



# The role of left inferior frontal cortex during audiovisual speech perception in infants



Nicole Altvater-Mackensen<sup>a,\*</sup>, Tobias Grossmann<sup>a,b</sup>

<sup>a</sup> Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

<sup>b</sup> University of Virginia, Charlottesville, USA

## ARTICLE INFO

### Article history:

Received 12 January 2016

Accepted 21 February 2016

Available online 3 March 2016

### Keywords:

Audiovisual speech perception

Visual speech cues

Eye tracking

fNIRS

Language development

## ABSTRACT

In the first year of life, infants' speech perception attunes to their native language. While the behavioral changes associated with native language attunement are fairly well mapped, the underlying mechanisms and neural processes are still only poorly understood. Using fNIRS and eye tracking, the current study investigated 6-month-old infants' processing of audiovisual speech that contained matching or mismatching auditory and visual speech cues. Our results revealed that infants' speech-sensitive brain responses in inferior frontal brain regions were lateralized to the left hemisphere. Critically, our results further revealed that speech-sensitive left inferior frontal regions showed enhanced responses to matching when compared to mismatching audiovisual speech, and that infants with a preference to look at the speaker's mouth showed an enhanced left inferior frontal response to speech compared to infants with a preference to look at the speaker's eyes. These results suggest that left inferior frontal regions play a crucial role in associating information from different modalities during native language attunement, fostering the formation of multimodal phonological categories.

© 2016 Elsevier Inc. All rights reserved.

## Introduction

In the course of the first year of life, infants attune their speech perception to the language(s) spoken around them by refining their sensitivity to native sound contrasts while losing sensitivity to sound contrasts that are not relevant in their native language(s). This holds for auditory as well as visual and audiovisual speech perception (for a review, see Maurer and Werker, 2014). Although the behavioral changes in speech perception during the first year of life have been studied extensively, the neural mechanism and processes underlying perceptual attunement are still only poorly understood. Using fNIRS and eye tracking, the current study adds to the understanding of the underpinnings of infant speech perception by investigating the processing of multimodal speech information in infants undergoing native language attunement.

Most studies investigating the neural foundations of speech perception in infants focused on auditory speech perception and activation patterns elicited in brain regions involved in auditory processing. In these studies, infants show activation in response to speech in temporal (auditory) cortices soon after birth and left-lateralized responses within temporal regions have been shown to emerge around 2 months of age (for a review, see Minagawa-Kawai et al., 2011). More recently, research

has also begun to explore the recruitment of anterior frontal and inferior frontal brain regions during speech perception and their potential role for phonological learning in the first year of life. In line with evidence that frontal cortex is involved in phonological learning and processing in adults (e.g., Callan et al., 2004; Ojanen et al., 2005; Skipper et al., 2007; Wilson et al., 2004), these studies provide increasing evidence for the activation of frontal brain regions during speech perception in early infancy.

For instance, in a study by Dehaene-Lambertz et al. (2006), 3-month-old infants show activation in frontal brain regions in response to native speech sentences (see also Dehaene-Lambertz et al., 2002; Peña et al., 2003). In particular, the neural response to repeated sentences is enhanced in left inferior frontal regions, suggesting their involvement in verbal memory processes. Their findings further suggest that specifically Broca's area in left inferior frontal cortex serves as a node involved in integrating information from different domains, which they argue might help to establish a speech perception–production link during development. In line with this view, Imada et al. (2006) report increasing activation of left inferior frontal regions during speech perception over the course of the first year of life, indicating that frontal regions become increasingly activated with development and experience. Because activation in frontal and temporal regions showed similar timing for speech but not for non-speech stimuli at 6 and 12 months of age, they further suggest that increasing frontal activation reflects the recruitment of motor areas during speech perception, binding perception and production in the course of early development. In contrast to Imada et al. (2006), Perani et al. (2011) find frontal activation in

\* Corresponding author at: Max Planck Research Group Early Social Development, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr, 1a, Leipzig 04103, Germany.

E-mail address: [altvater@cbs.mpg.de](mailto:altvater@cbs.mpg.de) (N. Altvater-Mackensen).

response to speech already in newborns. They show that newborns activate a neural network that is similar to that of adults in response to speech, including regions in temporal and inferior frontal cortex. Their results further indicate that inter- and intra-hemispheric connections between temporal and frontal pre-motor areas are present from birth, while intra-hemispheric connections between temporal cortex and Broca's area are yet to develop. They therefore suggest that a dorsal pathway linking auditory and motor areas is present from birth and is recruited for phonological learning to establish a perceptuomotor link that couples auditory and articulatory speech patterns when the infant starts to babble (see also [Homae et al., 2011](#)). Recent results from [Kuhl et al. \(2014\)](#) further support the notion that frontal brain regions are involved in phonological learning. Testing 7- and 11-month-old infants, they found activation in frontal and temporal brain regions in response to both native and non-native speech sounds in 7-month-olds, while at 11 months of age, activation patterns are modulated by language familiarity with native sounds eliciting stronger activation than non-native sounds in auditory areas and weaker activation in motor areas. [Kuhl et al. \(2014\)](#) interpret their results as evidence that activation in temporal (auditory) brain regions reflects the perceptual attunement to native speech, while activation in inferior frontal (motor) regions reflects the recruitment of articulatory information during speech perception (which is more challenging for non-native speech because there are no motor representations readily available, reflecting the formation of a language-specific articulatory-auditory map).

