Developmental Cognitive Neuroscience 34 (2018) 130–138

Contents lists available at ScienceDirect

Developmental Cognitive Neuroscience

journal homepage: www.elsevier.com/locate/dcn

Modality-independent recruitment of inferior frontal cortex during speech processing in human infants

Nicole Altvater-Mackensen[a,c,*], Tobias Grossmann[b,c]

[a] Department of Psychology, Johannes-Gutenberg-University Mainz, Germany
[b] Department of Psychology, University of Virginia, USA
[c] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

ARTICLE INFO

Keywords:
Infant speech perception
Modality differences
Inferior frontal cortex
fNIRS

ABSTRACT

Despite increasing interest in the development of audiovisual speech perception in infancy, the underlying mechanisms and neural processes are still only poorly understood. In addition to regions in temporal cortex associated with speech processing and multimodal integration, such as superior temporal sulcus, left inferior frontal cortex (IFC) has been suggested to be critically involved in mapping information from different modalities during speech perception. To further illuminate the role of IFC during infant language learning and speech perception, the current study examined the processing of auditory, visual and audiovisual speech in 6-month-old infants using functional near-infrared spectroscopy (fNIRS). Our results revealed that infants recruit speech-sensitive regions in frontal cortex including IFC regardless of whether they processed unimodal or multimodal speech. We argue that IFC may play an important role in associating multimodal speech information during the early steps of language learning.

1. Introduction

Language processing starts long before infants utter their first words. Already in utero, babies start to extract rhythmic regularities of the ambient language: newborns prefer to listen to speech with prosodic characteristics of the native language and their crying matches their native language's stress pattern (Mampe et al., 2009; Mehler et al., 1988; Moon et al., 1993). With only little experience, infants discriminate a range of speech sound contrasts (e.g., Eimas et al., 1971; Werker and Tees, 1984) and match auditory and visual speech cues (e.g., Kuhl and Meltzoff, 1982; Patterson and Werker, 2003). They considerably refine and extend this knowledge in the first year of life and attune perception to the characteristics of their native language. This *native language attunement* is evidenced by changes in the perception of speech sounds, particularly by a reduced sensitivity to non-native sound contrasts and enhanced sensitivity to native sound contrasts (for reviews, see Maurer and Werker, 2014; Jusczyk, 1998). However, most work on the development of infant speech perception has focused on the auditory domain even though large portions of the language input to babies is conveyed through multimodal face-to-face communication.

Indeed, it has been argued that infants exploit visual and social cues inherent in multimodal speech to facilitate language learning and processing (e.g., Kuhl, 2007; Csibra and Gergeley, 2009). In a seminal study, Kuhl and Meltzoff (1982) showed that infants process cross-modal speech information. In this study, when 18- to 20-weeks-old infants were presented with two articulating faces side by side and an auditory stream that matched one of the visual articulations, infants preferred to look at the matching face. Subsequent work revealed that despite this early sensitivity for multimodal speech cues, audiovisual speech perception considerably develops during infancy and this development further extends into childhood. Attunement to the native language can also be observed for audio-visual speech. While young infants are still sensitive to the match between auditory and visual speech cues for both familiar and unfamiliar languages, towards the end of the first year of life, they lose sensitivity for the cross-modal segmental match of non-native audiovisual speech (Pons et al., 2009; Kubicek et al., 2014; Shaw et al., 2015). Similarly, 4- to 6-month-olds discriminate their native from a non-native language when presented with silent articulations, based on visual speech cues alone, whereas monolingual 8-month-olds no longer do so (Weikum et al., 2007). Nevertheless, visual cues continue to be exploited for phonetic learning as they can be used to enhance sound discrimination in infants and adults (Teinonen et al., 2008; Ter Schure et al., 2016; Mani and Schneider, 2013; but see Danielson et al., 2017). Indeed, infants appear to actively seek out visual speech information by increasing attention to

* Corresponding author at: Department of Psychology, Johannes-Gutenberg-University Mainz, Binger Str. 14-16, 55122, Mainz, Germany.
E-mail address: altvater@uni-mainz.de (N. Altvater-Mackensen).

https://doi.org/10.1016/j.dcn.2018.10.002
Received 27 July 2017; Received in revised form 25 August 2018; Accepted 25 October 2018
Available online 30 October 2018
1878-9293/ © 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/BY-NC-ND/4.0/).

a speaker's mouth during the phase of native language attunement (Lewkowicz and Hansen-Tift, 2012; Tenenbaum et al., 2013), and such attention has been associated with more mature speech processing (Kushnerenko et al., 2013). Taken together, these studies demonstrate that infants associate speech sound information from different modalities, suggesting that phonemic representations are inherently multimodal (see also Bristow et al., 2008). They do not, however, necessarily imply that infants automatically integrate information from audition and vision during speech processing (see Shaw and Bortfeld, 2015). Evidence for limited audiovisual integration early in development comes from studies showing reduced sensitivity to the temporal synchrony of auditory and visual speech streams in infants (e.g., Lewkowicz, 2010) and from studies finding considerable variation in the likelihood to fuse auditory and visual speech cues into one percept until well into childhood (Desjardins and Werker, 2004; McGurk and MacDonald, 1976). Interestingly, the developing ability to map auditory and visual speech cues seems to be modulated by articulatory knowledge (Desjardins et al., 1997; Mugitani et al., 2008; Altvater-Mackensen et al., 2016), which has been taken to suggest a role for sensomotoric information in phonological learning and processing (for a recent discussion see Guellai et al., 2014).

Despite this growing body of behavioural work, research on the neural processing of audiovisual speech in infants is relatively scarce and the neural underpinnings of infant audiovisual speech perception are still only poorly understood. Most studies investigating the neural correlates of infant speech perception focused on auditory speech perception and activation patterns elicited in temporal brain areas involved when processing auditory stimuli (for reviews, see Minagawa-Kawai et al., 2011; Rossi et al., 2012). Only more recently, research has also begun to explore the processing of (a) multimodal input and (b) the recruitment of frontal brain regions during infant speech perception (e.g., Altvater-Mackensen and Grossmann, 2016). The former is particularly relevant because most language input to infants is inherently multimodal and – as pointed out above – this multimodality seems to be reflected in infants' phonemic representations. Understanding the processes of audiovisual speech perception is thus crucial for an ecologically valid account of infant speech perception. The latter seems particularly interesting in the light of behavioural findings suggesting a link between articulatory and perceptual abilities. There is a long-standing debate on the potential influence of production processes on speech perception and its link to frontal brain regions, specifically the (left) inferior frontal cortex (for reviews see Galantucci et al., 2006; Poeppel and Monahan, 2011). Investigating the involvement of frontal brain regions during speech processing in infants seems especially relevant as the perceptuo-motor link is thought to be rooted in early language development: infants might use perception to guide and develop their production (Hickok and Poeppel, 2007), they might use their motor knowledge to interpret phonemic information in perception (Pulvermüller and Fadiga, 2010), and/or they might form multimodal representations by establishing a mapping between auditory and articulatory information (Westermann and Miranda, 2004). The current study builds on this line of work and investigates infants' processing of unimodal and multimodal speech in frontal brain regions using functional near-infrared spectroscopy (fNIRS) to add to our understanding of the neural processes underlying infant speech perception.

