

# Reading-Related Brain Changes in Audiovisual Processing: Cross-Sectional and Longitudinal MEG Evidence

 Sendy Caffarra,<sup>1,2,3</sup>  Mikel Lizarazu,<sup>3</sup> Nicola Molinaro,<sup>3,4</sup> and  Manuel Carreiras<sup>3,4,5</sup>

<sup>1</sup>Division of Developmental-Behavioral Pediatrics, Stanford University School of Medicine, Stanford, California 94305-5101, <sup>2</sup>Stanford University Graduate School of Education, Stanford, California 94305, <sup>3</sup>Basque Center on Cognition, Brain and Language, 20009 San Sebastian, Spain, <sup>4</sup>Ikerbasque Basque Foundation for Science, 48009 Bilbao, Spain, and <sup>5</sup>University of the Basque Country (UPV/EHU), 48940 Bilbao, Spain

The ability to establish associations between visual objects and speech sounds is essential for human reading. Understanding the neural adjustments required for acquisition of these arbitrary audiovisual associations can shed light on fundamental reading mechanisms and help reveal how literacy builds on pre-existing brain circuits. To address these questions, the present longitudinal and cross-sectional MEG studies characterize the temporal and spatial neural correlates of audiovisual syllable congruency in children (age range, 4–9 years; 22 males and 20 females) learning to read. Both studies showed that during the first years of reading instruction children gradually set up audiovisual correspondences between letters and speech sounds, which can be detected within the first 400 ms of a bimodal presentation and recruit the superior portions of the left temporal cortex. These findings suggest that children progressively change the way they treat audiovisual syllables as a function of their reading experience. This reading-specific brain plasticity implies (partial) recruitment of pre-existing brain circuits for audiovisual analysis.

**Key words:** audiovisual congruency effect; MEG; reading acquisition

## Significance Statement

Linking visual and auditory linguistic representations is the basis for the development of efficient reading, while dysfunctional audiovisual letter processing predicts future reading disorders. Our developmental MEG project included a longitudinal and a cross-sectional study; both studies showed that children's audiovisual brain circuits progressively change as a function of reading experience. They also revealed an exceptional degree of neuroplasticity in audiovisual neural networks, showing that as children develop literacy, the brain progressively adapts so as to better detect new correspondences between letters and speech sounds.

## Introduction

Literacy is a relatively recent cognitive achievement in human evolution for which there are no specialized neural circuits already in place. Learning this life-changing skill thus requires considerable modulation of pre-existing brain networks, such as the visual object recognition and spoken language networks (Carreiras et al., 2009; Dehaene et al., 2015). A considerable amount of research on reading-related brain

changes has examined this plasticity in either visual and auditory brain circuits (Goswami and Ziegler, 2006; Ziegler and Muneaux, 2007; Dehaene et al., 2010, 2015). However, the core of reading acquisition lies in the interaction between these two modalities. Efficient reading skills crucially depend on the ability to compare and connect visual and auditory representations of letters (Blomert, 2011). The present MEG study focused on these audiovisual processes, testing how they changed as a function of developing reading abilities. We hypothesized that during reading acquisition pre-existing brain circuits for audiovisual processing should become progressively tuned to the arbitrary relationships between letters and speech sounds (Blomert, 2011).

The processing of natural audiovisual associations (e.g., the correspondence between speech and lip movements) has been widely explored in the literature. The effects of audiovisual integration (i.e., the absolute difference between bimodal and unimodal presentations) and audiovisual congruency (i.e., the absolute difference between matching and mismatching bimodal presentations) have mainly been localized in the auditory cortex and the

Received Nov. 30, 2020; revised May 10, 2021; accepted May 16, 2021.

Author contributions: S.C., N.M., and M.C. designed research; S.C. performed research; S.C. and M.L. analyzed data; S.C. wrote the paper.

This project received funding from the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie Grant Agreement No. 837228 (H2020-MSCA-IF-2018-837228-ENGRAVING). The project was also funded by the Spanish Ministry of Economy, Industry and Competitiveness (Grant PSI2017-82941-P), the Basque Government through the BERC 2018-2021 Program, and the Agencia Estatal de Investigación through BCBL (Basque Center on Cognition, Brain and Language) Severo Ochoa excellence accreditation SEV-2015-0490.

The authors declare no competing financial interests.

Correspondence should be addressed to Sendy Caffarra at [caffarra@stanford.edu](mailto:caffarra@stanford.edu).