Taken together, infants recruit inferior frontal brain regions from early on during the development of speech perception, and with development, their involvement is affected by experience. The role of left inferior frontal brain regions during speech perception is primarily seen as reflecting the coupling of auditory and articulatory speech information during development. This interpretation fits well with findings that frontal regions are involved in the phonological processing of audiovisual speech in adults. In particular, [Ojanen et al. \(2005\)](#) report that Broca's area and left inferior prefrontal cortex are activated more strongly during processing of mismatching than processing of matching audiovisual speech sounds. They suggest that stronger activation in Broca's area reflects higher demands of processing during the mismatch condition, as auditory and visual phonetic features cannot be integrated but have to be processed separately, while prefrontal activation when processing mismatching information is taken to reflect increased attention to conflicting cues. [Ojanen et al. \(2005\)](#) therefore interpret their results as evidence that Broca's area mediates the mapping of acoustic and visual input onto (motor) representations of sounds during audiovisual speech perception (see also [Skipper et al., 2007](#)). Similar reports for phonological categorization tasks in adult second-language learners have further been taken to suggest that listeners recruit motor information to evaluate and facilitate the perception of ambiguous or difficult sound contrasts more generally (e.g., [Callan et al., 2004](#)).

Thus, (left) inferior frontal regions are thought to be involved in binding and evaluating speech information from different domains, in particular during language learning and audiovisual speech perception. Since studies so far have focused on auditory speech perception in infants, it is particularly important to examine the role of inferior frontal regions during infants' perception of audiovisual speech. This will help to examine the role of frontal cortex during phonological processing and its potential involvement in the formation of multimodal sound representations during native language attunement. The current study therefore investigated whether infants undergoing native language attunement recruit inferior frontal brain regions to a different extent during perception of congruent versus conflicting auditory and visual native speech sounds. Based on the previous research discussed above, we predicted a left-lateralized response to speech stimuli in general. Given that infants at the age tested in this study have been shown to discriminate between matching and mismatching auditory and visual speech cues for the stimuli used in the current study ([Altvater-Mackensen and Grossmann, 2015](#); [Altvater-Mackensen](#)

[et al., in press](#)), we further hypothesized that activation in inferior frontal regions might be modulated by audiovisual congruency.

We also assessed how much infants attended to visual speech cues during audiovisual speech perception and how this relates to the recruitment of inferior frontal brain regions. Because infants shift their attention from a speaker's eyes to her mouth during the phase of native language attunement, it has been suggested that attention to visual speech cues supports the formation of multimodal native language categories ([Lewkowicz and Hansen-Tift, 2012](#); see also [Kushnerenko et al., 2013](#)). We therefore hypothesized that infants who show increased attention to visual speech cues might also recruit inferior frontal brain regions more strongly in order to integrate auditory and visual speech information. To examine whether the matching of audiovisual speech cues and infants' attention to visual speech information selectively modulates activation of inferior frontal brain regions, we also assessed brain responses in anterior frontal regions. Although anterior frontal regions have been found to be involved in speech processing, they have mainly been associated with the processing of socially relevant aspects of speech, such as speaker familiarity ([Dehaene-Lambertz et al., 2010](#); [Naoi et al., 2012](#)), and are taken to serve the processing of communicative signals more generally (for a review on functions of prefrontal cortex in infancy see [Grossmann, 2013, 2015](#)).

## Methods

### Participants

Thirty German 5.5- to 6-month-olds (15 girls) from a monolingual language environment participated in the experiment (age range = 5;15 (months; days) to 6;0, mean age = 5;22). All infants were born full term with normal birth weight (>2500 g) and had no reported hearing or vision impairment. Five additional infants could not be tested because they started to cry, and one additional infant had to be excluded because she did not pay attention to the stimuli. Infants were recruited via a large existing infant and child database at the first author's host institution. Parents gave informed consent for their infants to participate in the study and received 7.50 euros and a toy for their infant for participation. The study was approved by the local ethics committee and conducted according to the Declaration of Helsinki.

### Stimuli

Stimuli were taken from a behavioral study testing infants' ability to match auditory and visual vowel cues ([Altvater-Mackensen and Grossmann, 2015](#); [Altvater-Mackensen et al., in press](#)). This study demonstrated 5.5- to 6-month-old infants' successful discrimination of the matching and mismatching speech stimuli used in the current study. Specifically, infants in this previous study showed a familiarity preference for matching over mismatching speech. Furthermore, this study revealed considerable variability in infants' sensitivity to the congruency between auditory and visual speech cues, indicating that the stimuli can be used to track individual differences in speech perception at this age.

Speech stimuli consisted of audiovisual videos of a female native speaker of German, uttering the vowels /a/ and /o/ in hyper-articulated, infant-directed speech (for specific details, see [Altvater-Mackensen and Grossmann, 2015](#); [Altvater-Mackensen et al., in press](#)). For each vowel, two stimulus videos were created that contained three successive repetitions of the respective vowel. Each utterance started and ended with the mouth completely shut in neutral position. Each vowel articulation was separated by approximately 3 s in which the woman kept a friendly open face, leading to a video length of 15 s. The eye gaze was always directed toward the infant. All videos were edited so that they only showed the woman's head against a light-gray background. Video frames were 1024 pixels wide and 1000 pixels high, resulting in a width of 27 cm and a height of 26 cm on screen.

