotkina et al., 2018, 2021), which emerge around the same time, support the hypothesis that specialization might be a modality-general, multisensory process (Lewkowicz & Ghazanfar, 2009; Ujiie et al., 2020; for a review, see Maurer & Werker, 2014). In the present study, we took infants' age as a proxy for their level of face specialization. As such, we tested two groups of infants (younger, aged 5–6.5 months; older, aged 9–10.5 months), which according to previous literature should differ in their processing of upright versus inverted faces.

The integration of AV speech stimuli has previously been tested behaviorally as well electrophysiologically in infants. Behavioral studies reported that integration between auditory and visual information is possible (Burnham & Dodd, 2004) from 4.5 months of age but not mandatory (Desjardins & Werker, 2004; Desjardins et al., 1997). On the neural level, AV speech perception, just like face processing, undergoes a developmental transition around 6–8 months of age (Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013). Specifically, previous infant studies that have used electroencephalography (EEG) in the McGurk paradigm have shown that perception of VbaAga (illegal combination) in 5-month-olds elicits an event-related component sensitive to the bimodal incongruence, the audiovisual mismatch response (AVMMR). The presence of AVMMR supports the view that early experience with visual speech may have an impact on speech perception skills (Kushnerenko et al., 2008). Importantly, in a group of infants aged 6–9 months, the AVMMR was observed only in those infants who showed less looking to the mouth relative to the eyes (Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013). Using both EEG and eye-tracking methods, these researchers showed that infants with a more mature pattern of visual scanning (increased attention to the mouth area relative to the eyes) did not have the AVMMR in response to the VbaAga stimulus suggesting that AVMMR may be transient and indexes the lack of integration between signals coming from two modalities. In younger, 5-month-old infants and those showing decreased attention to the mouth, the AVMMR reflects a less mature pattern of processing of AV speech information (Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013). Given this complex developmental trajectory, the use of AVMMR may index how information from different modalities is integrated at different stages of infant development, while infants accumulate experience and build proficiency in speech processing.

Kushnerenko, Tomalski, Ballieux, Ribeiro, et al. (2013) were first to demonstrate that the AVMMR disappears in typical development as AV speech processing matures between 6 and 9 months of age. On the basis of their results, they hypothesized that the AVMMR is a transitory component and its diminishing reflects neural reorganization of AV speech processing in late infancy. This hypothesis is further supported by our recent functional near-infrared spectroscopy (fNIRS) results, which investigated neural correlates of AV integration of speech cues in similar age groups, but in a different paradigm. Dopierała et al. (2023) found differences in the spatial distribution of neural responses to bimodal (audiovisual) versus alternating unimodal (auditory + visual) syllables in 5- and 10-month-olds (Dopierała et al., 2023). Since responses were widely distributed across the cortex, they used multivariate pattern analysis (MVPA) to identify areas, which selectively responded to stimuli, which require AV speech integration (bimodal synchronized AV syllables). MVPA correctly classified responses at 5 months of age, with key input from inferior frontal and superior temporal channels of the right hemisphere. Surprisingly, MVPA classification was not successful at 10 months of age in identifying selective channels for bimodal AV syllables. This lack of successful classification is consistent with previous studies from another method (EEG/event-related potentials [ERPs]), which showed diminishing AV mismatch response by the age of 9 months (Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013). Both these studies suggest that around the age of 9–10 months, there is functional cortical reorganization in the perception of AV speech occurring in typical development (for further discussion, see Dopierała et al., 2023).

Our current question is focused on investigating the sources of this reorganization in terms of emerging face processing skills in late infancy. As previously mentioned, the developmental timing of the AVMMR disappearance seems to coincide with the emergence of configural face processing, raising the possibility of shared neural and developmental mechanisms of speech and face perception as suggested by earlier studies (Belin et al., 2011; Krasotkina et al., 2018, 2021; Pascalis et al., 2014). Therefore, to better understand the relations between the specialization for speech and face processing, we investigated a group of infants (aged 5–6.5 months) at the onset of the postulated period of configural face processing (before the age of 7 months; Cohen & Cashon, 2001). We tested another, older, group (aged 9–10.5 months) at a stage when specialization for upright faces has likely already emerged. The older group was also expected to have more mature AVSI, as suggested by a previous electrophysiological study (Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013). To measure AV speech integration, we used the McGurk stimuli (congruent VbaAba and incongruent nonfusible VbaAga) and measured the amplitude of the AVMMR component as its index. To investigate the relationship between developmental changes in AV speech integration and developmental changes in configural face processing, these stimuli were presented in the upright and the inverted orientation.

The intertwining of configural face processing and AV speech conflict has already been studied using an oddball paradigm in adults (Eskelund et al., 2015). However, the interdependence of AV speech integration and configural face processing was not tested when both processes emerge: during the first year of life. Our main goal was to test whether the presence of the AVMMR is modulated by face inversion in infants at different stages of the process of face specialization. We propose that differential responses to incongruent upright versus inverted faces would reflect the intertwining of those two mechanisms. If these processes are in some way interconnected from early on, then we hypothesize that only the younger group of infants would process the AV mismatch, regardless of the face orientation. The growing proficiency in the processing of upright faces might affect AVSI; thus, we studied an older group (9- to 10.5-month-olds) to test if their electrophysiological responses resemble more adultlike patterns, assuming that they process upright faces configurally (Cohen & Cashon, 2001). For control purposes, we also studied a group of adults to test if face inversion affects their AVSI. Previous findings showed that the frontocentral AVMMR in response to AV mismatch was not observed in adult participants (Control Study S3 in Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013); thus, it was suggested that AVMMR is a developmentally transient ERP component. Therefore, we wanted to test whether face inversion will result in a differential response to upright versus inverted AV mismatch, as no study has tested those two manipulations in a single paradigm. We hypothesize that proficiency in the processing of upright faces facilitates the integration of conflicting AV speech cues in the VbaAga mismatch conditions, thus no AVMMR would be observed in adults. However, the less proficient AV processing when observing AV mismatch in inverted faces will not result in the integration of conflicting AV stimuli and adults will show AVMMR for inverted faces.

2 | METHODS

2.1 | Participants

2.1.1 | Infants

The final sample consisted of 42 infants in two age groups: 20 infants (seven girls) at the age of 5–6.5 months ($M = 5.74$; $SD = 0.41$) and 22 infants (eight girls) at the age of 9–10.5 months ($M = 9.88$; $SD = 0.44$). Additional 19 infants were tested but excluded from analysis due to preterm birth ($N = 1$), bad electrode signal ($N = 3$), excessive movements and failure to reach the criterion of minimum of 20 segments per condition ($N = 12$), and refusal to keep on an EEG net ($N = 3$). All infants came from monolingual Polish-speaking families and were born full-term (37–42 gestational age). Families of the infants were asked questions about the vision/hearing deficits, diagnosis of any type of disorders, for example, autism spectrum disorder, major complications during pregnancy and delivery, and maternal age to control for other influencing factors.

The study was approved by the Research Ethics Committee at the Faculty of Psychology, University of Warsaw, Poland, and conformed to the standards of the Declaration of Helsinki. Prior to the testing, all parents gave written informed consent. As a thank-you gift for their participation, the families received a diploma, a small baby book, and a video recording of their play in the laboratory.