<https://doi.org/10.1523/JNEUROSCI.3021-20.2021>

Copyright © 2021 the authors

**Table 1. Behavioral description of participants in the cross-sectional and the longitudinal studies**

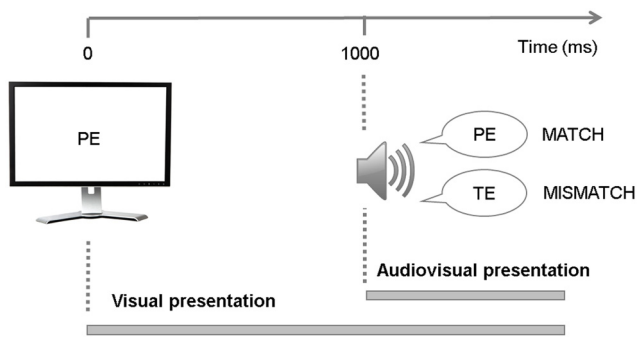
	Cross-sectional study, mean (SD)		Longitudinal study, mean (SD)	
	Prereaders	Readers	Session 1	Session 2
<i>N</i>	20	22	15	15
Females, <i>n</i>	9	11	7	7
Right handed, <i>n</i>	12*	21	8*	12
Age, years	4.8 (0.8)	7.7 (0.7)	4.5 (0.7)	7.1 (0.3)
Formal reading instruction duration, months	0	14.2 (4.7)	0	8.1 (1.7)
Nonverbal IQ (0–48)	14.1 (5.3)	25.6 (4.2)	12.3 (4.6)	24.8 (3.0)
	5–24	15–34	5–22	20–30
Vocabulary size (0–45)	20.4 (5.1)	32.9 (4.2)	18.9 (4.4)	30.3 (4.2)
	12–31	23–38	12–28	23–38
Letter knowledge (0–27)	13.7 (10.0)	27 (0.0)	10.8 (8.6)	27 (0.0)
	0–27	27–27	0–27	27–27
Syllables reading times, s		21.6 (6.9)		26.7 (8.7)
		14–43		17–46
Word reading errors (0–80)		2.4 (2.4)		4.6 (4.0)
		0–9		0–15
Word reading times, s		85.1 (42.2)		133.7 (54.7)
		41–248		62–245
Pseudoword reading errors (0–80)		3.3 (1.9)		6.6 (5.5)
		1–8		0–20
Pseudoword reading times, s		90.5 (42.6)		125.9 (49.6)
		42–236		61–245

For each measure, we included the mean (SD) and range of raw scores. The nonverbal intelligence quotient (IQ) was assessed using Raven's Progressive Matrices test, and vocabulary size was measured using a picture-naming task (both taken from the Kaufman Brief Intelligence Test (K-BIT); Spanish version, 2009). Reading performance was assessed by measuring the speed and number of errors on 30 Basque syllables (the MEG experimental stimuli), 80 Basque high-frequency words, and 80 Basque pseudowords.

\*The rest of the children did not show a clear hand preference at this time point.

superior temporal cortex (Amedi et al., 2005; Hocking and Price, 2008), with possible left lateralization (Calvert et al., 1998; Calvert, 2001). Research on fluent adult readers has shown that these brain areas seem to be (at least partially) recruited even in processing arbitrary associations between letters and speech sounds, indicating a certain degree of plasticity in audiovisual brain areas during reading acquisition (Amedi et al., 2005; Hocking and Price, 2008; Blomert and Froyen, 2010). Neuroimaging studies comparing matching and mismatching letter–sound pairs reported effects in the superior temporal and auditory cortex (van Atteveldt et al., 2004, 2007; Blau et al., 2008, 2010; Karipidis et al., 2017, 2018), which were often left lateralized and appeared within the first 500 ms of stimulus presentation (Karipidis et al., 2017, 2018; Raji et al., 2000; Xu et al., 2019, 2020; for even earlier effects, see Herdman et al., 2006). Importantly, cross-sectional designs have revealed a relation between these audiovisual effects and reading skills (Blau et al., 2010; Jost et al., 2014; Karipidis et al., 2017,

2018), indicating that cross-modal brain responses are affected by literacy experience. Studies on normal reading acquisition in children seem to suggest that automatic effects of audiovisual letter processing are rare in beginning readers (Xu et al., 2018) and may emerge only after a few years of formal reading instruction under facilitated experimental conditions (e.g., nonsimultaneous bimodal presentations; Froyen et al., 2009). However, the scarce research on these plastic brain changes during development has so far been documented only by means of between-group comparisons. Longitudinal designs overcome the potential limitations—related to the difficulty of establishing perfectly matching groups—in between-group designs. The present MEG study is the first to adopt a longitudinal (alongside a cross-sectional) design to characterize the progressive emergence of audiovisual congruency effects as children learn to read. Matching and mismatching audiovisual syllables were presented to children. We predicted that the audiovisual congruency effect should be localized in the left superior



**Figure 1.** Schematic representation of an experimental trial.

temporal cortex and left auditory cortex and emerge within 500 ms after stimulus onset. We expected this effect to be reading specific and, thus, to correlate with children's reading scores.

## Materials and Methods

### Participants

Forty-two Basque-Spanish early bilingual children participated in the cross-sectional study (20 females; mean age, 6.3 years; SD, 1.7; age range, 4–9 years). Data from five additional participants were excluded because of poor data quality ( $n = 4$ ) or the presence of a hearing disorder ( $n = 1$ ). Participants were divided in two groups (prereaders and readers) based on whether they had already received formal reading instruction (Table 1). Fifteen children from the prereaders group also participated in the longitudinal study, returning for a second MEG recording session. The mean time between session 1 and session 2 was 32 months (SD, 5; age range, 4–8 years; Table 1).