Auditory stimuli were recorded separately from the same woman. The length of the vowels was timed to match the length of the mouthing in the videos. The final stimuli were created by dubbing the audio recordings of the vowels onto the videos of the woman mouthing the vowels. For matching trials, visual and auditory vowel matched, namely, seen /a/ was accompanied by heard /a/, and seen /o/ by heard /o/. For mismatching trials, visual and auditory vowel did not match, namely, seen /a/ was accompanied by heard /o/, and vice versa. Note that visual and auditory stimuli were always time-locked and that tokens used in matching and mismatching trials were identical, only their pairing changed across trial types. See Fig. 1A for an example frame of the mouth position for each of the fully articulated vowels and examples of the sounds' spectrograms. Four additional example trials were created using different recordings of the same woman uttering each vowel in a block of three repetitions followed by an engaging smile and raise of her eyebrows. Each example stimulus had a length of approximately 20 s.

Non-speech stimuli consisted of successive repetitions of three complex sounds, accompanied by time-locked visualizations. Non-speech sounds were melodies that matched the speech sounds in length and volume and represented a bouncing ball, a ringing bell, and a blowing whistle. Non-speech visual stimuli were created by using iTunes' visualizer and showed visual objects (light bubbles) against a black background that changed in color and intensity corresponding to the sound stimuli. The onset of the sounds and video length of the non-speech stimuli were matched to the vowel stimuli, leading to a stimulus length of 15 s. Video frames were 1024 pixels wide and 1000 pixels high, resulting in a width of 27 cm and a height of 26 cm on screen (see Fig. 1B for an example frame of the visual objects used in the non-speech stimuli with the spectrogram of the corresponding sound).

### Procedure

Infants were seated on their parent's lap in a quiet experimental room, facing a 52-cm wide and 32.5 cm high TV screen at a distance of 60 cm from the screen. Visual stimuli were presented in the center of the screen. Auditory stimuli were presented via loudspeakers that were located behind the screen. Infants were first presented with the four example videos, showing the woman uttering /a/ and /o/, in order to introduce the infant to the testing situation and to the speaker and her characteristics. This seemed especially important given that every speaker has a particular way of speaking and detecting

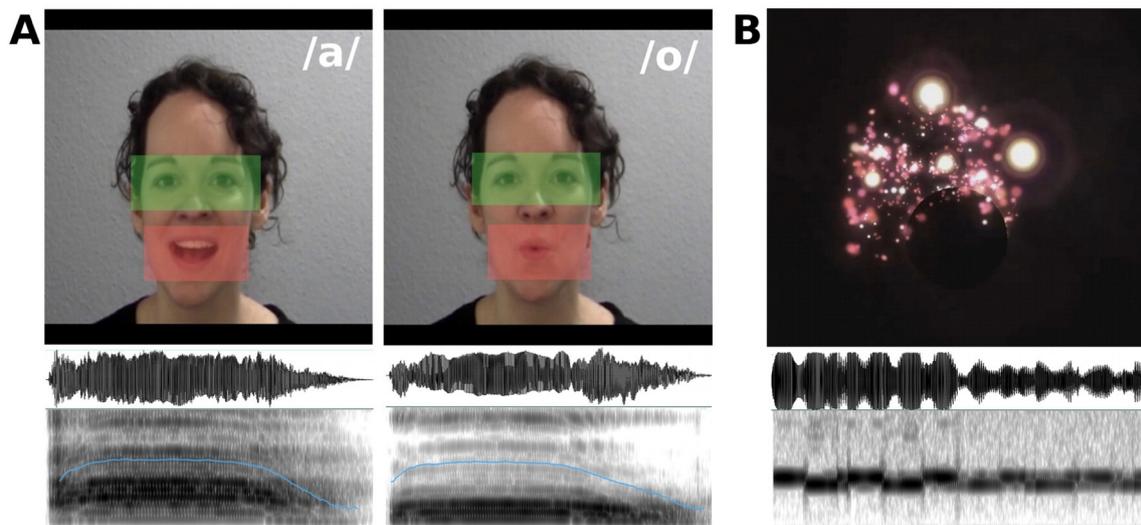
mismatches between auditory and visual speech cues relies on the infant's ability to recognize the auditory and visual stimulus as belonging to a specific vowel category. During presentation of the initial example videos, eye-tracking data were obtained.

Infants were then presented with a maximum of 16 speech–non-speech sequences. Each sequence presented one of eight different audiovisual speech videos (two videos per vowel and matching condition), immediately followed by an audiovisual non-speech video. During presentation of these sequences, fNIRS data were recorded. Sequences were pseudo-randomized so that no more than two consecutive sequences contained the same auditory and/or visual vowel, and so that no more than three consecutive sequences belonged to the same condition. To ensure that infants looked at the screen, each experimental sequence was started manually by the experimenter when the infant attended to the screen, resulting in variable inter-stimulus intervals. On average, the experiment took approximately 15 min.