2.1.2 | Adults

The adult group consisted of 20 participants (women, $N = 15$), aged between 21 and 43 years ($M = 26.26$; $SD = 5.14$). All participants reported that they were free of neurological disorders and had normal or corrected-to-normal visual ability. The group of adults was tested for control purposes. All participants were volunteers who were not paid for participation and had signed a consent form before participation in the study.

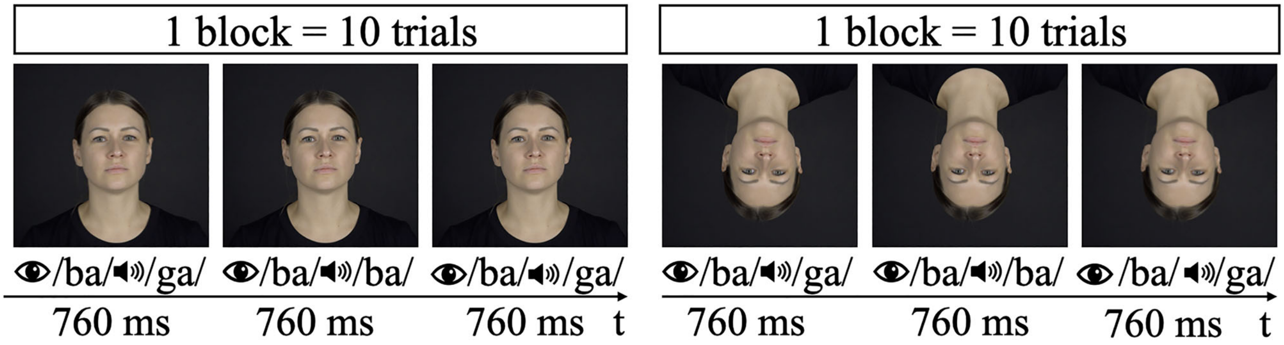

FIGURE 1 Schematic diagram of the procedure. Each block consisted of five congruent and five incongruent trials presented in pseudorandom order.

2.2 | Facilities and equipment

EEG signals were recorded in Netstation 4.2 software and an EGI NetAmps200 amplifier with 64-channel water-based electrolyte Hydrocel Sensor Nets (EGI, Inc.) referenced to vertex (Cz). Stimuli were displayed on a 24″ monitor (1920 × 1200 pixels) with a 60-Hz refresh rate. Auditory stimuli were presented with two symmetrically spaced speakers (embedded in the monitor) at the level of approximately 65 dB SPL. The session was recorded with two remote-controlled CCTV color cameras in HD quality to code infant-looking behavior during EEG.

2.3 | Experimental design and stimuli

Video recordings (frontal view) of three female professional actresses (native Polish speakers) articulating isolated /ba/ and /ga/ sounds were recorded in a professional audio recording studio. AV stimuli consisted of 760-ms-long videos of faces producing syllables with sound onset at 360 ms after the stimulus appearance on the screen. To generate two types of stimuli, the congruent sound stimulus /ba/ as well as incongruent /ga/ stimulus were dubbed onto a visual representation of syllable /ba/ (see Figure S2 for a schematic illustration of an incongruent upright stimulus). The combination of visual /ba/ and auditory /ga/ typically leads to AV mismatch perceived as /bga/ or /baga/ as was reported in adults (McGurk & MacDonald, 1976).

The created dynamic videos were displayed in two configurations: upright or inverted (stimuli created from the original stimuli by rotating them 180° upside-down). The digitization rate of the video clips was 25 frames per second, and the stereo sound digitization was 44.1 kHz with 16-bit resolution. Each block was displayed in one of two face conditions—upright or inverted—containing 10 repetitions of the visual syllable accompanied by auditory stimuli (pseudorandomized presentation of /ba/ or /ga/; see Figure 1 for the schematic diagram of the procedure). The procedure contained a maximum of 42 blocks; however, in most cases, the procedure was terminated earlier, due to a lack of infants' interest or fussiness. In the present study, four types of AV stimuli were presented with equal probability. During testing, infants sat approximately 65 cm away from the screen, with the monitor position centered on the infant's eye level. The faces were centered on the middle of the screen, approximately life-size, and covered 16° visual angle in width.

2.4 | Testing procedure

After the parent's written consent, the infant was seated on the parent's lap in a dimly lit room. Before the experiment, an animation of *Elmo's World* was displayed on the screen to grab the infants' attention and/or a research assistant was talking to the infant to distract him/her from the researcher who was placing the net on the infant's head.

After the net correct placement, the parent was instructed not to speak to the infant and not to distract him/her from the stimuli on the screen. The experiment lasted 7 min or until the infant lost interest in the stimuli or became excessively fussy.

2.5 | EEG signal preprocessing

2.5.1 | Infant data

The data were preprocessed with a standard procedure (Kushnerenko et al., 2008) using EGI Netstation 4.2 software. The EEG signal was amplified, digitized at 500 Hz, and band-pass filtered from 0.1 to 200 Hz. Continuous EEG recordings were off-line low-pass filtered at 30 Hz and segmented into 1100-ms-long epochs. Each segment included 760-ms-long stimulus, a period of 100 ms prior to stimulus onset and 240 ms after stimulus offset. Channels contaminated by eye or motion artifacts were rejected manually, and segments with more than 10 bad channels were excluded. In addition, video recordings of the infants' behavior were coded frame by frame, and segments during which the infant did not attend to the stimulus on the screen were excluded from further analysis. The remaining marked channels were replaced via trial-by-trial channel interpolation. Artifact-free segments were re-referenced to the overall average and then averaged for each infant within each condition. A baseline correction was performed by subtracting mean amplitudes in the 260–360 ms window from the stimulus onset (i.e., the period of 100 ms immediately before the sound onset). The average number of segments per condition contributed by individuals had to reach a minimum of 20 (criterion based on previous studies of Kushnerenko et al., 2008; Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013) for the participant to be included in the analysis: upright VbaAba ($M = 29.45$; $SD = 12.93$), inverted VbaAba ($M = 28.25$; $SD = 11.63$), upright VbaAga ($M = 29.50$; $SD = 11.12$), and inverted VbaAba ($M = 28.80$; $SD = 11.65$) for the younger group of infants; upright VbaAba ($M = 26.27$; $SD = 8.19$), inverted VbaAba ($M = 25.91$; $SD = 7.52$), upright VbaAga ($M = 26.36$; $SD = 8.26$), and inverted VbaAba ($M = 26.14$; $SD = 8.09$) for the older group of infants. The number of segments included in the analysis did not differ across conditions in both younger ($F_{(3, 57)} = 2.10$; $p = .135$) and older ($F_{(3, 63)} = .73$, $p = .541$) groups of infants.

2.5.2 | Adult data

The adult data were preprocessed offline utilizing Matlab and the EEGLAB toolbox (Delorme & Makeig, 2004) with custom-made scripts. The EEG signal was amplified, digitized at 500 Hz, and band-pass filtered from 0.1 to 200 Hz. Three, second-order Butterworth filters were implemented with 12 dB/octave roll-off to filter continuous EEG data with the high-pass cutoff of 1 Hz, the low-pass cutoff of 30 Hz, and the notch for the 49.5–50.5 Hz band. For one participant, five channels contaminated with high noise were removed (E27, E28, E45, E46, and E48). Next, missing channels were reconstructed using spherical interpolation.