All participants were learning to read in Basque. Basque has a transparent orthography, such that the consistent correspondences between letter and speech sounds are usually mastered within 1 year of reading instruction. Readers' school attendance was regular, and none of them were repeating or had skipped a grade. All participants had normal or corrected-to-normal vision and normal hearing. Their parents reported no neurologic disorders and did not suspect developmental reading problems. The BCBL (Basque Center on Cognition, Brain and Language) ethical committee approved the experiment (following the principles of the Declaration of Helsinki) and all parents (or the tutors) of the children compiled and signed the written informed consent.

### Materials and procedure

Thirty consonant–vowel syllables were created using one of six consonants (f, k, l, m, p, t) followed by one of five vowels (a, e, i, o, u) from the Basque alphabet. We used syllables rather than single letters to make the stimuli more ecological. Basque children learn to name Basque letters using syllables and the consonant–vowel syllable structure is highly common in the Basque lexicon. We did not expect this choice to affect our results as audiovisual congruency effects have been reported for a wide range of linguistic (e.g., letters, words, ideograms; Amedi et al., 2005; Hocking and Price, 2008; Xu et al., 2019) and nonlinguistic (pictures; Hocking and Price, 2008) stimuli. The syllables were presented four times both in the visual and the auditory modality to create 120 cross-modal pairs. Spoken syllables were recorded by a female voice at 44.1 kHz. The audiovisual correspondence of cross-modal pairs was manipulated to produce 60 matching and 60 mismatching pairs. The mismatching pairs were pseudorandomly selected so that they always differed in the initial consonant while sharing the final vowel. Sixteen cross-modal syllable pairs were added for a target detection task. They contained the image of a cat in between the letters in the visual presentation and/or the sound of a cat meowing in between the letter sounds in the auditory presentation.

During the experimental trial, the visual stimulus (written syllable) was first presented at the center of the screen. After a 1 s, the auditory stimulus (spoken syllable) was also presented, while the written syllable remained displayed on the screen. The visual and auditory stimuli offsets

coincided and the interstimulus interval was 1000 ms (Fig. 1). The onsets of the visual and auditory syllable presentations were shifted to create a facilitated experimental situation where it was more likely to observe early audiovisual congruency effects (Froyen et al., 2009). Moreover, this temporal sequence better reflected children's everyday experience, such as listening to stories read aloud, where they hear language after seeing it in print. Auditory stimuli were presented between 70 and 80 dB through plastic tubes and silicon earpieces (mean duration, 700 ms; SD, 95). The task consisted of pressing a button whenever the current stimulus corresponded to a cat either in the visual or in the auditory modality. Stimuli were randomized across participants. The recording session lasted ~10 min.

### MEG data recording and preprocessing

MEG data were recorded in a magnetically shielded room (Maxshield<sup>†</sup>, Elekta) using an Elekta Neuromag MEG device (including 102 sensors with two planar gradiometers and one magnetometer each). MEG recordings were acquired continuously with children in a sitting position, with a bandpass filter at 0.03–330 Hz and a sampling rate of 1 kHz. Head position inside the helmet was continuously monitored using five head position indicator coils. The location of each coil relative to the anatomic fiducials (nasion, and left and right preauricular points) was defined with a 3D digitizer (Fastrak, Polhemus). This procedure is critical for head movement compensation during the data recording session. In addition, ~200 head surface points were digitized and later used to spatially align the MEG sensors with an age-based pediatric T1 template (Fonov et al., 2011).

Eye movements were monitored with bipolar vertical electro-oculogram (VEOG) and horizontal electro-oculogram (HEOG). MEG data were individually corrected for head movements and subjected to noise reduction using MaxFilter (version 2.2.15; Neuromag, Elekta) and the temporally extended signal space separation method (Taulu and Kajola, 2005; Taulu and Hari, 2009). On average, 10 bad channels were automatically identified using Xscan (Neuromag, Elekta). Bad channels were substituted with interpolated values. There was no difference between the number of channels interpolated between readers (10.2; SD, 2.2) and prereaders (9.1; SD, 2.2;  $t < 1$ ), or between session 1 (10.1; SD, 2.3) and session 2 (10.2; SD, 3.3;  $t < 1$ ).