Before we discuss previous research with infants in more detail, we will briefly summarise relevant findings from the adult literature to put the current study into context. Research with adults has shown that visual speech processing (i.e., silent lip reading) recruits areas in auditory cortex (Calvert et al., 1997; Sams et al., 1991). The response to speech in sensory cortices, such as auditory cortex, and in areas associated with multisensory processing, such as the superior temporal sulcus (STS), is enhanced during audiovisual as compared to unimodal speech perception (Calvert et al., 1999; see also Callan et al., 2001). Further, Broca's area and left inferior prefrontal cortex seem to be critically involved in the processing of audiovisual speech. Both regions are activated more strongly in response to congruent as compared to incongruent audiovisual speech (Ojanen et al., 2005; see also Skipper et al., 2007). While activation in Broca's area is thought to reflect the mapping of auditory cues onto motor representations (e.g., Wilson et al., 2004), activation in prefrontal cortex is thought to reflect processes of attention allocation (e.g., Miller and Cohen, 2001). In sum, these studies show enhanced processing of audiovisual as compared to unimodal speech in sensory areas, the STS and left inferior frontal and prefrontal cortex. This provides a neural substrate to behavioural improvements in speech perception when a speaker's mouth is visible (e.g., Schwartz et al., 2004) and concurs with findings that auditory processing is affected by visual cues and vice versa (see also McGurk and MacDonald, 1976). While temporal areas and specifically the STS have been demonstrated to be involved in binding and integrating information from different modalities (e.g., Beauchamp et al., 2004; but see Hocking and Price, 2008), the role of frontal regions during speech perception is more controversially discussed. In particular, enhanced activation in left inferior frontal cortex has been taken to reflect the mapping of speech percepts onto motor schemes – either through a direct action-perception link (e.g., Pulvermüller and Fadiga, 2010) or in the context of a predictive coding mechanism (e.g., Hickok et al., 2011). Regardless of the proposed nature of sensomotoric mapping, the reported frontal effects in adult speech perception are commonly assumed to result from the association of speech cues from different modalities during language development (see Dick et al., 2010).

Indeed, neuroimaging studies with infants show activation of left inferior frontal regions in response to speech already in newborns and 3-month-old infants (Dehaene-Lambertz et al., 2002, 2006; Peña et al., 2003). Specifically, over the course of the first year of life, Broca's area has been shown to become more prominently recruited during speech processing, suggesting the establishment of a perceptuo-motor link for speech categories during early language development (Imada et al., 2006; Kuhl et al., 2014; Perani et al., 2011). Yet, only very few studies explicitly tested infants' processing of multimodal speech cues. When auditory speech is presented alongside a visual non-speech stimulus, such as a checkerboard pattern, 2- to 4-month-olds show similar activation of sensory cortices for unimodal and multimodal stimuli (Taga and Asakawa, 2007; Taga et al., 2003). This is thought to indicate that early in ontogeny auditory and visual components are processed in separate neural circuits with little cross-talk between sensory regions (but see Watanabe et al., 2013). In contrast, 6- to 9-month-olds show left-lateralised enhancement of activation in temporal areas in response to auditory speech paired with a visual stimulus as compared to auditory speech alone (Bortfeld et al., 2007, 2009). Taken together, these results suggest that enhanced processing of multimodal speech might rely on language experience. However, visual cues in these studies were non-linguistic. Thus, it is unclear if results generalise to the processing of audiovisual speech. Yet, similarly enhanced processing for synchronous multimodal as compared to unimodal or asynchronous speech has been reported in ERP studies with 3- and 5-month-old infants (Hyde et al., 2010, 2011; Reynolds et al., 2014). Moreover, Fava et al. (2014) report increased left-lateralised activation in response to audiovisual native as compared to non-native speech in temporal brain regions by the end of the first year of life, providing evidence for a neural correlate of perceptual native language attunement in the audiovisual domain (for similar findings with auditory-only speech see, e.g., Kuhl et al., 2014).

While these studies suggest improved processing of native audiovisual as compared to auditory-only and non-native speech, no study so far directly compared the processing of auditory, visual and audiovisual speech in infants. Furthermore, brain responses were only assessed for sensory cortices and temporal regions. This concurs with well-established findings in the adult literature, highlighting the importance of temporal regions in speech processing and specifically the role of STS for binding information from different modalities during speech perception (e.g., Nath and Beauchamp, 2012; Baum et al., 2012). Yet, the

results cannot speak to the potential role of inferior frontal cortex (IFC) in processing speech information from different modalities during language development. Currently, only one study directly assessed infants' recruitment of frontal cortex during audiovisual speech perception (Altvater-Mackensen and Grossmann, 2016). This study reported enhanced activation of regions in IFC in the left hemisphere in response to congruent – but not incongruent – native audiovisual speech in 6-month-old infants. Furthermore, results from this study show that infants' response to audiovisual speech in inferior frontal brain regions is impacted by their general attention to a speaker's mouth during speech perception. This finding is in line with the notion that left IFC is involved in mapping information from different modalities during infant speech perception.

To further illuminate the role of IFC during infant language learning and speech perception, the current study examined infants' processing of auditory, visual and audiovisual speech using fNIRS. In particular, we investigated 6-month-olds' neural response to speech across modalities at frontal and prefrontal sites in both hemispheres. Prefrontal sites were included because prefrontal cortex has been suggested to be involved in processes of attention control during audiovisual speech perception in adults (Ojanen et al., 2005) and in processing of socially – but not necessarily linguistically – relevant aspects of speech in infants (Dehaene-Lambertz et al., 2010; Naoi et al., 2012). We hypothesised that infants might differentially recruit areas in IFC in response to multimodal as compared to unimodal speech, reflecting differences in sensory stimulation, attention and/or task demands (for discussion of crossmodal, additive effects in adults see Calvert, 2001). This hypothesis is also based on previous reports of enhanced temporal activation for multimodal stimuli in infants (Bortfeld et al., 2007, 2009). Indeed, if areas in IFC are associated with integration processes during audiovisual speech perception (e.g., Dehaene-Lambertz et al., 2002, 2006), activation should be stronger for multimodal as compared to unimodal speech because only audiovisual speech input requires the evaluation and integration of information from different modalities. However, it has also been suggested that activation of areas in left IFC, such as Broca's area, during auditory speech perception reflect the mapping of motor schemes onto speech percepts (e.g., Kuhl et al., 2014). This implies that speech perception leads to the automatic retrieval and processing of information from different domains, such as auditory, visual and motoric information, regardless of the modality of the input (see also Westermann and Miranda, 2004). That is, infants might retrieve (visuo-) motoric information not only when they are presented with a talking face but also when they are presented with auditory-only speech (as suggested by motor theory of speech perception, for instance, see Liberman, 1957). If so, processing demands should be similar across modalities and we might thus find similar patterns of activation for unimodal and multimodal speech.

2. Method and materials

2.1. Participants

Twenty-eight German 5.5- to 6-month-olds (13 girls) from a monolingual language environment participated in the experiment (age range: 5;15 (months; days) to 6;0, mean age 5;24). Infants were recruited via a large existing infant and child database at the Max Planck Institute for Human Cognitive and Brain Sciences in Leipzig, Germany. All infants were born full term with normal birth weight (> 2500 g) and had no reported hearing or vision impairment. Eight additional infants could not be tested because they started to cry, four additional infants had to be excluded due to technical failure during fNIRS recording, and three additional infants were later excluded from analysis because they contributed less than 50% of valid data (see data analysis). Parents gave written informed consent to participate in the study and received 7.50 Euro 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.