The continuous EEG signal was segmented into 1100-ms-long epochs, ranging from −100 to 1000 ms, with 0 being the onset of the visual stimulus. Signals were passed through automatic artifact detection followed by visual inspection. Eye blinks were detected when the value of the difference between the maximum and the minimum amplitude, in the 0–900 ms time window, exceeded 70 μV. When searching for motion artifacts, all channels were inspected, and the threshold between maximal and minimal amplitude was set to 150 μV. Epochs contaminated by eye or motion artifacts were removed. Artifact-free segments were re-referenced to the overall average and then averaged for each participant within each condition. Averaged signals from each participant were baseline corrected using 260–360 ms intervals (i.e., the period of 100 ms immediately before the sound onset).

2.6 | ERP analysis

To obtain the average amplitudes from areas of interest—frontal and central areas on each hemisphere (see Figure S1; based on Kushnerenko et al., 2008; Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013)—the mean voltage potential was averaged from the groups of electrodes in a 100-ms-long time window. Previously, AVMMR was reported in the time window of 190–290 ms after sound onset (i.e., 550–650 ms after stimulus onset; Kushnerenko et al., 2008); thus, we conducted the main analysis within the exact same time window as in prior work. The analysis for additional time windows of infant data, namely, 90–190 ms and 290–390 ms after sound onset, was also conducted and is presented in Supporting Information S3. Adult data from a 90- to 190-ms time window were added to the results section, due to time differences between infant and adult electrophysiological responses (de Haan et al., 2002). Left and right frontal channels were centered around FC3 (channels 12, 14, 15) and FC4 (channels 53, 57, 60), respectively. The left and right central area channels included electrodes around C3 (channels 20, 21, 22) and C4 (channels 41, 49, 50; see Supporting Information for channels location map), respectively.

Statistical analysis was conducted separately for each age group. As a first step, repeated-measures analyses of variance (ANOVAs) were conducted with three within-subject factors: area (frontal left, frontal right, central left, and central right), stimulus (congruent, incongruent), and orientation (upright, inverted). As a second step, we used planned contrasts to test the presence of previously reported effects of the mismatch condition (VbaAga vs. rest). Additional pairwise comparisons (with Bonferroni correction) were used to further test differences between conditions.

3 | RESULTS

3.1 | Younger group

See descriptive statistics of mean amplitude within the 190–290 ms interval in Table 1. We first conducted a three-way ANOVA with area (4) × stimulus (2) × orientation (2) to investigate the presence of the AVMMR in frontal and central areas. The analysis revealed a significant main effect of the area ($F_{(1, 19)} = 11.13$, $p < .001$, $\eta^2 = .369$) and a three-way interaction of area × stimulus × orientation ($F_{(1, 19)} = 3.74$, $p = .016$, $\eta^2 = .164$).

Visual inspection of the grand-averaged plots (see Figure 2) suggested that the three-way interaction is driven by specific responses to the upright and inverted mismatch stimuli over two areas, the left central and right frontal areas, respectively. Thus, we compared amplitudes for each stimulus separately in each of these two areas. Over the left central area, the amplitude for the upright VbaAga was significantly higher than that for the upright VbaAba ($p = .008$; Figure 2C). Also, there was a marginal effect suggesting a difference between the upright and the inverted incongruent stimulus over the left central area, with the amplitude in response to the upright VbaAga tending to be higher than to the inverted VbaAga ($p = .069$; Figure 2C). This pattern of results was confirmed using contrast in a one-way ANOVA with all four conditions entered into the model as a single within-subjects factor. The amplitude for the upright incongruent VbaAga was significantly higher than the other conditions over the central left area ($F_{(1, 19)} = 7.42$, $p = .013$, $\eta^2 = .281$; see Figure 2C). The corresponding central channels over the right hemisphere showed no significant differential response to the upright mismatch stimulus (VbaAga, $p = .664$, $\eta^2 = .009$; see Figure 2D).

Pairwise comparisons revealed significant differences over the frontal right area, showing that the amplitude for the inverted incongruent VbaAga was higher than for the upright one ($p = .019$; see

TABLE 1 Descriptive statistics of mean amplitude (in μV) within the 190–290 ms time window after sound onset.

			Upright		Inverted	
Group	**Area**	**Condition**	**Mean (*SD*)**	**[Minimum, Maximum]**	**Mean (*SD*)**	**[Minimum, Maximum]**
Younger (5- to 6.5-month-olds)	Left frontal	Congruent VbaAba	1.51 (3.74)	[−4.48, 11.16]	1.689 (2.94)	[−3.33, 6.47]
		Incongruent VbaAga	2.49 (3.71)	[−3.55, 8.99]	1.00 (3.37)	[−5.95, 9.41]
	Right frontal	Congruent VbaAba	0.61 (4.18)	[−13.66, 5.58]	0.38 (4.36)	[−8.06, 9.77]
		Incongruent VbaAga	0.12 (3.81)	[−8.02, 6.20]	2.34 (3.72)	[−3.44, 9.36]
	Left central	Congruent VbaAba	−0.28 (3.39)	[−7.08, 7.82]	0.24 (3.36)	[−5.92, 5.52]
		Incongruent VbaAga	1.91 (3.75)	[−3.58, 11.50]	0.01 (3.71)	[−6.22, 7.74]
	Right central	Congruent VbaAba	−2.10 (2.94)	[−8.37, 3.57]	−2.53 (4.99)	[−12.65, 7.87]
		Incongruent VbaAga	−1.56 (4.40)	[−10.78, 6.68]	−2.22 (3.34)	[−6.67 4.47]
Older (9- to 10.5-month-olds)	Left frontal	Congruent VbaAba	1.02 (2.83)	[−5.01, 5.08]	2.18 (3.48)	[−4.00, 10.99]
		Incongruent VbaAga	2.70 (4.03)	[−2.34, 11.31]	1.69 (2.84)	[−4.37, 7.39]
	Right frontal	Congruent VbaAba	0.9686 (3.49)	[−4.17, 8.55]	0.9673 (3.71)	[−4.89, 11.07]
		Incongruent VbaAga	0.80 (3.96)	[−5.09, 8.57]	1.40 (2.89)	[−2.68, 8.40]
	Left central	Congruent VbaAba	−0.29 (2.51)	[−4.20, 4.70]	−1.24 (3.21)	[−6.91, 5.76]
		Incongruent VbaAga	0.90 (3.56)	[−5.09, 9.44]	0.32 (3.78)	[−7.68, 10.20]
	Right central	Congruent VbaAba	0.01 (2.77)	[−5.09, 4.08]	−1.25 (2.56)	[−6.51, 4.96]
		Incongruent VbaAga	−0.92 (2.73)	[−7.59, 5.44]	−0.64 (3.19)	[−6.59, 6.83]