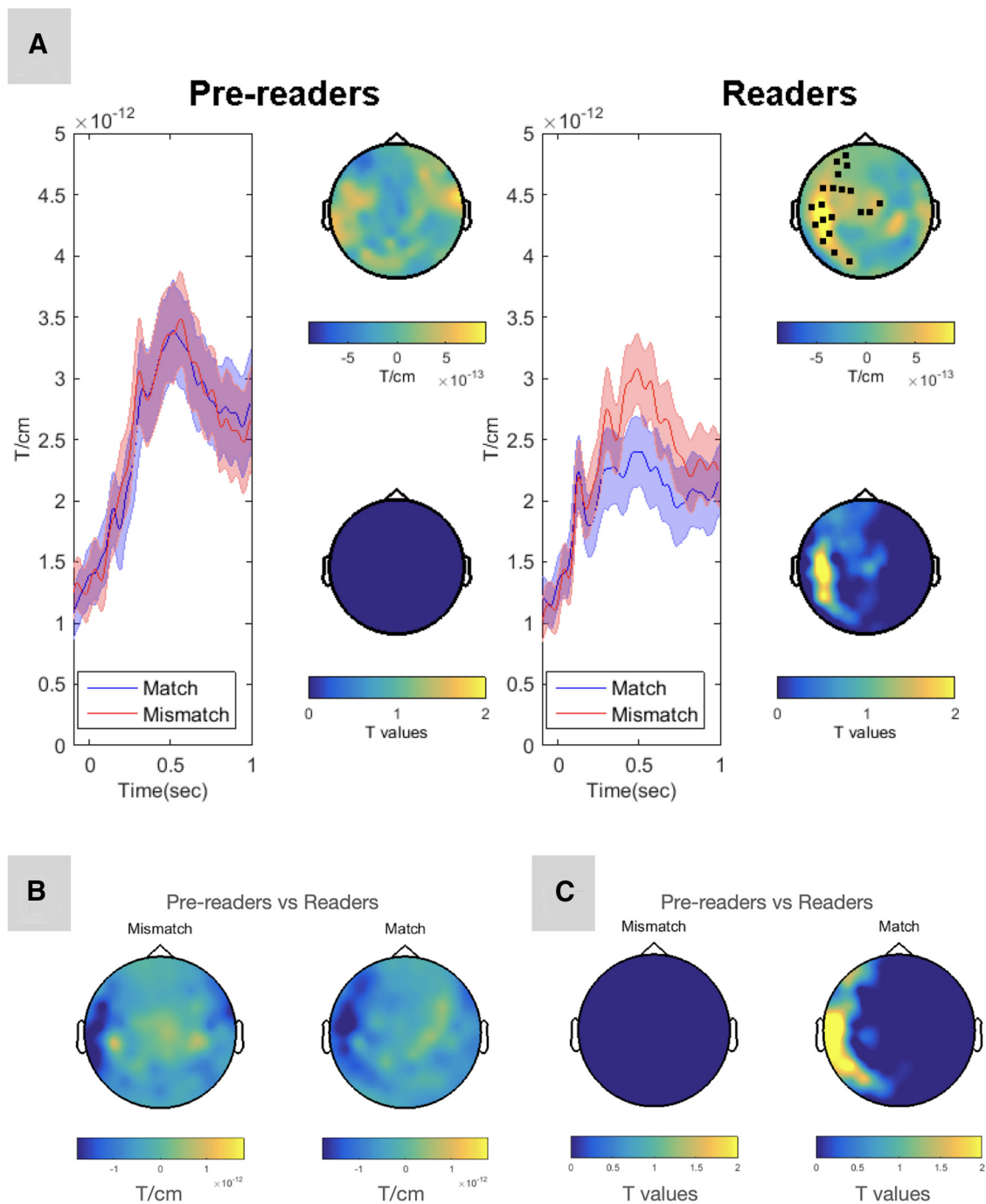
Subsequent analyses were performed using MATLAB R2014 (MathWorks) and the Fieldtrip toolbox (Oostenveld et al., 2011). MEG epochs of 2.5 s were obtained, including 1.5 s before and 1.0 s after the auditory presentation onset. High-frequency muscle artifacts (110–140 Hz) were automatically rejected: average  $z$ -values over sensors and time points in each trial were calculated, and trials exceeding the threshold of a  $z$  score equal to 30 were removed. To suppress eye movement artifacts, 70 independent components were identified by applying independent component analysis (Jung et al., 2000) to the MEG data. Independent components corresponding to ocular artifacts were identified and removed based on the correlation values between each component and the VEOG/HEOG channels (rejected components range, 0–2).

Finally, MEG epochs were visually inspected to discard any remaining artifacts. On average, 28.1% (SD, 13.1) of trials were rejected [cross-sectional study: 26.7% (SD, 11.6); longitudinal study: 30.0% (SD, 14.9)], with no significant difference between conditions ( $F$  values  $< 3$ ;  $p$  values  $> 0.05$ ) or groups ( $F$  values  $< 5$ ;  $p$  values  $> 0.05$ ).

### MEG experimental design and statistical analysis

**Sensor-level event-related fields.** The artifact-free MEG data were lowpass filtered at 35 Hz. Trials were grouped together for each condition and then averaged to obtain the event-related fields (ERFs). ERFs were quantified as the absolute amplitude of the 102 orthogonal planar gradiometer pairs by computing the square root of the sum of squares of the amplitudes of the two gradiometers in each pair. A baseline correction for the data preceding the stimulus by 500 ms was performed.