2.2. Stimuli

Stimuli were adapted from a previous study testing infants' audiovisual speech perception (Altvater-Mackensen and Grossmann, 2016). Speech stimuli consisted of audiovisual recordings of a female native speaker of German, uttering the vowels /a/, /e/ and /o/ in hyper-articulated, infant-directed speech (for details on the visual and acoustic characteristics see Altvater-Mackensen et al., 2016). 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 three seconds in which the woman kept a friendly facial expression, leading to a video length of 15 s. The eye-gaze was always directed towards the infant. All videos were zoomed and cropped so that they only showed the woman's head against a light-grey wall. 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. For the auditory-only stimuli, the visual stream of the videos was replaced by a blank (black) screen. For the visual-only stimuli, the auditory stream of the videos was replaced by silence. Three additional example trials were created using different recordings of the same woman uttering each vowel twice in a block of three repetitions followed by an engaging smile and raise of her eyebrows. Each example stimulus had a length of approximately 45 s.

Non-speech stimuli were created to mimic the speech stimuli. Each stimulus consisted of three successive repetitions of a complex sound, accompanied by a time-locked visualisation. Non-speech sounds were three different melodies that matched the three speech sounds in length and volume and represented a bouncing ball, a ringing bell, and a whistle. Non-speech visual stimuli were created using i-Tunes' visualiser and showed visual objects (light bubbles) against a black background that changed in colour and intensity corresponding to the sound stimuli. Auditory and visual streams of the non-speech stimuli were contingent, i.e. bubble explosions were time-locked to the sounds, to mimic the synchrony between auditory and visual speech streams in the speech stimuli. Timing 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. Again, auditory-only stimuli were created by replacing the visual stream of the audiovisual videos with a blank (black) screen, and visual-only stimuli were created by replacing the auditory stream of the audiovisual videos with silence. Fig. 1 shows an example frame of the mouth position for each of the fully articulated vowels and examples of the sounds' spectrograms (Fig. 1A) and an example frame of the visual objects used in the non-speech stimuli with the spectrogram of the corresponding sound (Fig. 1B).

2.3. 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 on screen. Auditory stimuli were presented via loudspeakers that were located behind the screen. Infants were first presented with the three example videos showing the woman uttering /a/, /e/ and /o/, to introduce infants to the testing situation and to the speaker and her characteristics. Infants were then presented with a maximum of 18 speech – non-speech sequences. Each sequence presented one of six different auditory, visual and audiovisual speech videos (two videos per vowel and modality), immediately followed by a modality-matched non-speech video. Speech and non-speech videos were paired so that each specific speech video was always followed by the same modality-matched non-speech video. Sequences were pseudo-randomised so that no more than two

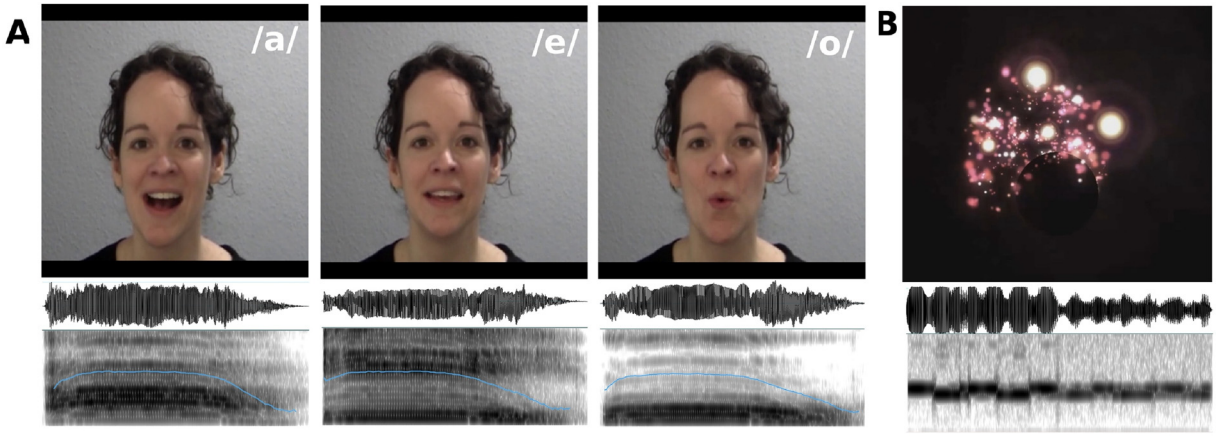

Fig. 1. Stimulus examples. (A) Example frames for each of the fully articulated vowels used as speech stimuli and the corresponding sounds' spectrograms with pitch contours outlined in blue. (B) Example frame for one of the exploding bubbles used as non-speech stimuli with the corresponding sound's spectrogram.

consecutive sequences belonged to the same modality, and so that no more than two consecutive sequences contained the same vowel. Thus, stimuli from the three different modalities were not presented in a blocked design but intermixed throughout the experiment. Although this might decrease the likelihood of task-dependent modality-specific processing, pilot testing showed that an intermixed presentation increased infants' interest in the stimuli and considerably reduced dropout rates. To ensure that infants looked at the screen, each sequence was started manually by the experimenter when the infant attended to the screen, resulting in variable inter-stimulus intervals (median = 8 s, range 1–60 seconds; note that this excludes longer breaks which were occasionally taken to reposition or cheer up the infant). On average, the experiment took approximately 15 min.

2.4. Data acquisition

A camera mounted below the screen recorded infants' behaviour during the experiment to allow offline coding of attention and movement throughout the experiment. The speech – non-speech sequences were presented using Presentation software (Neurobehavioral Systems), and infants' fNIRS data were recorded using a NIRx Nirscout system and NirStar acquisition software. The NIRS method relies on determination of changes in oxygenated and deoxygenated haemoglobin concentration in cerebral cortex based on their different absorption spectra of near-infrared light (for a detailed description see Lloyd-Fox et al., 2010). 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). To ensure comparable placement of channels irrespective of infants' head size, several caps of different size were prepared with optode holders. Based on an individual infant's head size, the best fitting cap was used for testing. 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 NIRS recording system to provide optimal gain.

2.5. Data analysis

Infants' attention to speech and non-speech stimuli and their movements during fNIRS recordings were coded offline from video. If an infant looked away from the screen for more than 5 s, which 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 artefacts in the data (based on visual inspection), the data for this particular stimulus were also excluded. Three infants were excluded from analysis because they did not contribute data for at least 50% of the stimuli according to these criteria (for similar rejection criteria, see Altvater-Mackensen and Grossmann, 2016). The final sample consisted of data from 28 infants that contributed on average data from 15 speech sequences (range 9–18). Infants' tended to look away more often during auditory-only sequences and consequently contributed fewer auditory trials to the analysis compared to audiovisual trials (audiovisual: mean 5.36, SE 0.18, range 3–6; auditory: mean 4.61, SE 1.1, range 2–6; visual: mean 5.04, SE 1.04, range 3–6; audiovisual vs. auditory: $t(27) = 2.938$, $p = .01$; other $ps > .05$).