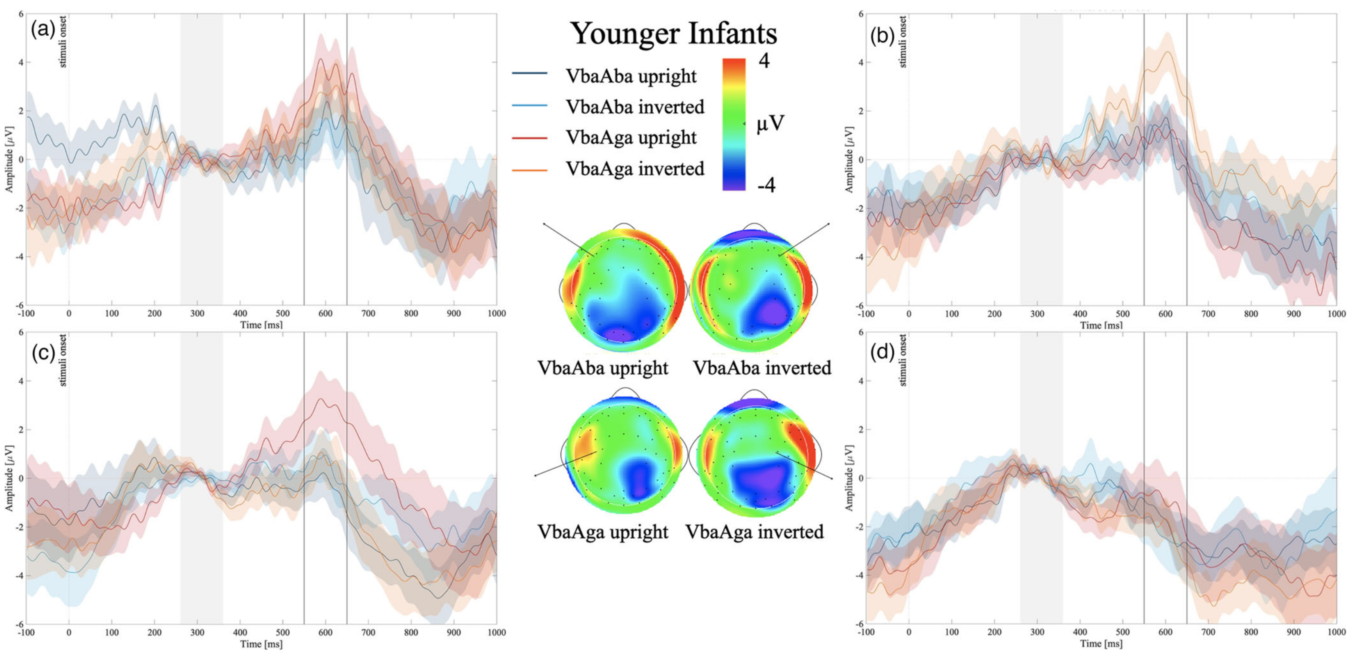

FIGURE 2 Grand-averaged event-related potential (ERP) responses for the younger group (panels: [A] channel 14; [B] channel 57; [C] channel 20; [D] channel 50) to the audiovisual (AV) stimuli: upright VbaAba (dark blue), inverted VbaAba (light blue), upright VbaAga (red), and inverted VbaAga (orange). The gray-shaded area represents the baseline correction period. The analysis time window of 190–290 ms after the sound onset is marked by solid black lines. Topographic maps (top projection) represent grand-averaged responses to each condition within the 190–240 ms time window (from the sound onset).

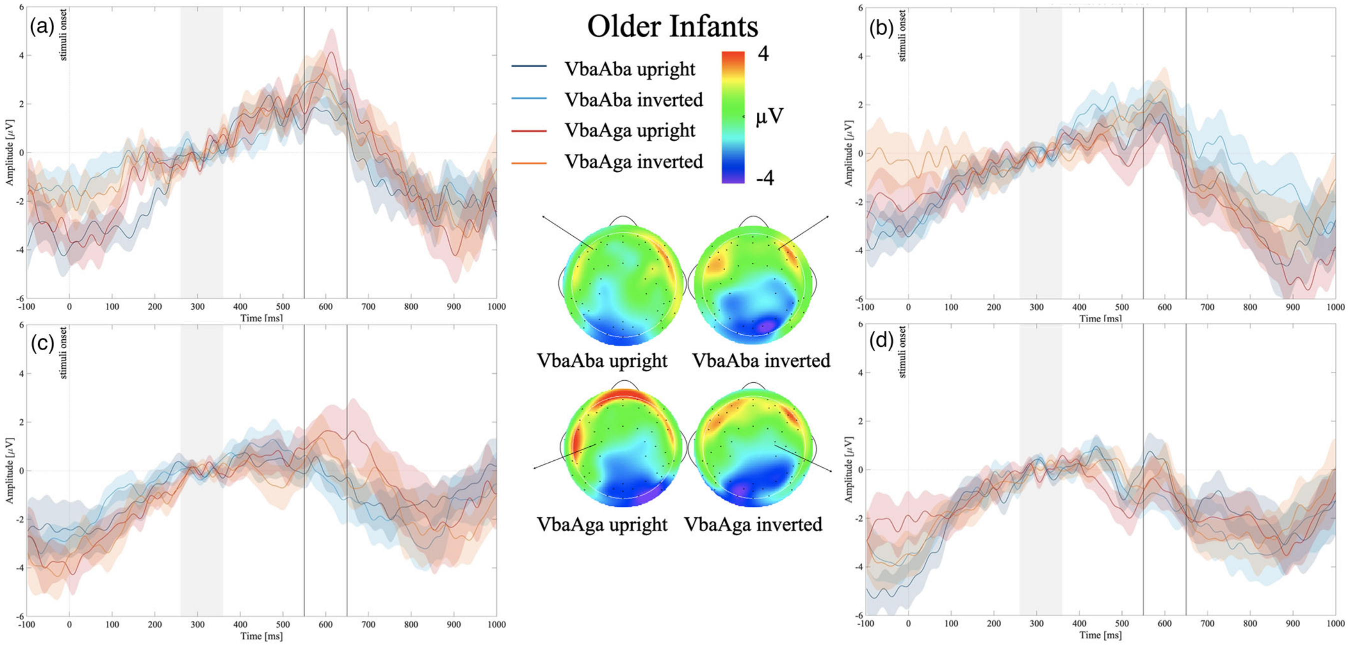

FIGURE 3 Grand-averaged event-related potential (ERP) responses for the older group (panels: [A] channel 14; [B] channel 57; [C] channel 20; [D] channel 50) to the audiovisual (AV) stimuli: upright VbaAba (dark blue), inverted VbaAba (light blue), upright VbaAga (red), and inverted VbaAga (orange). The gray-shaded area represents the baseline correction. The analysis time window of 190- 290 ms after the sound onset is marked by solid black lines. Topographic maps (top projection) represent grand-averaged responses to each condition within the 190–240 ms time window (from the sound onset).

Figure 2B). Again, we confirmed this result using contrast in a one-way ANOVA. Over the right frontal area, the amplitude for the inverted VbaAga was significantly higher than the other conditions ($F_{(1, 19)} = 8.35$, $p = .009$, $\eta^2 = .305$). The contrast for the corresponding channels over the right frontal area did not show a significantly higher amplitude of responses to inverted VbaAga condition relative to other conditions ($p = .501$, $\eta^2 = .040$; Figure 2A).

3.2 | Older group

To test whether the AVMMR is elicited by incongruent stimulus in the upright and inverted orientation, a 4 (area) × 2 (stimulus) × 2 (orientation) ANOVA was conducted for the older group of infants, which showed only a significant main effect of area ($F_{(3,63)} = 14.10$, $p < .001$, $\eta^2 = .40$). No other effect was significant.