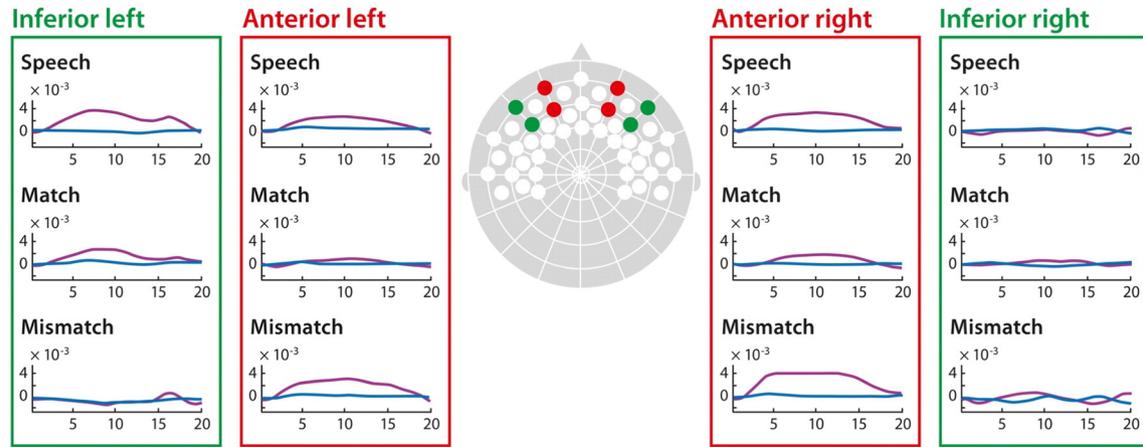
### Data acquisition

A camera mounted below the screen recorded infants' behavior during the experiment to allow offline coding of attention and movement throughout the experiment. The initial four example videos were presented using TobiiStudio software, and infants' eye movements during presentation of these videos were recorded using a Tobii X120 eye-tracker mounted below the screen.

The following speech–non-speech sequences were presented using Presentation software (Neurobehavioral Systems), and infants' fNIRS data during presentation of these sequences were recorded using a NIRx Nirxcout system and NirStar acquisition software. The NIRS method relies on the determination of changes in oxygenated and deoxygenated hemoglobin concentration in cerebral cortex based on their different absorption spectra of near-infrared light (for a detailed description, see Lloyd-Fox et al., 2010). NIRS data were recorded from 16 source-detector pairs, placed at a distance of approximately 2.5 cm within an elastic cap (Easycap) in order to record brain responses from anterior and inferior frontal brain regions. The source-detector arrangement resulted in a total of 49 channels, placed with reference to the 10–20 system (see Fig. 2 for details). Data were recorded at a sampling rate of 6.25 Hz. Near-infrared lasers used two wavelengths at 760 nm and 850 nm with a power of 5 mW/wavelength. Light intensity was automatically adjusted by the system to provide optimal gain.



**Fig. 1.** Stimulus examples. (A) Example frames for each of the fully articulated vowels used as speech stimuli and the corresponding sounds' spectrograms with pitch contours outlined in blue. The rectangles indicate the areas of interest used in the eye-tracking analysis to determine infants' fixations on the speaker's eyes (green) and mouth (red). They were not visible for the infants. (B) Example frame for one of the exploding bubbles used as non-speech stimuli with the corresponding sound's spectrogram.



**Fig. 2.** Channel placement and time course plots of the hemodynamic response to speech stimuli. Dots on the topographic head mark the placement of fNIRS channels. Channels in inferior frontal and anterior frontal areas included in the regions of interest for analysis are colored in green (inferior frontal channels) and red (anterior frontal channels). The panels show the hemodynamic response to speech (upper row), matching speech (middle row), and mismatching speech (lower row) in an exemplary channel in inferior frontal (green panels) and anterior frontal (red panels) regions in the left (left panels) and right (right panels) hemisphere. The depicted inferior frontal channels correspond in placement to F7/F8, targeting the inferior frontal gyrus (Kabdebon et al., 2014), while the depicted anterior frontal channels correspond in placement to FP1/FP2, targeting the prefrontal cortex. The graphs plot the change in oxyHb (red line) and deoxyHb (blue line) from the onset of the speech stimulus for 20 s, i.e. up to 5 s after speech stimulus offset.

### Data analysis

Infants' preference to look at the speaker's mouth compared to her eyes was estimated by analyzing the eye-tracking data collected during the presentation of the initial example videos. Infants' looking pattern across these videos is assumed to reflect a robust individual preference (see also Lewkowicz and Hansen-Tift, 2012; Kushnerenko et al., 2013). Areas of interest of similar size were created that covered the eye and mouth regions of the speaker's face (see Fig. 1A). Infants' total fixation time (FT) during familiarization videos was calculated for each area of interest. For each infant, we further calculated the proportion of time that the infant looked at the speaker's mouth in comparison to her eyes:  $FT_{\text{mouth}} / (FT_{\text{mouth}} + FT_{\text{eyes}})$ . Five infants of the final sample were excluded from the eye-tracking analysis because of technical difficulties to track the infants' eyes (3) or because infants did not pay attention to the videos (2). Thus, only 25 out of 30 infants contributed eye-tracking data to the final analysis.