The fNIRS data were analysed using the Matlab-based software nilab2 (see Grossmann et al., 2010, 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 (equalling the length of each speech and non-speech sequence), measurements were converted into oxygenated haemoglobin (oxyHb) and deoxygenated haemoglobin (deoxyHb) concentrations using the modified Beer-Lambert law. We then calculated changes in oxyHb and deoxyHb concentration in response to speech relative to the non-speech baseline (for a similar method applied to fNIRS data obtained from infants of similar ages, see Grossmann et al., 2008, 2010; Altvater-Mackensen and Grossmann, 2016). Note that we used modality-matched non-speech sequences as a baseline rather than silence and a blank screen so that any (relative) change in oxyHb and deoxyHb concentration can be interpreted as a response to the speech stimuli rather than to sensory auditory, visual or audiovisual stimulation *per se* (but see 4. Discussion for alternative interpretations). Pilot testing further showed that a modality-matched baseline considerably reduced infant movement and fussiness compared to a baseline without stimulation (for the use of a similar baseline, see Altvater-Mackensen and Grossmann, 2016). For subsequent statistical analysis, we averaged the resulting concentration changes in oxyHb and deoxyHb by participant for each channel. Note that we report results on concentration changes in deoxyHb, but that it is not unusual for studies with infants to find no or inconsistent changes in deoxyHb concentration in response to functional stimuli (cf. Lloyd-Fox et al., 2010; Meek, 2002)

3. Results

Based on previous research (Altvater-Mackensen and Grossmann, 2016), we conducted one-sample t-tests on left- and right-hemispheric

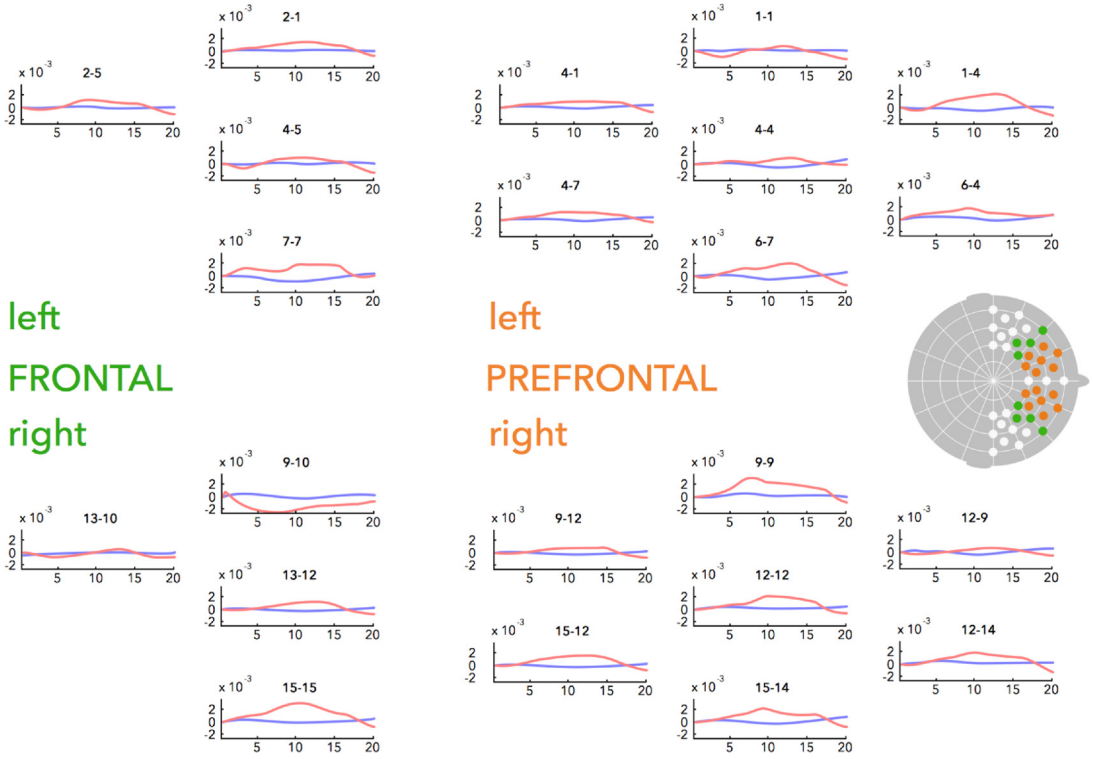

Fig. 2. Channel placement and time course plots of the hemodynamic response to speech stimuli. Dots on the topographic head mark the placement of fNIRS channels. Channels in frontal and prefrontal areas included in the regions of interest for analysis are coloured in green (frontal channels) and red (prefrontal channels). The panels show the hemodynamic response to speech for all channels included in the analysis in the left (upper panels) and right (lower panels) hemisphere. Note that not all depicted channels show significant changes in oxyHb in response to speech (see 3. Results for details). The depicted frontal channels correspond in placement approximately to F7/F8, targeting the inferior frontal gyrus and lower parts of the middle frontal gyrus (Kabdebon et al., 2014), while the depicted prefrontal channels roughly correspond in placement to FP1/FP2, targeting the medial prefrontal cortex. The graphs plot the change in oxyHb (red line) and deoxyHb (blue line) from the onset of the speech stimulus (averaged across all speech conditions) for 20 s, i.e. up to 5 s after speech stimulus offset.

frontal channels to detect significant increases in oxyHb in response to speech (collapsed across modalities) to isolate channels of interest that covered speech sensitive brain regions. This revealed two clusters of speech-sensitive channels: four adjacent channels in left frontal regions and seven adjacent channels in right prefrontal regions (all channels $t(27) \geq 2.539$, $p \leq .017$, no corrections applied). To test for potential lateralisation effects, we included the corresponding channels in the opposite hemisphere for subsequent analysis. Our analysis was thus conducted on the resulting two regions of interest (frontal and prefrontal) in each hemisphere (see Fig. 2).[1] 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 prefrontal regions mainly targeted the medial prefrontal cortex whereas the frontal regions mainly targeted the inferior frontal gyrus and lower parts of the middle 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 for all channels included in the analysis.

A repeated-measures ANOVA on mean concentration changes in oxyHb in response to speech with site (frontal, prefrontal), hemisphere (left, right) and modality (audiovisual, auditory, visual) as within-subject factors revealed an interaction between site and hemisphere ($F(1,27) = 8.175$, $p = .008$, $\eta_p^2 = .232$). No other interactions or main effects reached significance ($ps \geq .21$). Separate repeated-measures ANOVAs with hemisphere (left, right) and modality (audiovisual, auditory, visual) as within-subject factors revealed a main effect of hemisphere at frontal sites ($F(1,27) = 5.135$, $p = .032$, $\eta_p^2 = .160$) and prefrontal sites ($F(1,27) = 6.205$, $p = .019$, $\eta_p^2 = .187$). No other interactions or main effects reached significance ($ps \geq .46$). Corresponding analysis on mean concentration changes in deoxyHb in response to speech revealed no significant main effects or interactions at frontal sites, and a main effect of hemisphere ($F(1,27) = 11.874$, $p = .002$, $\eta_p^2 = .305$) and modality ($F(1,26) = 4.055$, $p = .029$, $\eta_p^2 = .238$), but no interaction at prefrontal sites.

Follow-up analysis showed significant increases in oxyHb concentration for speech at both left ($t(27) = 4.084$, $p \leq .001$, $d = 0.77$) and right ($t(27) = 2.498$, $p = .019$, $d = 0.47$) frontal sites, and both left ($t(27) = 3.875$, $p = .001$, $d = 0.73$) and right ($t(27) = 7.365$, $p \leq .001$, $d = 1.39$) prefrontal sites.[2] Concentration changes were stronger

[1] The identified speech-sensitive channels spanned inferior and superior frontal brain regions. Based on channel placement, we split the resulting channels in inferior and superior clusters for initial analysis. Because there were no main effects or interactions with respect to the anatomical inferior-superior distinction ($ps \geq .09$ for oxyHb and $ps \geq .07$ for deoxyHb), we collapsed the data across inferior and superior channels at frontal and prefrontal sites in each hemisphere for analysis.