In both the cross-sectional and longitudinal studies, the ERFs for the match and mismatch conditions of prereaders and readers were statistically compared using a nonparametric cluster-based permutation test (Maris and Oostenveld, 2007). Specifically,  $t$  statistics were computed for each sensor (combined gradiometers) and time point during the 0–1000

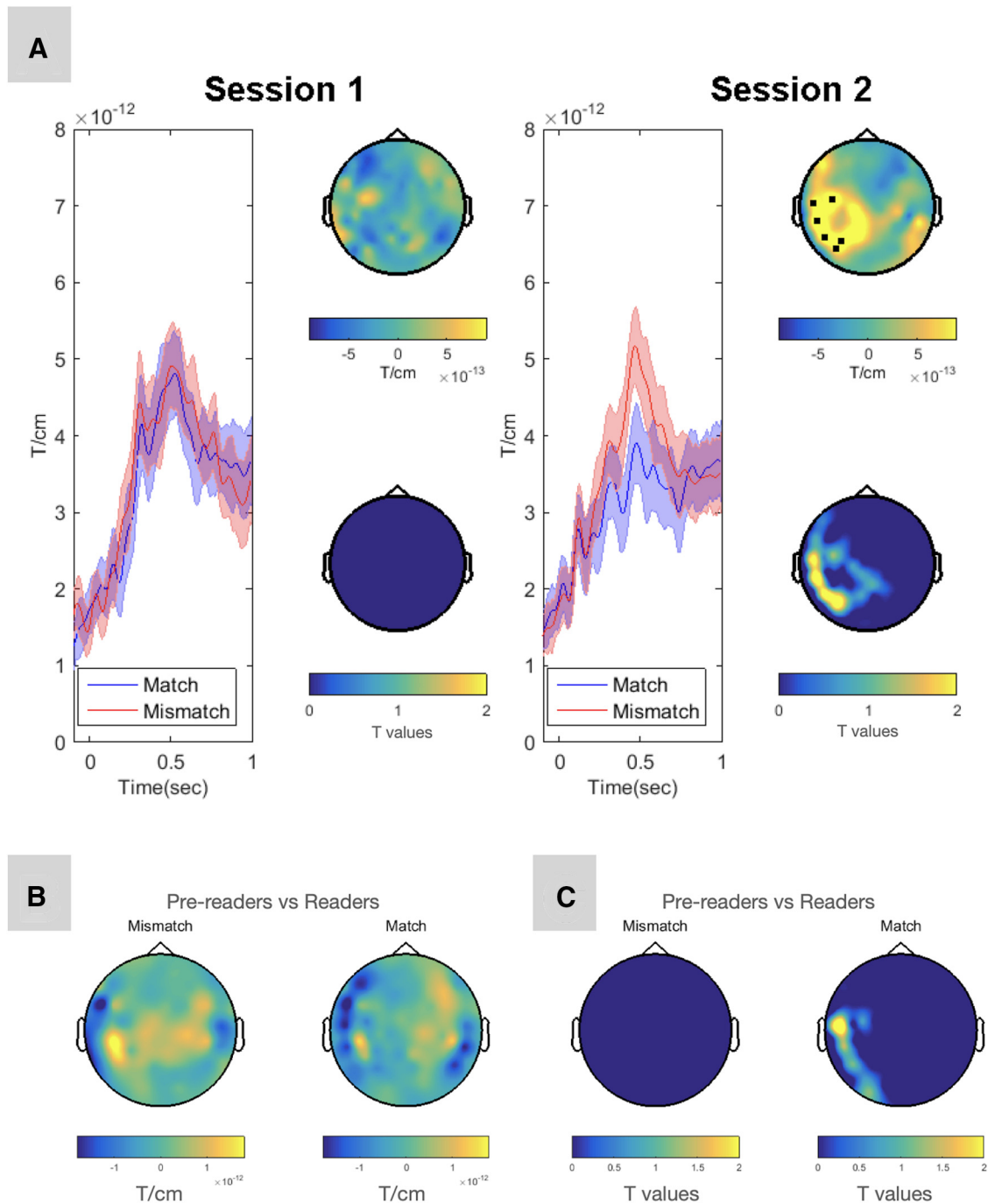


**Figure 2.** *A*, ERFs for the cross-sectional study. *A*, Grand average ERF responses to spoken syllables for the match (blue) and the mismatch (red) condition in prereaders (left) and readers (right). Shaded edges represent  $\pm 1$  SE. ERF waveform averages were calculated based on the group of left sensors displayed on the map in the top left corner. The top maps represent the topographic distribution of the audiovisual congruency effect (calculated by subtracting the match from the mismatch condition) within the time window when the effect reached its maximum. The topographic maps at the bottom show the spatial distribution of the statistically significant cluster in the same time window (yellow color scale indexes the magnitude of  $t$  values that passed the statistical threshold of 0.05). *B*, Topographic maps of the difference between readers and prereaders. *C*, Spatial distribution of the statistically significant cluster when comparing readers and prereaders (yellow color scale indexes significant  $t$  values magnitude).

ms time window, and a clustering algorithm formed groups of channels over time points based on these tests. The neighborhood definition was based on the template for combined gradiometers of the Neuromag-306 provided by the toolbox. In order for a data point to become part of a cluster, a threshold of  $p = 0.05$  was used (based on a two-tailed dependent  $t$  test, using probability correction). The sum of the  $t$  statistics in a sensor group was then used as a cluster-level statistic (e.g., the maxsum option in Fieldtrip), which was then tested with a randomization test using 1000 runs. Moreover, we used a two-tailed  $t$  test to perform a between-group comparison of the audiovisual congruency effects (ERF

differences between mismatch and match conditions) in the cross-sectional and the longitudinal study. Finally, partial correlations were calculated to evaluate the relationship between the magnitude of the audiovisual congruency effect and reading performance after correcting for age, vocabulary size, and nonverbal intelligence.