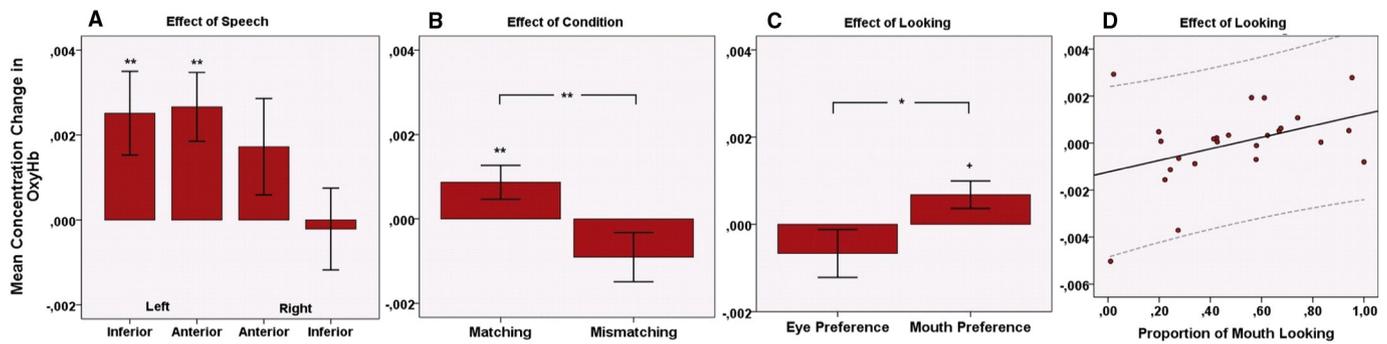
Infants' attention to speech and non-speech stimuli and their movements during the fNIRS recordings was coded offline from the experiment videos. Infants' engagement with the speaker was coded throughout the experimental session by counting visible facial expressions and audible sounds that were (a) smiles or laughs and (b) vocalizations or mouth movements resembling speech movements. The number of videos during which infants engaged with the speaker were comparable across conditions (smiles: matching speech mean = 2.9, range = 1–5, mismatching speech mean = 2.6, range = 1–5; vocalizations: matching speech mean = 1.5, range = 1–2, mismatching speech mean = 1.7, range = 1–3). If an infant looked away from the screen for more than 5 s, that is, for more than one third of a stimulus video, the data for this particular stimulus were excluded from further analysis. If an infant showed severe head movement during presentation of a stimulus video, which resulted in movement artifacts in the data (based on visual inspection), the data for this particular stimulus were also excluded. The final sample consisted of data from 30 infants who on average contributed data from 13.4 non-speech sequences (range = 4–16) and 12.6 speech sequences (range = 5–16), evenly distributed across matching (mean = 6.2, range = 2–8) and mismatching (mean = 6.4, range = 3–8) sequences.

The fNIRS data were then analyzed using the Matlab-based software *nilab2* (see Grossmann et al., 2010a, for previously published fNIRS data using this analysis software). Data were filtered with a 0.2-Hz low-pass filter to remove fluctuations that were too fast and with a high-pass

filter of 30 s to remove changes that were too slow to be related to the experimental stimuli. Using a 15-s time window (equaling the length of each speech and non-speech sequence), measurements were converted into oxygenated hemoglobin (oxyHb) and deoxygenated hemoglobin (deoxyHb) concentrations using the Beer–Lambert law (for a similar method applied to fNIRS data obtained from comparably young infants see Grossmann et al., 2008, 2010a) to calculate relative changes in oxyHb and deoxyHb concentration in response to speech (as compared to the non-speech baseline). Note that we used non-speech audiovisual sequences as a baseline rather than silence and a blank screen so that any (relative) changes in oxyHb and deoxyHb concentration can be interpreted as a functional response to the speech input rather than to audiovisual stimulation *per se*. For further statistical analysis, we averaged the resulting concentration changes in oxyHb and deoxyHb in response to speech by participant for each channel. To isolate channels of interest that covered speech-sensitive brain regions, we conducted one-sample *t*-tests on left- and right-hemispheric frontal channels to detect significant increases in oxyHb in response to audiovisual speech. This revealed four speech-sensitive channels, two adjacent channels each in left inferior and left anterior frontal regions (all  $t(29) > 2.0$ ,  $p < .05$ ), while there were no speech-sensitive channels found in right frontal regions. Our further analysis focused on the resulting two regions of interest (anterior frontal and inferior frontal) in each hemisphere. For each region of interest, relative concentration changes were averaged by participant and experimental condition across the relevant channels for further analysis. According to NIRS channel placement with reference to the 10–20 system and the resulting anatomical correspondences (Kabdebon et al., 2014), the anterior frontal regions mainly targeted the prefrontal cortex whereas the inferior frontal regions mainly targeted the inferior frontal gyrus (corresponding in placement approximately to F7/F8 for the channels located over the inferior frontal gyrus and FP1/FP2 for the channels located over the prefrontal cortex). Fig. 2 indicates the channel placement with the four regions of interest and displays the time courses of the hemodynamic responses to the different speech stimuli for an example channel over right and left anterior frontal and inferior frontal regions of interest.

### Results

One-sample *t*-tests assessing concentration changes in response to audiovisual speech as compared to the audiovisual non-speech baseline



**Fig. 3.** Mean change in oxyHb concentration in left frontal brain regions in response to audiovisual speech. The bar graphs illustrate differences in concentration changes found (A) for speech at inferior and anterior left and right frontal sites, (B) depending on speech condition in left inferior frontal regions (averaged across eye- and mouth-looking infants), and (C) depending on infants' looking preference in left inferior frontal regions (averaged across matching and mismatching speech); error bars indicate  $\pm 1$  SE, asterisks indicate a significance level of  $**p \leq .025$ ,  $*p \leq .05$ ,  $+p \leq .055$ . The scatter plot (D) illustrates the correlation between the proportion of mouth looking and the concentration changes in left inferior frontal regions in response to speech (averaged across matching and mismatching speech); the solid line represents the linear regression line, and the dotted lines mark the 95% confidence interval.

revealed a left-lateralized response to speech with a significant increase in oxyHb in both left inferior ( $t(29) = 2.544$ ,  $p = .017$ ,  $d = .464$ ) and left anterior frontal brain regions ( $t(29) = 3.282$ ,  $p = .003$ ,  $d = .599$ ), but neither in right inferior nor right anterior frontal brain regions ( $ps \geq .14$ ; Fig. 3A).<sup>1</sup> There were no significant changes in concentration of deoxyHb ( $ps \geq .09$ ). Note that based on adult work, one would expect an increase in oxyHb concentration to be accompanied by a decrease in deoxyHb concentration (Obrig and Villringer, 2003). However, changes in deoxyHb seem to be far less consistent in infants, and it is not unusual for infant studies to find no or inconsistent changes in deoxyHb concentration in response to functional stimuli (cf. Lloyd-Fox et al., 2010; Meek, 2002).