[2] Note that results for the follow-up analysis and the planned comparisons remain similar when controlling for false positives in multiple comparisons through the Benajamini-Hochberg procedure with a false discovery rate of 0.05 (Benjamini and Hochberg, 1995). Controlling for the false discovery rate with the Benjamini-Hochberg procedure seems more appropriate than controlling for the familywise error rate with Bonferroni-corrections given the limited power of our data set. Using the more conservative Bonferroni-correction the increase in oxyHb for speech at right frontal sites and the difference in oxyHb between

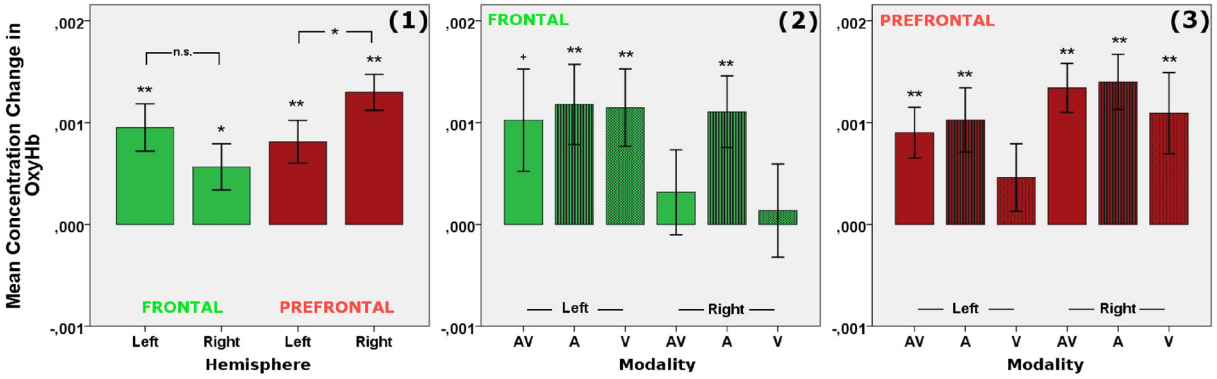

Fig. 3. Mean change in oxyHb concentration in frontal and prefrontal brain regions in response to speech. The bar graphs illustrate differences in concentration changes found at frontal and prefrontal sites for speech collapsed across modalities ((1), left panel), and at frontal ((2), middle panel) and prefrontal sites ((3), right panel) depending on modality with AV = audiovisual speech (solid bars), A = auditory speech (striped bars), and V = visual speech (dotted bars); error bars indicate +/- 1 SE, asterisks indicate a significance level of ** $p \leq .01$, * $p \leq .02$, $^{+}$ p = .05.

in the right than the left hemisphere for prefrontal sites ($t(27) = -2.505$, $p = .019$, $d = -0.47$), while there was no significant difference between hemispheres for frontal sites ($p = .10$, see Fig. 3.1). Corresponding analysis showed a significant decrease in deoxyHb concentration for speech at prefrontal left sites ($t(27) = -3.047$, $p = .005$, $d = -0.57$), but not at prefrontal right sites or at frontal sites in either hemisphere ($ps > .20$).

Planned analysis on the influence of modality showed significant increases in oxyHb concentration for audiovisual ($t(27) = 2.037$, $p = .052$, $d = 0.38$), auditory ($t(27) = 2.998$, $p = .006$, $d = 0.57$) and visual ($t(27) = 3.018$, $p = .005$, $d = 0.57$) speech at frontal left sites and for auditory ($t(27) = 3.141$, $p = .004$, $d = 0.59$) – but not for audiovisual or visual ($ps \geq .45$) – speech at frontal right sites (see Fig. 3.2). There was no significant difference between modalities within each hemisphere ($ps \geq .07$) or between hemispheres for either modality ($ps \geq .14$). Corresponding analysis showed no significant changes or differences in deoxyHb concentration at frontal sites ($ps > .07$).

At prefrontal sites, oxyHb concentration significantly increased for audiovisual ($t(27) = 3.611$, $p = .001$, $d = 0.68$) and auditory ($t(27) = 3.248$, $p = .003$, $d = 0.61$) – but not for visual ($p = .17$) – speech in the left hemisphere and for audiovisual ($t(27) = 5.579$, $p \leq .001$, $d = 1.05$), auditory ($t(27) = 5.159$, $p \leq .001$, $d = 0.97$) and visual ($t(27) = 2.739$, $p = .011$, $d = 0.52$) speech in the right hemisphere (see Fig. 3.3). Again, concentration changes did not differ between hemispheres for either modality ($ps \geq .08$) or between modalities within each hemisphere ($ps \geq .18$). Corresponding analysis showed a significant decrease in deoxyHb concentration for visual speech at prefrontal left sites ($t(27) = -3.570$, $p = .001$, $d = -0.67$), which was significantly different from changes in response to audiovisual ($t(27) = -3.032$, $p = .005$, $d = -0.56$) and auditory speech ($t(27) = 2.400$, $p = .024$, $d = -0.64$) and significantly stronger than in the right hemisphere ($t(27) = -2.711$, $p = .012$, $d = -0.56$). No other changes or differences in deoxyHb concentration at prefrontal sites reached significance ($ps > .07$).

Fig. 3 displays mean concentration changes in oxyHb in response to speech at frontal and prefrontal sites for both left and right hemisphere, collapsed across modality (Fig. 3.1) and separated by modality (Fig. 3.2 and 3.3).

4. Discussion

The current study tested 6-month-old infants' neural response to auditory, visual and audiovisual speech stimuli to assess modality-specific effects in speech processing. Our results revealed the recruitment of speech-sensitive regions in frontal and prefrontal cortex in both hemispheres for uni- and multimodal speech.

Before we discuss results in more detail, it is important to emphasise that we used a modality-matched baseline rather than silence and a blank screen to assess changes in response to speech. This ensures that the brain responses reported in the current study cannot be reduced to a basic response to sensory auditory, visual or audiovisual stimulation, but can be interpreted as a functional response to the speech input. One might, however, argue that the observed response is not specifically related to speech processing but to face-voice processing more generally. From birth, infants prefer faces and voices over other kinds of visual and auditory stimuli (Johnson et al., 1991; Vouloumanos et al., 2010). Thus, the salience of the speaker's face/voice in our speech stimuli might by itself increase attention and impact processing. Since we did not include a control condition using facial non-speech movements (such as gurns, e.g., Calvert et al., 1997) and non-speech vocal sounds (such as grunting), we cannot rule out this possibility. An experiment including such control conditions for all modalities would have been too long to run with 6-month-old infants. Nevertheless, it would be important for future studies to directly compare speech and non-speech conditions that both involve facial and vocal stimuli. Previous findings can indeed be taken to suggest that the distinction between speech and non-speech facial movements and articulatory gestures is not clear-cut early in infancy. First, infants do not only attune their speech perception to the ambient input in the first year of life but also their face perception (Maurer and Werker, 2014). The impact of face perception on audiovisual speech processing might thus change over the course of the first year of life. Second, infants are initially able to match auditory and visual cues not only for human speech but also for human non-speech sounds (Mugitani et al., 2008) and for monkey calls (Lewkowicz and Ghazanfar, 2006), suggesting a broad ability to match multimodal information. Nevertheless, we argue that the observed frontal activation is functional to speech for several reasons. Adult studies found stronger activation of IFC in response to visual speech as compared to facial non-speech movements (Calvert et al., 1997; Campbell et al., 2001; Hall et al., 2005), suggesting a response that is specific to speech rather than to face-voice processing more generally. Furthermore, adults' activation of IFC during processing of visual-only or degraded audiovisual speech is modulated by individual differences in speech reading and learning abilities (Paulescu et al., 2003; Eisner et al., 2010; McGettigan et al., 2012). This is in line with the notion that IFC activation is modulated by linguistic task demands.

(*footnote continued*)
prefrontal right and left sides are no longer significant in the follow-up analysis (adjusted $p = 0.11$), and only increases in oxyHb for audiovisual speech at left prefrontal sites and for audiovisual and auditory speech at right prefrontal sites remain significant in the planned comparisons (adjusted $p = 0.003$).