*Source-level ERFs.* Using MRILab (version 1.7.25; Neuromag, Elekta), the digitized points from the Fastrak digitizer (Polhemus) were coregistered to the skin surface obtained from an age-compatible T1 template (Fonov et al., 2011). The T1 template was segmented into scalp, skull, and brain components using the segmentation algorithms



**Figure 3.** *A*, ERFs for the longitudinal study. Grand average ERF responses to spoken syllables for the match (blue) and the mismatch (red) conditions in session 1 and session 2. Shaded edges represent  $\pm 1$  SE. ERF waveform averages were calculated based on the group of left sensors displayed on the map in the top left corner. The top maps represent the topographic distribution of the audiovisual congruency effect (calculated by subtracting the match from the mismatch condition) within the time window when the effect reached its maximum. The topographic maps at the bottom show the spatial distribution of the statistically significant cluster in the same time window (yellow color scale indexes the magnitude of  $t$  values that passed the statistical threshold of 0.05). *B*, Topographic maps of the difference between session 1 and session 2. *C*, Spatial distribution of the statistically significant cluster when comparing session 1 and session 2 (yellow color scale indexes significant  $t$  values magnitude).

implemented in Freesurfer (Dale et al., 1999). The source space was defined as a regular 3D grid with a 5 mm resolution, and the lead fields were performed using a realistic three-shell model. Both planar gradiometers and magnetometers were used for inverse modeling. Whole-brain source activity was estimated using the linearly constrained minimum variance (LCMV) beamformer approach (Van Veen et al., 1997). For each condition, LCMV beamformer was computed on the evoked data in the  $-400$  to  $0$  prestimulus and in the  $350$ – $750$  ms poststimulus time intervals. This poststimulus interval was chosen because it contained the audiovisual congruency effects at the sensor level. Statistical significance

was assessed by a paired  $t$  test (from SPM software) comparing mean amplitudes in the poststimulus and the prestimulus interval (SPM).

## Results

Participants were able to correctly identify the target stimuli (cross-sectional  $d'$ , 1.870; longitudinal  $d'$ , 2.077), with no differences across groups (cross-sectional:  $t_{(40)} = 1.114$ ,  $p = 0.272$ ; longitudinal:  $t_{(14)} = 1.872$ ,  $p = 0.082$ ).

For the cross-sectional study (Fig. 2A), cluster-based permutations on the ERF responses showed an audiovisual congruency effect ( $p = 0.001$ ; difference between mismatch and match condition) only for readers in a 350–790 ms time window following the auditory syllable onset over left temporal sensors (Fig. 2A). The magnitude of the audiovisual congruency effect differed between readers and prereaders ( $p = 0.005$ ). This difference was because of the suppressed amplitude of the match condition in readers compared with prereaders (match condition,  $p = 0.021$ ; mismatch condition,  $p = 0.105$ ; Fig. 2B and 2C).

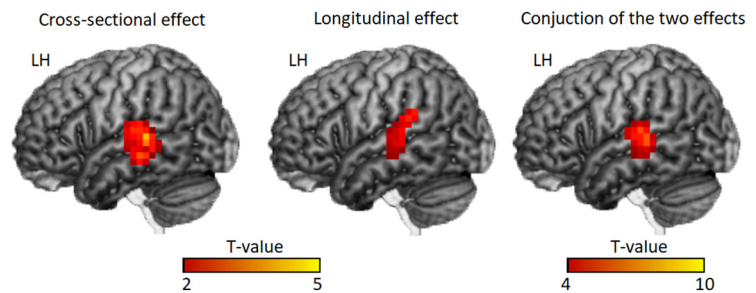
Similarly, for the longitudinal study (Fig. 3A), we observed an audiovisual congruency effect ( $p = 0.017$ ) in a 390–563 ms time window following the auditory syllable onset over left temporal sensors (Fig. 3A). The magnitude of the audiovisual congruency effect differed between sessions (session 1 vs session 2,  $p = 0.038$ ). Again, this difference was because of the suppressed amplitude of the match condition in the readers (session 2) compared with the prereaders (session 1: match condition,  $p = 0.021$ ; mismatch condition,  $p = 0.627$ ; Fig. 3B and 3C).

The ERF effects observed at the sensor level were source reconstructed in the 350–750 ms time window. In both the cross-sectional and longitudinal studies, the congruency effect ( $p < 0.05$ ) emerged in the posterior part of the left superior temporal cortex (Fig. 4).

The size of the audiovisual congruency effect negatively correlated with reading errors and reading speed measures after correcting for age, nonverbal intelligence, and vocabulary size (syllable reading times:  $r = -0.31$ ,  $p = 0.031$ ; number of errors per second while reading Basque words:  $r = -0.36$ ,  $p = 0.014$ ; number of errors per second while reading Basque pseudowords:  $r = -0.23$ ,  $p = 0.090$ ; Fig. 5).