From the 25 infants for whom we obtained eye-tracking data during presentation of the initial example videos, 12 infants showed a preference for looking at the speaker's mouth (mean % of mouth looking = 73.04, range = 56.08–100; 5 girls) and 13 infants showed a preference for looking at the speaker's eyes (mean % of eye looking = 72.86, range = 52.90–99.00; 7 girls). Critically, infants' summed attention to the speaker's eyes and mouth did not differ between groups (mean total fixation duration: eye group = 29.56 s, mouth group = 28.89 s). Note that there were also no differences between groups with respect to age, and data contributed to fNIRS analysis or engagement as evidenced by smiles or vocalizations (all  $ps \geq .30$ ). To further investigate how far the audiovisual congruency of speech and infants' individual preference to look at the speaker's mouth modulated the recruitment of left inferior and anterior frontal brain regions during speech perception, we ran separate repeated-measures ANOVAs with condition (matching/mismatching) as within-subject factors and looking preference (eyes/mouth) as between-subjects factor on changes in oxyHb concentration for each site. This analysis revealed no significant main effects or interactions for left anterior frontal brain regions ( $ps \geq .19$ ), and no interaction but a main effect of condition ( $F(1,23) = 5.778$ ,  $p = .025$ ,  $\eta_p^2 = .201$ ) and looking preference ( $F(1,23) = 4.321$ ,  $p = .049$ ,  $\eta_p^2 = .158$ ) for left inferior frontal brain regions.

<sup>1</sup> As can be seen from Figs. 2 and 3A, there appears to be some activation for right anterior frontal regions. However, this activation is not consistent across infants and the rather large variability may explain the absence of a significant effect. Note also, that the time courses in Fig. 2 only represent data from one channel, while Fig. 3 and the statistical analysis relies on data pooled across the two speech-sensitive channels found in each region of interest. As pointed out in the introduction and discussion, previous research indicates that although speech processing tends to be biased to the left hemisphere from early in development, it becomes more strongly left-lateralized in the course of development (cf. Minagawa-Kawai et al., 2011). We therefore suggest that infants who are more advanced in development, i.e., infants who already acquired more sophisticated knowledge about the phonological categories of their native language, will show a more left-lateralized response to speech compared to infants who are less advanced. This possibly leads to larger variation in response patterns in the right hemisphere rendering results insignificant.

Follow-up comparisons revealed that matching speech elicited a greater hemodynamic response than mismatching speech in left inferior frontal brain regions ( $t(24) = 2.460$ ,  $p = .021$ ,  $d = .711$ ; Fig. 3B). Furthermore, infants with a preference to look at the speaker's mouth showed a stronger left inferior frontal brain response to speech than infants with a preference to look at the speaker's eyes ( $t(23) = 2.079$ ,  $p = .049$ ,  $d = .867$ ; Fig. 3C). Pearson correlations further confirmed that concentration changes in oxyHb at left inferior frontal brain regions in response to speech were positively correlated with the proportion of infants' fixations on the speaker's mouth during the eye-tracking measurement ( $r(25) = .396$ ,  $p = .05$ ,  $R^2 = .157$ ; Fig. 3D).<sup>2</sup>

## Discussion

We examined 6-month-old infants' frontal brain responses to matching and mismatching audiovisual speech. Our results revealed that speech-sensitive left inferior frontal brain regions are involved in distinguishing between matching and mismatching audiovisual speech, with matching audiovisual speech evoking greater responses in these brain regions. The current findings further suggest that individual variation in infants' attention to visual speech cues is associated with differences in the recruitment of left inferior frontal brain regions during audiovisual speech perception. Specifically, preferential looking to the mouth when compared to the eyes during audiovisual speech processing correlated positively with infants' left inferior frontal brain responses. These findings suggest that left inferior frontal brain regions play an important role in the development of brain systems considered to be of critical importance for human language (Dehaene-Lambertz et al., 2002; Perani et al., 2011).

The finding that inferior frontal activation in response to speech in 6-month-old infants is left lateralized generally fits with previous reports showing that speech processing is biased to the left hemisphere from early in infancy but becomes more strongly lateralized over the course of the first year of life (cf. Minagawa-Kawai et al., 2011). The current finding that audiovisual speech activates inferior frontal brain regions in infants also agrees with findings in adults. Specifically, in adults, inferior frontal activation has been suggested to reflect the recruitment of

<sup>2</sup> Note that Fig. 3A and the remaining graphs of Fig. 3 do not represent the exact same data sets, but the data sets that formed the basis of each individual statistical analysis. Our initial analysis pooled the data in each region of interest in response to speech (not distinguishing between matching and mismatching speech) and included all 30 infants tested (Fig. 3A). Follow-up analysis on activation differences in left frontal regions pooled the data by speech condition and only included the 25 infants from which we obtained eye tracking data (Figs. 3B, C and D). The apparent difference in overall activation between Fig. 3A and the remaining graphs of Fig. 3 thus results (a) from the different number of infants contributing data to each analysis, (b) from differences in the way the data were pooled, and (c) from differences in variance across data sets.

motor representations during audiovisual speech perception (Ojanen et al., 2005; Skipper et al., 2007). However, contrary to the pattern that has been found in adults, where mismatching audiovisual speech resulted in greater responses, infants activated left inferior frontal brain regions more strongly in response to matching when compared to mismatching audiovisual speech. This might be explained by infants' increased attention to the more familiar matching speech as shown in prior behavioral studies with this age-group (Altvater-Mackensen and Grossmann, 2015; Mugitani et al., 2008). This behaviorally manifested preference may also be reflected in a more elaborate brain processing of familiar as opposed to unfamiliar speech stimuli (see also Cristia et al., 2014, for evidence suggesting that listening to speech in a familiar accent recruits the left fronto-temporal cortex at a younger age than listening to speech in an unfamiliar accent).