Analogously to the younger group, planned contrasts over the left central area were conducted to test the presence of the AVMMR. The amplitude of response to the upright VbaAga stimulus did not significantly differ from the remaining three conditions over the left ($p = .137$, $\eta^2 = .102$; Figure 3C) and right central channels ($p = .664$, $\eta^2 = .009$; Figure 3D), which may imply a lack of AVMMR for the upright VbaAga. Also, the contrast for the right frontal area did not show a significantly higher amplitude of responses to inverted VbaAga condition relative to other conditions over the left ($p = .730$, $\eta^2 = .006$; Figure 3A) and right hemisphere ($p = .581$, $\eta^2 = .015$; Figure 3B).

3.3 | Adults

In adults, we observed peaks of ERP response to AV stimuli earlier than in infants, which is consistent with the literature (de Haan et al., 2003). Thus, due to discrepancies in ERP timing between infants and adults, we analyzed an earlier window of 90–190 ms (results for the subsequent time window of 190–290 ms are presented in the Supporting Information). In the analysis of adult data, we focused on testing the presence of responses to upright and inverted VbaAga stimuli over two areas that showed differential responses in infants: the left central and the right frontal channels groups.

In the 90–190 ms time window, three-way ANOVA, 4 (area) × 2 (stimulus) × 2 (orientation), showed a significant interaction of area × stimulus ($F_{(3, 57)} = 4.17$, $p = .028$, $\eta^2 = .180$) and a significant three-way interaction of area × stimulus × orientation ($F_{(3, 57)} = 4.32$, $p = .023$, $\eta^2 = .185$). The interaction of area × orientation was approaching significance ($F_{(3, 57)} = 2.64$, $p = .058$, $\eta^2 = .122$). There were no significant main effects.

Data for the younger infant group showed two specific responses to the upright and inverted mismatch stimuli, respectively, over the left central and the right frontal areas. We used planned contrasts to test the presence of these responses in the adult data. The amplitude of response to the upright VbaAga stimulus did not significantly differ from the remaining three conditions over the left central area ($p = .640$, $\eta^2 = .012$; see Figure 4C) and over the right central area ($p = .109$, $\eta^2 = .130$; see Figure 4D). The planned contrast for the

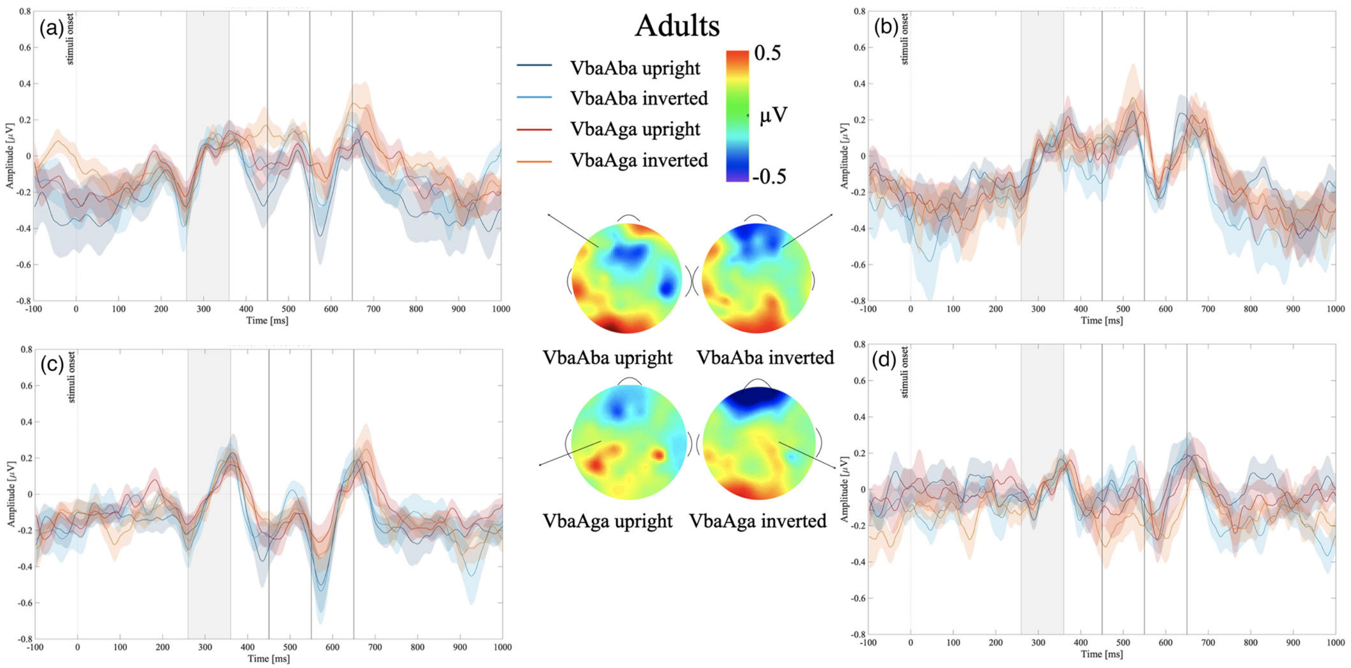

FIGURE 4 Grand-averaged event-related potential (ERP) responses of adults (panels: [A] channel 14; [B] channel 57; [C] channel 20; [D] channel 50) to the audiovisual (AV) stimuli: upright VbaAba (dark blue), inverted VbaAba (light blue), upright VbaAga (red), and inverted VbaAga (orange). The gray-shaded area represents the baseline correction. The time windows of 90–190 ms and 190–290 ms after the sound onset are marked by solid black lines. Topographic maps (top projection) represent grand-averaged responses to each condition within the 190–240 ms time window (from the sound onset).

TABLE 2 Descriptive statistics of mean amplitude (in μV) within the 90–190 ms time window after sound onset.

			Upright		Inverted	
Group	Area	Condition	Mean (*SD*)	[Minimum, Maximum]	Mean (*SD*)	[Minimum, Maximum]
Adults	Left frontal	Congruent VbaAba	−0.52 (0.14)	[−0.44, 0.25]	0.05 (0.20)	[−0.32, 0.52]
		Incongruent VbaAga	−0.04 (0.27)	[−0.77, 0.51]	0.22 (0.37)	[−0.12, 1.40]
	Right frontal	Congruent VbaAba	0.03 (0.44)	[−1.−1, 1.46]	−0.08 (0.29)	[−0.70, 0.37]
		Incongruent VbaAga	0.06 (0.28)	[−0.35, 0.90]	0.23 (0.37)	[−0.10, 1.29]
	Left central	Congruent VbaAba	−0.04 (0.42)	[−0.98, 1.32]	0.07 (0.31)	[−0.45, 1.13]
		Incongruent VbaAga	−0.1 (0.39)	[−1.48, 0.50]	−0.20 (0.35)	[−1.09, 0.26]
	Right central	Congruent VbaAba	0.01 (0.19)	[−0.56, 0.41]	−0.16(0.53)	[−2.31, 0.48]
		Incongruent VbaAga	−0.11 (0.22)	[−0.87, 0.11]	0.01 (0.17)	[−0.19, 0.56]

frontal areas showed that adults had a significantly higher amplitude to inverted VbaAga condition relative to other conditions over the left ($F_{(1, 19)} = 5.08$, $p = .036$, $\eta^2 = .211$; Figure 4A) and right hemisphere ($F_{(1, 19)} = 7.21$, $p = .015$, $\eta^2 = .275$; Figure 4B). See Table 2 for descriptive statistics of mean amplitude for both time windows.