## Discussion

The capacity to create strong associations between speech sounds and written representations is a key skill for reading. Audiovisual letter and audiovisual symbol processing predict future reading fluency (Horbach et al., 2015, 2018; Karipidis et al., 2018) and are often impaired in dyslexia (Fox, 1994; Vellutino et al., 2004; Froyen et al., 2011; Richlan, 2019). Understanding the developmental changes involved in letter-to-speech sound processing can shed light on the pivotal mechanisms of reading and can point to possible sources of reading disorders. With this aim, the present study investigated how audiovisual syllable analysis changed as a function of reading acquisition. The results showed a high degree of plasticity in neural responses to audiovisual syllable congruency, which was related to reading acquisition (as shown by partial correlations with reading performance). This neural adjustment was mainly localized in the left superior temporal cortex, which is in line with previous findings (Rajj et al., 2000; Blau et al., 2008, 2010; Karipidis et al., 2017, 2018; Xu et al., 2019, 2020). Importantly, this brain area is not exclusively involved in the processing of letter–speech sound correspondences, but is also sensitive to less arbitrary audiovisual associations available before reading acquisition (Calvert et al., 1998; Calvert, 2001; Amedi et al., 2005). This broad sensitivity is compatible with the idea that we do not have evolutionarily specialized circuits for reading, and literacy must build on pre-existing brain networks (Dehaene et al., 2010, 2015). In line with this hypothesis, previous findings have shown reading-related adjustment of naturally evolved brain mechanisms for visual and



**Figure 4.** Spatial localization of the audiovisual congruency effect for readers of the cross-sectional and the longitudinal study. The final plot shows the conjunction of the two effects. Paired  $t$  test comparing the mean source activity in the prestimulus and poststimulus intervals were calculated. Different color intensity indexes significant  $t$  values.

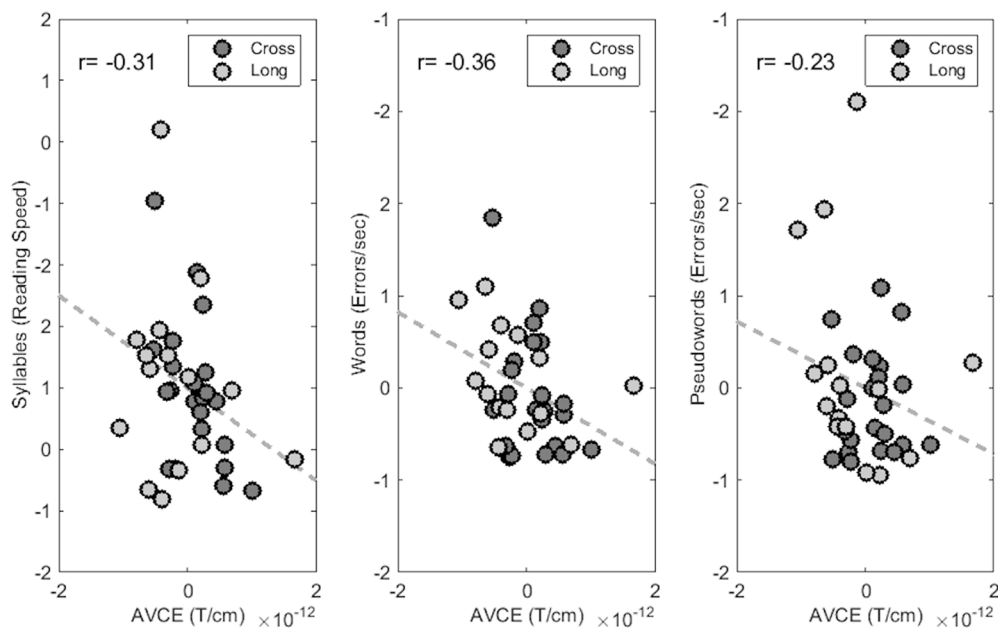
auditory processing (Goswami and Ziegler, 2006; Ziegler and Muneaux, 2007; Dehaene et al., 2010, 2015). The present findings extend this claim, suggesting that reading experience can also have an impact on naturally evolved brain mechanisms for audiovisual processing (Blomert, 2011).

The direction of the audiovisual congruency effect is also informative. Past research reveals considerable inconsistency: some studies have shown stronger responses for matching conditions; others report the opposite pattern (Table 2). Although it remains unclear what drives the direction of the effect (for some proposals, see Hocking and Price, 2008; Holloway et al., 2015; Plewko et al., 2018; Wang et al., 2020), we note that  $\sim 70\%$  of the studies reporting stronger matching responses are fMRI studies. The reverse pattern has been more frequently observed in electrophysiology and with experimental designs that include nonsimultaneous audiovisual presentations. This could indicate that temporal aspects of experimental design may affect the direction of the effect. The present MEG studies fully align with these trends found in the literature.