Our finding that infants' attention to a speaker's mouth correlates with their responses in left inferior frontal brain regions during audiovisual speech processing supports the notion that left inferior frontal regions are involved in integrating information from different domains, such as vision and audition. In particular, infants who attend more to a speaker's mouth will receive more detailed visual speech information about a specific speech sound, which can in turn be associated with the acoustic (and potentially articulatory) information of this sound. Left inferior frontal regions may play a critical role in associating and representing such multimodal speech information. Specifically, infants who attend more to visual speech information are likely to also recruit left inferior frontal brain regions more strongly in order to integrate and associate speech information from different domains. This scenario is in keeping with previous work suggesting that inferior frontal regions are involved (a) in the formation of acoustic-articulatory maps (Dehaene-Lambertz et al., 2002, 2006; Perani et al., 2011) and (b) in the recruitment of articulatory information during speech perception (Imada et al., 2006; Kuhl et al., 2014). This view further fits models of phonological category learning, stipulating that the perception of speech involves the activation of multimodal representations that link visual, articulatory, and auditory information during early language development (Westermann and Miranda, 2004; see also Bristow et al., 2008). We therefore take our data to suggest that the perception of speech triggers the activation of mental representations that encompass information from different domains, such as acoustic, motoric, and visual speech information, and that left inferior frontal brain regions are involved in the formation and recruitment of such multimodal representations.

Increased attention to visual speech cues (provided by the mouth region) and the associated increased recruitment of left inferior frontal brain regions during audiovisual speech processing might then also represent more mature or more advanced language development in infants undergoing native language attunement. Prior work has shown that while young infants tend to focus on the speaker's eyes, they shift their attention to the speaker's mouth around 6 months of age, which is when they start to attune to the characteristics of their native language (Lewkowicz and Hansen-Tift, 2012). Increased attention to the speaker's mouth is therefore considered to be the more mature looking pattern for infants undergoing native language attunement, and enhanced attention to visual speech information may help with the formation of native language categories. Supporting this notion, recent research finds that infants' recruitment of visual speech information is related to more mature processing of audiovisual speech by 6 to 9 months of age and to better receptive and productive language scores by 14 to 16 months of age (Kushnerenko et al., 2013). Similarly, stronger inferior frontal activation has been argued to represent the more mature neural processing pattern because over the course of the first year of life, infants have been shown to increasingly recruit inferior frontal regions in the left hemisphere during speech perception (Imada et al., 2006). Thus, the positive association between left inferior frontal brain activation and increased looking to the mouth during audiovisual speech perception evidenced in the current data might be a marker of advanced native language attunement.

In addition to the activation in left inferior frontal regions, we also found activation of left anterior frontal regions in response to audiovisual speech. This is in line with findings that medial prefrontal cortex is engaged in the processing of communicative signals (Grossmann et al., 2010b; Grossmann, 2015) and in the processing of socially relevant aspects of speech (Dehaene-Lambertz et al., 2010; Naoi et al., 2012). Brain responses in anterior frontal regions were, however, neither modulated by the match between auditory and visual speech cues nor by infants' looking preference. This suggests that – in contrast to inferior frontal activation – anterior frontal activation is not associated with the processing of the speech's segmental content but with the processing of non-segmental aspects of speech (see also the social gating hypothesis; Kuhl, 2000). This further supports the notion that left inferior frontal activation is specifically related to the processing and formation of multimodal representations of speech sounds.

To conclude, the current study examined audiovisual speech perception in 6-month-old infants, assessing infants' frontal brain responses to matching and mismatching speech cues as well as infants' attention to the speaker's mouth and eyes. Our results demonstrate that both the congruency of audiovisual speech cues and infants' attention to visual speech information modulates brain responses in speech-sensitive left inferior frontal regions. We take this to indicate that left inferior frontal brain regions play a crucial role in integrating information from audition and vision during native language attunement, thereby fostering the learning of multimodal phonological categories.