4 | DISCUSSION

The aim of the study was to investigate the effect of face orientation on the neural processing of AV speech in infants at two different stages of the development of face specialization. To this end, we presented two groups of infants (younger, aged 5–6.5 months; older, aged 9–10.5 months) and a group of adults with AV speech (actresses articulating syllables) consisting of congruent /ba/ syllable (VbaAba) and incongruent VbaAga (mismatching McGurk) stimuli in the upright and the inverted face orientation. In the younger group of infants, we found the AVMMR to the upright incongruent stimulus in the 190–290 ms time window. A notable novel finding is that the younger group of infants also showed the AVMMR to the inverted incongruent VbaAga stimulus relative to the upright one. By contrast, in the older group of infants, we found no AVMMR to the incongruent stimulus in either upright or inverted orientation. Finally, the AVMMR only for the inverted incongruent stimulus was also found over frontal areas in adults, although in an earlier time window, 90–190 ms. Altogether, our study shows different patterns of neural responses to AV mismatch in two infant age groups, whose age ranges were selected with respect to the timing of emerging specialization for configural face processing. We interpret these findings as suggesting that specialization for face processing may influence the development of AVSI.

The AVMMR was previously interpreted as evidence for the detection of conflicting cross-modal speech stimuli and was observed in 5-month-olds (Kushnerenko et al., 2008). Further research on AVSI provided evidence that the gradual decrease of the AVMMR can be related to more mature patterns of face scanning (Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013) and to later development of language (Kushnerenko, Tomalski, Ballieux, Potton, et al., 2013). In our current study, we replicated previous findings regarding the AVMMR in response to upright incongruent VbaAga stimulus in younger infants aged 5–6.5 months. Since the AVMMR in response to upright VbaAga stimulus is absent in adults, which was demonstrated both in our current study and in previous work (see Control Study S3 in Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013), we consider the absence of the AVMMR in the older group of infants as an index of more mature AVSI. We also propose that the diminishing AVMMR in response to the upright conflicting AV speech information is a result of AV speech integration overlapping with emerging configural face processing.

Existing behavioral data are in line with different patterns of responses to AV mismatch in the two age groups in our results. Longer looking to the mouth area of articulating face was observed in infants during the perception of McGurk stimuli in comparison to the congruent native speech stimuli (Mercure et al., 2019). Such increased visual attention to the mouth has been considered in the literature as an index of mature scanning patterns, related to speech perception development (Lewkowicz & Hansen-Tift, 2012; Lozano et al., 2022; Morin-Lessard et al., 2019). Mercure and collaborators (2019) concluded that longer fixations to the articulating mouth in younger infants (6.5- and 8-month-olds) in the incongruent condition may be due to novelty or surprise effect. That possible explanation implies that infants at that developmental stage may have an AV representation of syllables so they detect a novel, unpredicted sound that is incongruent with the mouth movement. However, a study conducted by the same research team on an older group of monolinguals (7- to 10-month-olds) showed the reverse effect: more looking to the eyes in the incongruent condition (Mercure et al., 2022). Interpretation of the latter result was that infants shifted their attention away from mismatching mouth movements to tolerate articulatory inconsistencies. Thus, this argumentation strengthens the possibility that our older group of infants ignored the conflicting cues to reduce uncertainty. As a result, older infants in our study would be able to integrate the conflicting AV cues, thus showing no AVMMR to either upright or inverted mismatch.

Another possible explanation for the lack of significant differences in response to incongruent stimuli in both upright and inverted conditions in the older group of infants is cortical reorganization for speech processing. Recent fNIRS study from our lab with two groups of infants of similar age ranges showed that younger infants showed specific cortical responses to stimuli, which required integration of auditory and visual speech cues relative to asynchronous auditory + visual stimuli, but older infants did not show such specific responses (Dopierała et al., 2023). MVPA did not show condition-specific neural responses to bimodal, AV speech in the older group of infants, while it did find such responses in younger infants, aged 5–6 months. This lack of a successful classification of cortical responses to speech stimuli found in the older group suggests a likely functional cortical reorganization in the perception of AV speech around the age of 9–10 months (Dopierała et al., 2023). These fNIRS findings closely resemble our current results, where we did not find significant differences in responses to AV mismatch stimuli in either orientation in infants aged 9–10.5 months. Altogether, we interpreted the current findings as further evidence for a lack of stable, consistent responses to AV speech in infants, suggestive of functional reorganization.

The scalp topography of the AVMMR to the upright mismatch stimulus observed in our study is consistent with previous ERP studies, reported in the same time window of analysis (Kushnerenko et al., 2008; Kushnerenko, Tomalski, Ballieux, Ribeiro, et al., 2013). Inversion of the incongruent VbaAga, however, resulted in a mismatch response in a different location—over right frontal channels—which is a novel finding. Differences in scalp location of responses to the upright versus inverted mismatch stimulus may suggest that AV speech processing mechanisms depend on face configuration from early on. We propose that frontal AVMMR is a response elicited due to the novelty effect of inverted faces, which occurred only for incongruent stimuli in the younger group of infants, with less advanced specialization of face processing in comparison with the older group. The involvement of frontal regions in relation to novel stimuli has been reported in 3-month-olds both in fNIRS (Nakano et al., 2009) and EEG (Bristow et al., 2009), where the right frontal activation in response to the incongruent cross-modal AV speech stimuli has been demonstrated. Both studies showed that frontal areas might respond to novelty processing early in infancy, so it is likely that in our study younger infants also activated this area while processing previously unseen incongruent VbaAga.

In our current study, adults also showed differential responses to inverted incongruent stimuli over frontal areas. We propose that this response is an effect of conflict detection in an inverted face, possible when the face is processed featurally. Existing studies suggest that when looking at inverted faces, featural processing precedes configural processing (Carbon & Leder, 2005). As a result of the AV conflict in likely featurally processed inverted faces, the AVMMR was observed on frontal channels in the 90–190 ms time window after sound onset. The differences in time windows are common because the latencies and morphology of infantile ERPs vary from the ones observed in adults (de Haan et al., 2002; Wunderlich et al., 2006). The visible frontal positivity in response to inverted AV mismatch can be discussed in relevance to the anterior cingulate cortex activity during conflict detection resulting in error positivity (e.g., Orr & Carrasco, 2011; van Veen & Carter, 2002); however, the large differences in procedures should be acknowledged between conflicting speech cues in this study and in the mentioned attention tasks. Taken together, the frontal AVMMR is a novel effect found in response to inverted conflicting AV speech in 5- to 6.5-month-olds and adults (in different time windows, but with similar scalp topography). The AVMMR in response to inverted VbaAga was not detected in infants aged 9–10.5 months, again possibly indicating a transition period in this age group.

Our results are not fully in line with Riva and collaborators (2022), who in a later time window (350–650 ms) reported responses to upright AV speech mismatch over the left temporal area in 12-month-olds—several months past the age at which we no longer observed mismatch response to the upright VbaAga in an earlier time window. Direct comparisons between the two studies are not possible due to important differences in the experimental paradigm: shorter duration of our stimuli, the native language of participants, the difference in speech sounds used in each study, and the use of inverted stimuli. Thus, more in-depth research is needed to fully understand the nuances of infant scalp topography as well as the time course of different neural responses in relation to both AV speech integration and face processing.