Of course, findings with adults cannot be generalised to infants. Yet, previous research with infants shows differential activation of IFC for matching and mismatching audiovisual speech with the same age group and similar stimuli as used in the current study (Altvater-Mackensen and Grossmann, 2016). This suggests that IFC is involved in mapping speech information from different modalities in infants and that the current experimental design taps into speech processing. This notion is further supported by the finding that in this prior study IFC activation in response to audiovisual speech stimuli correlated with infants' attention to the speaker's mouth (as assessed through eye tracking preceding the fNIRS recording; Altvater-Mackensen and Grossmann, 2016). In addition, infants' behavioural response to the same matching and mismatching speech stimuli has been shown to be modulated by infants' articulatory knowledge (Altvater-Mackensen et al., 2016), pointing to a potential role for production processes – which might be modulated by the IFC – in speech perception at this age and for these stimuli.

To summarise the main finding, six-month-olds were found to recruit frontal brain regions during processing of auditory-only, visual-only and audiovisual speech. Increased activation of regions in frontal cortex in response to speech was neither significantly modulated by modality nor were there significant differences in activation across hemispheres.[3] The finding that speech processing was not left-lateralised contrasts with our earlier findings on the processing of congruent compared to incongruent audiovisual speech in 6-month-olds (Altvater-Mackensen and Grossmann, 2016). Yet, there is considerable individual variation in the recruitment of IFC during infants' audiovisual speech perception (Altvater-Mackensen and Grossmann, 2016; see also Imada et al., 2006) and even though speech processing is biased to the left hemisphere from early on in infancy, it has been shown to become more strongly lateralised over the first year of life (Minagawa-Kawai et al., 2011). In general, our results replicate previous findings, demonstrating that infants recruit areas in IFC during speech perception (Dehaene-Lambertz et al., 2002; Imada et al., 2006; Kuhl et al., 2014; Peña et al., 2003; Perani et al., 2011) and support the notion that IFC plays a critical role for language learning and processing from early in ontogeny. Interestingly, we did not find systematic differences in the activation of frontal brain regions during speech processing with respect to modality. This is unexpected given that speech processing in adults is modulated by the congruency of audiovisual speech as well as by language modality (e.g., Calvert et al., 1999; Ojanen et al., 2005). Arguably, modality-specific processing might have been weakened by the fact that we presented auditory, visual and audiovisual stimuli in random order rather than in a blocked design. It would therefore be interesting for future studies to assess if modality-specific effects in infants are modulated by the specifics of stimulus presentation. However, given that there were no consistent differences across modalities, we take our results to suggest that IFC is recruited during uni- and multimodal speech processing in infants.

There are different possible interpretations of the results with respect to the functional role that the IFC plays during infant speech perception. In the adult literature, activation of IFC during speech perception has been found to be modulated by task demands, specifically by the congruence of audiovisual speech information (Ojanen et al., 2005), by stimulus clarity (McGettigan et al., 2012) and by stimulus complexity (syllables vs. words vs. sentences; Peelle, 2012). This is in line with the notion that IFC activation is associated with top-down processes related to attentional control and memory (Song et al., 2015; Friederici, 2002). On a more specific level, activation of the IFC has been taken to suggest activation of motor schemes in the service of speech perception.[4] However, theoretical accounts fundamentally differ in their conception of this perceptuo-motor link. Models taking an embodied approach to speech perception assume that the incoming speech signal is analysed in terms of the associated articulatory information (e.g., motor theory of speech perception, Liberman, 1957; and mirror neuron approaches, Pulvermüller and Fadiga, 2010). According to such models, phonemic information is represented in terms of motor schemes and the motor system is critical for speech perception (for reviews see Galantucci et al., 2006; Cappa and Pulvermüller, 2012). Other accounts assume that the motor system is activated during speech perception to inform production, i.e., to provide corrective feedback and to guide speech gestures (e.g., Hickok et al., 2011, for auditory speech perception; Venezia et al., 2016, for visual speech perception). In this view, the perceptuo-motor link results from the need to tune production to the native sound system in early development. This link may, however, be exploited to predict other's upcoming speech (see also Scott et al., 2009) and to limit potential interpretations of the speech signal (see also Skipper et al., 2007). Similar interpretations of the perceptuo-motor link in terms of phonological learning (Perani et al., 2011) and phonological analysis (Kuhl et al., 2014) can be found in the infant literature.

In the light of this discussion, it is interesting to note that there is increasing behavioural evidence that articulatory information modulates infants' speech perception. First, research suggests that articulatory knowledge acts as perceptual filter, focusing attention to sound contrasts relevant to concurrent phonological learning in production (Vihman, 1996; Majorano et al., 2014). The motor system might thus exert a top-down influence to guide productive development. Second, articulatory knowledge correlates with infants' ability to map auditory and visual speech cues during audiovisual speech perception (Desjardins et al., 1997; Altvater-Mackensen et al., 2016). This might suggest that infants recruit the motor system during speech perception to retrieve articulatory information that can be used to predict sensory outcomes of (visuo-motoric) mouth movements. Third, concurrent sensomotoric information affects infants' auditory and audiovisual speech perception: infants fail to discriminate sound contrasts that are associated with different tongue tip positions when the tongue is blocked by a pacifier (Yeung and Werker, 2013; Bruderer et al., 2015). This might indicate that infants use articulatory information to interpret the incoming speech signal, i.e., that the motor system is recruited to analyse speech information. Our data do not allow us to disentangle these positions. For future studies it will be important to directly assess to what extent infants' recruitment of IFC during speech perception is related to productive development in babbling and to (silent) imitation processes, for instance by measuring concurrent fascial muscle activity in order to investigate the contribution of articulatory information to infant speech perception.

Given the similar neural response to auditory, visual and audiovisual speech, we take our findings to support the view that infant speech processing involves retrieval and mapping of phonological information from different domains. In particular, our findings are in line with models assuming that phonological representations are inherently multimodal and reflect the association of auditory, visual and motor information in the course of early language development (Westermann and Miranda, 2004). According to this model, hearing or seeing a speech stimulus leads to automatic (co-)activation and retrieval of multimodal phonological information irrespective of the specific modality of the input stimulus itself. In combination with previous findings on the recruitment of the IFC during language processing in infants

[3] Note that visual inspection of the data suggests that the neural response to audiovisual and visual speech might be attenuated in the right compared to the left hemisphere in frontal regions (cf. Fig. 3.2). This difference was, however, not significant in direct planned comparisons.

[4] It should be noted that inferior frontal cortex and specifically Broca's area have mainly been associated with syntactic processing and higher-level unification processes (e.g. Hagoort, 2014; Friederici, 2011). Yet, this is not directly relevant to our study given infants' limited syntactic capabilities and the simple non-referential, syllabic structure of our stimuli.

(Altvater-Mackensen and Grossmann, 2016; Imada et al., 2006; Kuhl et al., 2014), we take our data to suggest that the IFC is pivotal in the learning and processing of such multimodal phonological representations.

In addition to activation in IFC, our results show increased activation of right and left prefrontal sites in response to speech with stronger effects in the right hemisphere. Again, we found no consistent differences in neural responses with respect to modality. This suggests that information from the face and the voice elicit similar patterns of activation in infants' prefrontal cortex. These findings are in agreement with previous reports of right-lateralised activation of prefrontal cortex in response to socially relevant stimuli, such as speech (for a review see Grossmann, 2013, 2015), and speak to theories that assign a central role to social information in infants' language learning and processing (e.g., *social gating hypothesis*, Kuhl, 2007). Prefrontal cortex might thus serve to evaluate the social relevance of the perceived speech input and to modulate attention to speech more generally. Such a mechanism of relevance evaluation and attention control with respect to language input is in line with theoretical proposals that view language development as an inherently social process (e.g., Kuhl, 2007) and may relate to behavioural findings showing that social information fosters language learning (e.g., Kuhl et al., 2003; Goldstein et al., 2003).