## References

- Altvater-Mackensen, N., Grossmann, T., 2015. Learning to match auditory and visual speech cues: social influences on the acquisition of phonological categories. *Child Dev.* 86, 362–378.
- Altvater-Mackensen, N., Mani, N., Grossmann, T., 2016. Audiovisual speech perception in infancy: the influence of vowel identity and productive abilities on infants' sensitivity to (mis)matches between auditory and visual speech cues. *Dev. Psychol.* in press.
- Bristow, D., Dehaene-Lambertz, G., Mattout, J., Soares, C., Gliga, T., Baillet, S., Mangin, J.-F., 2008. Hearing faces: how the infant brain matches the face it sees with the speech it hears. *J. Cogn. Neurosci.* 21, 905–921.
- Callan, D.E., Jones, J.A., Callan, A.M., Akahane-Yamada, R., 2004. Phonetic perceptual identification by native- and second-language speakers differentially activates brain region involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. *NeuroImage* 22, 1182–1194.
- Cristia, A., Minagawa-Kawai, Y., Egorova, N., Gervain, J., Filippin, L., Cabrol, D., Dupoux, E., 2014. Neural correlates of infant accent discrimination: a fNIRS study. *Dev. Sci.* 17, 628–635.
- Dehaene-Lambertz, G., Dehaene, S., Hertz-Pannier, L., 2002. Functional neuroimaging of speech perception in infants. *Science* 298, 2013–2015.
- Dehaene-Lambertz, G., Hertz-Pannier, L., Dubois, J., Meriaux, S., Roche, A., Sigman, M., Dehaene, S., 2006. Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proc. Natl. Acad. Sci.* 103, 14240–14245.
- Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allierol, L., Dubois, J., Hertz-Pannier, L., Dehaene, S., 2010. Language or music, mother or Mozart? Structural environmental influences on infants' language networks. *Brain Lang.* 114, 53–65.
- Grossmann, T., 2013. Mapping prefrontal cortex functions in human infancy. *Infancy* 18, 303–324.
- Grossmann, T., 2015. The development of social brain functions in infancy. *Psychol. Bull.* (Online First Publication).
- Grossmann, T., Oberecker, R., Koch, S.P., Friederici, A.D., 2010a. The developmental origins of voice processing in the human brain. *Neuron* 65, 852–858.
- Grossmann, T., Parise, E., Friederici, A.D., 2010b. The detection of communicative signals directed at the self in infant prefrontal cortex. *Front. Hum. Neurosci.* 4, 201.
- Grossmann, T., Johnson, M.H., Lloyd-Fox, S., Blasi, A., Deligianni, F., Elwell, C., Csibra, G., 2008. Early cortical specialization for face-to-face communication in human infants. *Proc. R. Soc. B* 275, 2803–2811.
- Homae, F., Watanabe, H., Nakano, T., Taga, G., 2011. Large-scale brain networks underlying language acquisition in early infancy. *Front. Psychol.* 2, 93.
- Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., Kuhl, P.K., 2006. Infant speech perception activates Broca's area: a developmental magnetoencephalography study. *Neuroreport* 17, 957–962.
- Kabdebon, C., Leroy, F., Simmonet, H., Perrot, M., Dubois, J., Dehaene-Lambertz, G., 2014. Anatomical correlations of the international 10–20 sensor placement system in infants. *NeuroImage* 99, 342–356.
- Kuhl, P.K., 2000. A new view on language acquisition. *Proc. Natl. Acad. Sci.* 97, 11850–11857.
- Kuhl, P.K., Ramirez, R.R., Bosseler, A., Lotus Lin, J.-F., Imada, T., 2014. Infants' brain responses to speech suggest analysis by synthesis. *Proc. Natl. Acad. Sci.* 111, 11238–11245.
- Kushnerenko, E., Tomalski, P., Bailieux, 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. *Front. Psychol.* 4, 432.
- Lewkowicz, D.J., Hansen-Tift, A.M., 2012. Infants deploy selective attention to the mouth of a talking face when learning speech. *Proc. Natl. Acad. Sci.* 109, 1431–1436.
- Lloyd-Fox, S., Blasi, A., Elwell, C.E., 2010. Illuminating the developing brain: the past, present and future of functional near infrared spectroscopy. *Neurosci. Biobehav. Rev.* 34, 269–284.
- Maurer, D., Werker, J.F., 2014. Perceptual narrowing during infancy: a comparison of language and faces. *Dev. Psychobiol.* 56, 154–178.
- Meek, J., 2002. Basic principles of optical imaging and application to the study of infant development. *Dev. Sci.* 5, 371–380.
- Minagawa-Kawai, Y., Cristia, A., Dupoux, E., 2011. Cerebral lateralization and early speech acquisition: a developmental scenario. *Dev. Cogn. Neurosci.* 1, 217–232.
- Mugitani, R., Kobayashi, T., Hiraki, K., 2008. Audio-visual matching of lips and non-canonical sounds in 8-month-old infants. *Infant Behav. Dev.* 31, 307–310.
- Naoi, N., Minagawa-Kawai, Y., Kobayashi, A., Takeuchi, K., Nakamura, K., Yamamoto, J., Kojima, S., 2012. Cerebral responses to infant-directed speech and the effect of talker familiarity. *NeuroImage* 59, 1735–1744.
- Obrig, H., Villringer, A., 2003. Beyond the visible—imaging the human brain with light. *J. Cereb. Blood Flow Metab.* 23, 1–18.
- Ojanen, V., Möttönen, R., Pekkola, J., Jääskeläinen, I.P., Joensuu, R., Autti, T., Sams, M., 2005. Processing of audiovisual speech in Broca's area. *NeuroImage* 25, 333–338.
- Peña, M., Maki, A., Kovacic, D., Dehaene-Lambertz, G., Koizumi, H., Bouquet, F., Mehler, J., 2003. *Proc. Natl. Acad. Sci.* 100, 11702–11705.
- Perani, D., Saccuman, M.C., Scifo, P., Anwander, A., Spada, D., Baldoli, C., Poloniato, A., Lohmann, G., Friederici, A.D., 2011. Neural language networks at birth. *Proc. Natl. Acad. Sci.* 108, 16056–16061.
- Skipper, J.I., van Wassenhove, V., Nusbaum, H.C., Small, S.L., 2007. Hearing lips and seeing voices: how cortical areas supporting speech production mediate audiovisual speech perception. *Cereb. Cortex* 17, 2387–2399.
- Westermann, G., Miranda, E.R., 2004. A new model of sensorimotor coupling in the development of speech. *Brain Lang.* 89, 393–400.
- Wilson, S.M., Saygin, A.P., Sereno, M.I., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702.