Our study is subject to several limitations. The number of infants included in the study is limited and does not allow a robust direct comparison between age groups. Another limitation is the number of segments included in infant analyses. It is comparable to the numbers achieved in similar studies (Kushnerenko, Tomalski, Ballieux, Ribeiro et al., 2013; Riva et al., 2022); however, with such a low number of repetitions compared to adults, there is an increased risk related to a lower signal-to-noise ratio. Potential solutions for future research would be in limiting the number of conditions, which might help to increase the number of repetitions, as well as the number of infants contributing sufficient data (Hoehl & Wahl, 2012).

The suggestion for future research is to investigate the fusible McGurk (visual /ga/ dubbed onto auditory /ga/) stimulus across the ages to understand how potential featural processing in inverted faces occurs for fusible AV speech. Rosenblum et al. (2000) suggested that facial configuration might be more important for some phonemes than others. The combination of visual /ba/ and auditory /ga/ results in the illegal Polish language percept of /bga/ and bilabial /ba/ viseme cannot result as /ga/ phoneme. The combination of syllables using velar /ga/ viseme may be more difficult to detect by the perceiver since the lip movement of /ga/ might be mistaken with other syllables. Future research should also explore older infants within similar paradigms to assess the age when the infant neural responses resemble more adultlike patterns. Also, the coregistration with an eye tracker may potentially answer how the scanning patterns of the speaking face are connected with the amplitude of AVMMR.

5 | CONCLUSION

Our electrophysiological study demonstrates that AVSI and face processing are related already in infancy, sharing some neural mechanisms resulting in differential responses dependent on face orientation. Our results shed new light on the development of speech processing by showing that neural responses to AV conflict diminish with age, a process likely related to emerging configural face processing. We also showed that face inversion of incongruent speech results in neural responses in the frontal areas in younger infants and adults. These differential neural responses suggest that AV integration depends on face orientation already in mid-infancy. The lack of consistent responses to AV speech in the older group of infants reflects the suggested potential functional cortical reorganization in the processing of AV speech.

ACKNOWLEDGMENTS

We would like to thank David Lopez Pérez for assistance with programming and Agata Kozioł for help in data visualization. We also thank Bogumił Karbowski for assistance with stimulus preparation and recruitment and Elena Kushnerenko for helpful discussions of the research paradigm at the early stages of this work. We acknowledge the generous contribution of all infants and caregivers who participated in this study. We also want to thank Karolina Babis and Aleksandra Dopierała for their assistance and support during data collection. Finally, we are grateful to Danijel Korzinek and the late Krzysztof Marasek from the Polish-Japanese Academy of Information Technology in Warsaw for access to the recording studio and assistance with stimulus recording.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in the OSF repository at https://osf.io/4yv6t/?view_only=96203b53a7ba4e3698462d033b6534ca.

ORCID

Magdalena Szmytke https://orcid.org/0000-0003-4714-0601
Zuzanna Laudańska https://orcid.org/0000-0001-6790-9559
Przemysław Tomalski https://orcid.org/0000-0002-0390-5759

REFERENCES

Belin, P., Bestelmeyer, P. E. G., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*(4), 711–725. https://doi.org/10.1111/j.2044-8295.2011.02041.x

Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., & Mangin, J. F. (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. https://doi.org/10.1162/jocn.2009.21076

Burnham, D., & Dodd, B. (2004). Auditory-visual speech integration by prelinguistic infants: Perception of an emergent consonant in the McGurk effect. *Developmental Psychobiology*, *45*(4), 204–220. https://doi.org/10.1002/dev.20032

Carbon, C.-C., & Leder, H. (2005). When feature information comes first! Early processing of inverted faces. *Perception*, *34*(9), 1117–1134. https://doi.org/10.1068/p5192

Cohen, L. B., & Cashon, C. H. (2001). Do 7-month-old infants process independent features or facial configurations? *Infant and Child Development*, *10*(1–2), 83–92. https://doi.org/10.1002/icd.250

de Haan, M., Johnson, M. H., & Halit, H. (2003). Development of face-sensitive event-related potentials during infancy: a review. *International Journal of Psychophysiology*, *51*(1), 45–58. https://doi.org/10.1016/s0167-8760(03)00152-1

De Haan, M., Pascalis, O., & Johnson, M. H. (2002). Specialization of neural mechanisms underlying face recognition in human infants. *Journal of Cognitive Neuroscience*, *14*(2), 199–209. https://doi.org/10.1162/089892902317236849

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Desjardins, R. N., Rogers, J., & Werker, J. F. (1997). An exploration of why preschoolers perform differently than do adults in audiovisual speech perception tasks. *Journal of Experimental Child Psychology*, *66*(1), 85–110. https://doi.org/10.1006/jecp.1997.2379

Desjardins, R. N., & Werker, J. F. (2004). Is the integration of heard and seen speech mandatory for infants? *Developmental Psychobiology*, *45*(4), 187–203. https://doi.org/10.1002/dev.20033

Dopierała, A. A., Pérez, D. L., Mercure, E., Pluta, A., Malinowska-Korczak, A., Evans, S., Wolak, T., & Tomalski, P. (2023). The development of cortical responses to the integration of audiovisual speech in infancy. *Brain Topography*, *36*(4), 459–475.

Eskelund, K., MacDonald, E. N., & Andersen, T. S. (2015). Face configuration affects speech perception: Evidence from a McGurk mismatch negativity study. *Neuropsychologia*, *66*, 48–54. https://doi.org/10.1016/j.neuropsychologia.2014.10.021

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns' preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(47), 17245–17250. https://doi.org/10.1073/pnas.0502205102

Grant, K. W., & Seitz, P.-F. (2000). The use of visible speech cues for improving auditory detection of spoken sentences. *The Journal of the Acoustical Society of America*, *108*(3 Pt 1), 1197–1208. https://doi.org/10.1121/1.1288668

Guellaï, B., Streri, A., & Yeung, H. H. (2014). The development of sensorimotor influences in the audiovisual speech domain: Some critical questions. *Frontiers in Psychology*, *5*, Article 812. https://doi.org/10.3389/FPSYG.2014.00812

Halit, H., de Haan, M., & Johnson, M. H. H. (2003). Cortical specialisation for face processing: Face-sensitive event-related potential components in 3- and 12-month-old infants. *Neuroimage*, *19*(3), 1180–1193. https://doi.org/10.1016/S1053-8119(03)00076-4

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*(3), 360–378. https://doi.org/10.1016/j.specom.2005.04.007

Hietanen, J. K., Manninen, P., Sams, M., & Surakka, V. (2001). Does audiovisual speech perception use information about facial configuration? *European Journal of Cognitive Psychology*, *13*(3), 395–407. https://doi.org/10.1080/09541440126006

Hoehl, S., & Wahl, S. (2012). Recording infant ERP data for cognitive research. *Developmental Neuropsychology*, *37*(3), 187–209. https://doi.org/10.1080/87565641.2011.627958

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1–2), 1–19. https://doi.org/10.1016/0010-0277(91)90045-6

Johnson, M. H., Senju, A., & Tomalski, P. (2015). The two-process theory of face processing: Modifications based on two decades of data from infants and adults. *Neuroscience & Biobehavioral Reviews*, *50*, 169–179. https://doi.org/10.1016/j.neubiorev.2014.10.009