To conclude, we tested 6-month-old infants' neural response to auditory, visual and audiovisual speech stimuli using fNIRS. Our results show that infants recruit areas in frontal and prefrontal cortex during speech perception of unimodal and multimodal speech. In combination with previous findings, we take our data to indicate that frontal and prefrontal cortex play a critical role in language learning and processing. In particular, we suggest that inferior frontal cortex is involved in the learning and processing of multimodal speech information and that prefrontal cortex serves to evaluate the significance of the speech input.

Acknowledgments

We thank Caterina Böttcher for her help with data collection and coding. We also thank all families who participanted in this study. This work was supported by funding awarded by the Max Planck Society (to T.G.).

References

Altvater-Mackensen, N., Grossmann, T., 2016. The role of left inferior frontal cortex during audiovisual speech perception in infants. NeuroImage 133, 14–20.
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. 52, 191–204.
Baum, S., Martin, R., Hamilton, C., Beauchamp, M., 2012. Multisensory speech perception withouth the left superior temporal sulcus. NeuroImage 62, 1825–1832.
Beauchamp, M., Argall, B., Bodurka, J., Duyn, J., Martin, A., 2004. Unraveling multisensory integration: patchy organization within human STS multisensory cortex. Nat. Neurosci. 7, 1190–1192.
Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B 57, 289–300.
Bortfeld, H., Wruck, E., Boas, D.A., 2007. Assessing infants' cortical response to speech using near-infrared spectroscopy. NeuroImage 34, 407–415.
Bortfeld, H., Fava, E., Boas, D.A., 2009. Identifying cortical lateralization of speech processing in infants using near-infrared spectroscopy. Dev. Neuropsychol. 34, 52–65.
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.
Bruderer, A., Danielson, D., Kandhadai, P., Werker, J., 2015. Sensorimotor influences on speech perception in infancy. Proc. Natl. Acad. Sci. 112, 13531–13536.
Callan, D.E., Callan, A.M., Kroos, C., Vatikiotis-Bateson, E., 2001. Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. Cogn. Brain Res. 10, 349–353.
Calvert, G.A., et al., 1997. Activation of auditory cortex during silent lip reading. Science 276, 593–596.
Calvert, G.A., 2001. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb. Cortex 11, 1110–1123.
Calvert, G.A., Brammer, M., Bullmore, E., Campbell, R., Iversen, S.D., David, A., 1999. Response amplification in sensory-specific cortices during crossmodal binding. NeuroReport 10, 2619–2623.
Campbell, R., et al., 2001. Cortical substrates for the perception of face actions: an fMRI study of the specificity of activation for seen speech and for meaningless lower-face acts (gurning). Cogn. Brain Res. 12, 233–243.
Cappa, S., Pulvermüller, F., 2012. Language and the motor system. Cortex 48, 785–787.
Csibra, G., Gergeley, G., 2009. Natural pedagogy. Trends Cognit. Sci. 13, 148–153.
Danielson, D., Bruderer, A., Khandhadai, P., Vatikiotis-Bateson, E., Werker, J., 2017. The organization and reorganization of audiovisual speech perception in the first year of life. Cogn. Dev. 42, 37–48.
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., Merlaux, 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., Allirol, 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.
Desjardins, R., Rogers, J., Werker, J., 1997. An exploration of why pre-schoolers perform differently than do adults in audiovisual speech perception tasks. J. Exp. Child Psychol. 66, 85–110.
Desjardins, R., Werker, J., 2004. Is the integration of heard and seen speech mandatory for infants? Dev. Psychobiol. 45, 187–203.
Dick, A., Solodkin, A., Small, S., 2010. Neural development of networks for audiovisual speech comprehension. Brain Lang. 114, 101–114.
Eimas, P., Siqueland, E., Jusczyk, P., Vigorito, J., 1971. Speech perception in infants. Science 209, 1140–1141.
Eisner, F., McGettigan, C., Faulkner, A., Rosen, S., Scott, S., 2010. Inferior frontal gyrus activation predicts individual differences in perceptual learning of cochlear-implant simulations. J. Neurosci. 30, 7179–7186.
Fava, E., Hull, R., Bortfeld, H., 2014. Dissociating cortical activity during processing of native and non-native audiovisual speech from early to late infancy. Brain Sci. 4, 471–487.
Friederici, A., 2002. Towards a neural basis of auditory sentence processing. Trends Cognit. Sci. 6, 78–84.
Friederici, A., 2011. The brain basis of language processing: from structure to function. Physiol. Rev. Suppl. 91, 1357–1392.
Galantucci, B., Fowler, C.A., Turvey, M.T., 2006. The motor theory of speech perception reviewed. Psychon. Bull. Rev. 13, 361–377.
Goldstein, M.H., King, A.P., West, M.J., 2003. Social interaction shapes babbling: testing parallels between birdsong and speech. Proc. Natl. Acad. Sci. 100, 8030–8035.
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. 144, 1266–1287.
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.
Grossmann, T., Oberecker, R., Koch, S.P., Friederici, A.D., 2010. The developmental origins of voice processing in the human brain. Neuron 65, 852–858.
Guellai, B., Streri, A., Young, H.H., 2014. The development of sensomotor influences in the audiovisual speech domain: some critical questions. Front. Psychol. 5, 1–7.
Hagoort, P., 2014. Nodes and networks in the neural architecture for language: Broca's region and beyond. Curr. Opin. Neurobiol. 28, 136–141.
Hall, D., Fussell, C., Summerfield, A., 2005. Reading fluent speech from talking faces: typical brain networks and individual differences. J. Cogn. Neurosci. 17, 939–953.
Hickok, G., Houde, J., Rong, F., 2011. Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69, 407–422.
Hickok, G., Poeppel, D., 2007. The cortical organization of speech. Nat. Rev. Neurosci. 8, 393–402.
Hocking, J., Price, C., 2008. The role of the posterior superior temporal sulcus in audiovisual processing. Cereb. Cortex 18, 2439–2449.
Hyde, D., Jones, B., Porter, C., Flom, R., 2010. Visual stimulation enhances auditory processing in 3-month-old infants and adults. Dev. Psychobiol. 52, 181–189.
Hyde, D., Jones, B., Flom, R., Porter, C., 2011. Neural signatures of face-voice synchrony in 5-month-old human infants. Dev. Psychobiol. 53, 359–370.
Imada, T., Zhang, Y., Cheour, M., Taulu, S., Ahonen, A., Kuhl, P.K., 2006. Infant speech perception activates Broca's area: a developmental magnetoencehalography study. NeuroReport 17, 957–962.
Johnson, M.H., Dziurawiec, S., Ellis, H., Morton, J., 1991. Newborns preferential tracking of face-like stimuli and its subsequent decline. Cognition 40, 1–19.
Jusczyk, P., 1998. Constraining the search for structure in the input. Lingua 106, 197–218.
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.
Kubicek, C., Boisferon, A., Dupierrix, E., Pascalis, O., Loevenbruck, H., Gervain, J., Schwarzer, G., 2014. Cross-modal matching of audio-visual German and French fluent speech in infancy. PLoS One 9 e89275.
Kuhl, P.K., 2007. Is speech learning' gated' by the social brain? Dev. Sci. 10, 110–120.
Kuhl, P., Meltzoff, A., 1982. The bimodal perception of speech in infancy. Science 218, 1138–1141.