Jordan, T. R., & Bevan, K. (1997). Seeing and hearing rotated faces: Influences of facial orientation on visual and audiovisual speech recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 388–403. https://doi.org/10.1037/0096-1523.23.2.388

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Ge, L., & Pascalis, O. (2007). The other-race effect develops during infancy: Evidence of perceptual narrowing. *Psychological Science*, *18*(12), 1084–1089. https://doi.org/10.1111/j.1467-9280.2007.02029.x

Krasotkina, A., Götz, A., Höhle, B., & Schwarzer, G. (2018). Perceptual narrowing in speech and face recognition: Evidence for intra-individual cross-domain relations. *Frontiers in Psychology*, *9*, Article 1711. https://doi.org/10.3389/fpsyg.2018.01711

Krasotkina, A., Götz, A., Höhle, B., & Schwarzer, G. (2021). Perceptual narrowing in face- and speech-perception domains in infancy: A longitudinal approach. *Infant Behavior and Development*, *64*, Article 101607. https://doi.org/10.1016/j.infbeh.2021.101607

Kushnerenko, E., Teinonen, T., Volein, A., & Csibra, G. (2008). Electrophysiological evidence of illusory audiovisual speech percept in human infants. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11442–11445. https://doi.org/10.1073/pnas.0804275105

Kushnerenko, E., Tomalski, P., Ballieux, H., Potton, A., Birtles, D., Frostick, C., & Moore, D. G. (2013). Brain responses and looking behavior during audiovisual speech integration in infants predict auditory speech comprehension in the second year of life. *Frontiers in Psychology*, *4*, Article 432. https://doi.org/10.3389/fpsyg.2013.00432

Kushnerenko, E., Tomalski, P., Ballieux, H., Ribeiro, H., Potton, A., Axelsson, E. L., Murphy, E., & Moore, D. G. (2013). Brain responses to audiovisual speech mismatch in infants are associated with individual differences in looking behaviour. *European Journal of Neuroscience*, *38*(9), 3363–3369. https://doi.org/10.1111/ejn.12317

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., & Ghazanfar, A. A. (2009). The emergence of multisensory systems through perceptual narrowing. *Trends in Cognitive Sciences*, *13*(11), 470–478. https://doi.org/10.1016/j.tics.2009.08.004

Lewkowicz, D. J., & Hansen-Tift, A. M. (2012). Infants deploy selective attention to the mouth of a talking face when learning speech. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(5), 1431–1436. https://doi.org/10.1073/pnas.111478310

Lozano, I., López Pérez, D., Laudańska, Z., Malinowska-Korczak, A., Szmytke, M., Radkowska, A., & Tomalski, P. (2022). Changes in selective attention to articulating mouth across infancy: Sex differences and associations with language outcomes. *Infancy*, *27*(6), 1132–1153. https://doi.org/10.1111/infa.12496

Massaro, D. W., & Cohen, M. M. (1996). Perceiving speech from inverted faces. *Perception & Psychophysics*, *58*(7), 1047–1065.

Maurer, D., & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178. https://doi.org/10.1002/dev.21177

McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0

Mercure, E., Bright, P., Quiroz, I., & Filippi, R. (2022). Effect of infant bilingualism on audiovisual integration in a McGurk task. *Journal of Experimental Child Psychology*, *217*, Article 105351. https://doi.org/10.1016/j.jecp.2021.105351

Mercure, E., Kushnerenko, E., Goldberg, L., Bowden-Howl, H., Coulson, K., Johnson, M. H., & MacSweeney, M. (2019). Language experience influences audiovisual speech integration in unimodal and bimodal bilingual infants. *Developmental Science*, *22*(1), Article e12701. https://doi.org/10.1111/desc.12701

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

Nakano, T., Watanabe, H., Homae, F., & Taga, G. (2009). Prefrontal cortical involvement in young infants' analysis of novelty. *Cerebral Cortex*, *19*(2), 455–463. https://doi.org/10.1093/cercor/bhn096

Orr, J. M., & Carrasco, M. (2011). The role of the error positivity in the conscious perception of errors. *The Journal of Neuroscience*, *31*(16), 5891–5892. https://doi.org/10.1523/JNEUROSCI.0279-11.2011

Pascalis, O., Loevenbruck, H., Quinn, P. C., Kandel, S., Tanaka, J. W., & Lee, K. (2014). On the links among face processing, language processing, and narrowing during development. *Child Development Perspectives*, *8*(2), 65–70. https://doi.org/10.1111/cdep.12064

Peykarjou, S., & Hoehl, S. (2013). Three-month-olds' brain responses to upright and inverted faces and cars. *Developmental Neuropsychology*, *38*(4), 272–280. https://doi.org/10.1080/87565641.2013.786719

Riva, V., Riboldi, E. M., Dondena, C., Piazza, C., Molteni, M., & Cantiani, C. (2022). Atypical ERP responses to audiovisual speech integration and sensory responsiveness in infants at risk for autism spectrum disorder. *Infancy*, *27*(2), 369–388. https://doi.org/10.1111/infa.12456

Rosenblum, L. D., Yakel, D. A., & Green, K. P. (2000). Face and mouth inversion effects on visual and audiovisual speech perception. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 806–819. https://doi.org/10.1037/0096-1523.26.2.806

Sams, M., Aulanko, R., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. (1991). Seeing speech: Visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, *127*(1), 141–145. https://doi.org/10.1016/0304-3940(91)90914-F

Schwarzer, G., Zauner, N., & Jovanovic, B. (2007). Evidence of a shift from featural to configural face processing in infancy. *Developmental Science*, *10*(4), 452–463. https://doi.org/10.1111/j.1467-7687.2007.00599.x

Teinonen, T., Aslin, R. N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition*, *108*(3), 850–855. https://doi.org/10.1016/j.cognition.2008.05.009

Tomalski, P. (2015). Developmental trajectory of audiovisual speech integration in early infancy. A review of studies using the McGurk paradigm. *Psychology of Language and Communication*, *19*(2), 77–100. https://doi.org/10.1515/plc-2015-0006

Ujiie, Y., Kanazawa, S., & Yamaguchi, M. K. (2020). The other-race-effect on audiovisual speech integration in infants: A NIRS study. *Frontiers in Psychology*, *11*, Article 971. https://doi.org/10.3389/fpsyg.2020.00971

van Veen, V., & Carter, C. S. (2002). The anterior cingulate as a conflict monitor: fMRI and ERP studies. *Physiology & Behavior*, *77*(4-5), 477–482. https://doi.org/10.1016/s0031-9384(02)00930-7

Weikum, W. M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., & Werker, J. F. (2007). Visual language discrimination in infancy. *Science*, *316*(5828), 1159. https://doi.org/10.1126/science.1137686

Werker, J. F., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*(1), 49–63. https://doi.org/10.1016/S0163-6383(84)80022-3

Wunderlich, J. L., Cone-Wesson, B. K., & Shepherd, R. (2006). Maturation of the cortical auditory evoked potential in infants and young children. *Hearing Research*, *212*(1–2), 185–202. https://doi.org/10.1016/j.heares.2005.11.010

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: Szmytke, M., Ilyka, D., Duda-Goławska, J., Laudańska, Z., Malinowska-Korczak, A., & Tomalski, P. (2023). The effect of face orientation on audiovisual speech integration in infancy: An electrophysiological study. *Developmental Psychobiology*, *65*, e22431. https://doi.org/10.1002/dev.22431