Kuhl, P.K., Tsao, F., Liu, H., 2003. Foreign-language experience in infancy: effects of short-term exposure and social interaction on phonetic learning. Proc. Natl. Acad. Sci. 100, 9096–9101.
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., Bailleux, 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., 2010. Infant perception of audio-visual speech synchrony. Dev. Psychol. 46, 66–77.
Lewkowicz, D., Ghazanfar, A., 2006. The decline of cross-species intersensory perception in human infants. Proc. Natl. Acad. Sci. 103, 6771–6774.
Lewkowicz, D., Hansen-Tift, A., 2012. Infants deploy selective attention to the mouth of a talking face when learning speech. Proc. Natl. Acad. Sci. 109, 1431–1436.
Liberman, A., 1957. Some results of research on speech perception. J. Acoust. Soc. Am. 29, 117–123.
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.
McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748.
Mampe, B., Friederici, A., Christophe, A., Wermke, K., 2009. Newborns' cry melody is shaped by their native language. Curr. Biol. 19, 1994–1997.
Majorano, M., Vihman, M., DePaolis, R., 2014. The relationship between infants' production experience and their processing of speech. Lang. Learn. Dev. 10, 179–204.
Mani, N., Schneider, S., 2013. Speaker identity supports phonetic category learning. J. Exp. Psychol. Hum. Percept. Perform. 39, 623–629.
Maurer, D., Werker, J.F., 2014. Perceptual narrowing during infancy: a comparison of language and faces. Dev. Psychobiol. 56, 154–178.
McGettigan, C., Faulkner, A., Altarelli, I., Obleser, J., Baverstock, H., Scott, S., 2012. Speech comprehension aided by multiple modalities: behavioural and neural interactions. Neuropsychologia 50, 762–776.
Meek, J., 2002. Basic principles of optical imaging and application to the study of infant development. Dev. Sci. 5, 371–380.
Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, G., Bertoncini, J., Amiel-Tison, C., 1988. A precursor of language acquisition in young infants. Cognition 29, 143–178.
Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202.
Minagawa-Kawai, Y., Cristia, A., Dupoux, E., 2011. Cerebral lateralization and early speech acquisition: a developmental scenario. Dev. Cogn. Neurosci. 1, 217–232.
Moon, C., Cooper, R., Fifer, W., 1993. Two-day-olds prefer their native language. Infant Behav. Dev. 16, 495–500.
Mugitani, R., Kobayashi, T., Hiraki, K., 2008. Audiovisual matching of lips and non-canonical sounds in 8-month-old infants. Infant Behav. Dev. 31, 307–310.
Nath, A., Beauchamp, M., 2012. A neural basis for interindividual differences in the McGurk effect, a multisensory speech illusion. NeuroImage 59, 781–787.
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.
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.
Patterson, M.L., Werker, J.F., 2003. Two-month-old infants match phonetic information in lips and voices. Dev. Sci. 6, 191–196.
Paulescu, E., et al., 2003. A functional-anatomical model for lipreading. Neurophysiology 90, 2005–2013.
Peelle, J., 2012. The hemispheric lateralization of speech processing depends on what "speech" is: a hierarchical perspective. Front. Hum. Neurosci. 6, 309.
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., et al., 2011. Neural language networks at birth. Proc. Natl. Acad. Sci. 108, 16056–16061.
Poeppel, D., Monahan, P.J., 2011. Feedforward and feedback in speech perception: revisiting analysis by synthesis. Lang. Cogn. Process. 26, 935–951.
Pons, F., Lewkowicz, D., Soto-Faroco, S., Sebastian-Galles, N., 2009. Narrowing intersensory speech perception in infancy. Proc. Natl. Acad. Sci. 106, 10598–10602.
Pulvermüller, F., Fadiga, L., 2010. Active perception: sensorimotor circuits as a cortical basis for language. Nat. Rev. Neurosci. 11, 351–360.
Reynolds, G., Bahrick, L., Lickliter, R., Guy, M., 2014. Neural correlates of intersensory processing in 5-month-old infants. Dev. Psychobiol. 56, 355–372.
Rossi, S., Telkemeyer, S., Wartenburger, I., Obrig, H., 2012. Shedding light on words and sentences: near-infrared spectroscopy in language research. Brain Lang. 121, 152–163.
Sams, M., Aulanko, R., Hamalainen, 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. Neurosci. Lett. 127, 141–145.
Schwartz, J.-L., Berthommier, F., Savariaux, C., 2004. Seeing to hear better: evidence for early audio-visual interactions in speech identification. Cognition 93, B69–B78.
Scott, S., McGettigan, C., Eisner, F., 2009. A little more conversation, a little less action – candidate roles for the motor cortex in speech perception. Nat. Rev. Neurosci. 10, 295–302.
Shaw, K., Baart, M., Depowski, N., Bortfeld, H., 2015. Infants' preference for native audiovisual speech dissociated from congruency preference. PLoS One 10 e0126059.
Shaw, K., Bortfeld, H., 2015. Sources of confusion in infant audiovisual speech perception research. Front. Psychol. 6, 1814.
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.
Song, J., Lee, H., Kang, H., Lee, D., Chang, S., Oh, S., 2015. Effects of congruent and incongruent visual cues on speech perception and brain activity in cochlear implant users. Brain Struct. Funct. 220, 1109–1125.
Taga, G., Asakawa, K., 2007. Selectivity and localization of cortical response to auditory and visual stimulation in awake infants aged 2 to 4 months. NeuroImage 36, 1246–1252.
Taga, G., Asakaw, K., Hirasawa, K., Konishi, Y., Koizumi, H., 2003. Brain imaging in awake infants by near-infrared optical topography. Proc. Natl. Acad. Sci. 100, 10722–10727.
Teinonen, T., Aslin, R., Alku, P., Csibra, G., 2008. Visual speech contributes to phonetic learning in 6-month-olds. Cognition 108, 850–855.
Tenenbaum, E., Sha, R., Sobel, D., Malle, B., Morgan, J., 2013. Increased focus on the mouth among infants in the first year of life: a longitudinal eye-tracking study. Infancy 18, 534–553.
Ter Schure, S., Junge, C., Boersma, P., 2016. Discriminating non-native vowels on the basis of multimodal, auditory or visual information: effects on infants' looking patterns and discrimination. Front. Psychol. 7, 525.
Venezia, J., Fillmore, P., Matchin, W., Isenberg, A., Hickok, G., Fridriksson, J., 2016. Perception drives production across sensory modalities: a network for sensorimotor integration of visual speech. NeuroImage 126, 196–207.
Vihman, M.M., 1996. Phonological Development. The Origins of Language in the Child. Blackwell, Oxford.
Vouloumanos, A., Hauser, M.D., Werker, J.F., Martin, A., 2010. The tuning of human neonates' preference for speech. Child Dev. 81, 517–527.
Watanabe, H., Homae, F., Nakano, T., Tsuzuki, D., Enkthur, L., Nemoto, K., Dan, I., Taga, G., 2013. Effect of auditory input on activation in infant diverse cortical regions during auidovisual processing. Hum. Brain Mapp. 34, 543–565.
Weikum, W.M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N., Werker, J.F., 2007. Visual language discrimination in infancy. Science 316, 1159.
Werker, J., Tees, R., 1984. Developmental changes across childhood in the perception of nonnative speech sounds. Can. J. Psychol. 37, 278–286.
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., Iacoboni, M., 2004. Listening to speech activates motor areas involved in speech production. Nat. Neurosci. 7, 701–702.
Yeung, H., Werker, J., 2013. Lip movements affect infants' audiovisual speech perception. Psychol. Sci. 24, 603–612.