In both the longitudinal and cross-sectional study, we observed progressive suppression of the audiovisual matching response as a function of reading skills. Given that the congruency effect was found in auditory areas and the lack of modulation in the mismatch condition, it is unlikely that attention mechanisms accounted for this effect. This pattern is more likely the result of cross-modal integration since audiovisual correspondences can only be detected given successful interaction between two unimodal inputs. However, not all brain areas showing a congruency effect are necessarily the source of integrative operations (van Atteveldt et al., 2004, 2007; van Atteveldt and Ansari, 2014). Neuroimaging studies on adults comparing unimodal and bimodal letters proposed a finer functional distinction within subareas of the left superior temporal cortex. According to this view, the superior temporal sulcus is the neural hub for audiovisual convergence and integration, which sends feedback to superior auditory areas signaling letter–sound congruency (van Atteveldt et al., 2004, 2007). This functional distinction is further confirmed by cytoarchitectonic studies in human and nonhuman primates, which have shown a difference in the cellular structure of dorsolateral and ventromedial temporal regions (Ding et al., 2009; Insausti, 2013; Zachlod et al., 2020). The reduced response of the superior temporal cortex to matching audiovisual syllables might reflect the sharpening of neuronal tuning (i.e., responses to overlearned audiovisual associations are suppressed; Hurlbert, 2000); cross-modal repetition suppression (Henson, 2003); or neural adaptation (Grill-Spector and Malach, 2004).

The present MEG results also support the idea that written letters systematically modulate children's responses to speech

## Correlations between behavioral and brain measures



**Figure 5.** Correlation between the residuals of the audiovisual congruency effect (AVCE) and the residuals of reading scores (after correction for age, nonverbal intelligence, and vocabulary size). From left to right: syllable reading times, number of errors per second while reading Basque words, number of errors per second while reading Basque pseudowords. All readers are displayed in the scatterplots ( $n = 37$ ; dark gray, cross-sectional study; light gray, longitudinal study).

**Table 2.** Quick summary of the direction of audiovisual congruency effects previously reported in the literature

	Technique	Participants	SOA for AVCE (ms)	Location (time, ms)	Location (space)
<b>Match &gt; mismatch</b>					
van Atteveldt et al., 2004	fMRI	Adults	0		Bilateral PT/HS
van Atteveldt et al., 2007	fMRI	Adults	Only at 0		Bilateral PT/aSTP
Blau et al., 2008	fMRI	Children (~9 years)	0		Left STS
Blau et al., 2010	fMRI	Adults	0		Bilateral STS/left PT/HS
Plewko et al., 2018	fMRI	Children with risk for dyslexia (~7 years)	0		Left PT/STG
Karipidis et al., 2018	fMRI	Prereaders (~6 years, future normal readers)	0		Left PT
Wang et al., 2020	fMRI	Children (poor readers, ~8 years)	0		Right MFG
Raij et al., 2000	MEG	Adults	0	415–515	Bilateral STS
Herdman et al., 2006	MEG	Adults	0	0–250	Left PT/HS
Xu et al., 2020	MEG	Adults	0	500–800	Bilateral STC
<b>Mismatch &gt; match</b>					
Froyen et al., 2008	EEG	Adults	Maximum at 0	~160	
Froyen et al., 2009	EEG	Children (11 years)	Only at 200	~160	
Froyen et al., 2011	EEG	Dyslexic children (11 years)	Only at 200	~690	
Karipidis et al., 2017	EEG/fMRI	Prereaders (~6 years)	0	382–442/644–704	Right ITG
Karipidis et al., 2018	EEG	Prereaders (~6 years, future normal readers)	0	382–442	
Xu et al., 2019	MEG	Adults	0	490–890	Left STC/MTC
Hocking and Price, 2008	fMRI	Adults	Maximum at 0		Left STS
Holloway et al., 2015	fMRI	Adults	0		Bilateral frontoparietal network
Plewko et al., 2018	fMRI	Children with no risk for dyslexia (7 years)	0		Left PT/STG
Wang et al., 2020	fMRI	Children (~8 years, poor and normal readers)	0		Right MTG/Left IFG/ITG/bilateral SFG

A summary of the timing and the MRI coordinates of the audiovisual congruency effects is reported in the last two columns (whenever available). AVCE, Audiovisual congruency effects; SOA, stimulus-onset asynchrony; aSTP, anterior superior temporal plane; HS, Heschl's sulcus; IFG, inferior frontal gyrus; ITG, inferior temporal gyrus; MFG, middle frontal gyrus; MTC, middle temporal cortex; PT, planum temporale; STC, superior temporal cortex; STS, superior temporal sulcus.

sounds in the left superior temporal cortex (Herdman et al., 2006; van Atteveldt et al., 2007; Froyen et al., 2008, 2009). Our longitudinal findings suggest that this effect is already present after a few months of formal reading instruction. A longer training period might be needed to reach a high degree of automaticity (and a shorter time window for audiovisual integration; Laasonen et al., 2000, 2002; Froyen et al., 2009). In the present study, the long stimulus-onset asynchrony (SOA) between the visual and auditory onsets, together with the relatively late latency of our audiovisual

congruency effect, point to a low degree of automaticity. This is in line with a slow developmental trajectory for automatic letter–speech integration that extends beyond the first years of reading instruction (Froyen et al., 2009).

While the superior temporal cortex became progressively more sensitive to audiovisual letter congruency, other reading-related brain areas, such as the visual word form area (VWFA), did not show similar tuning. The shifted time onset between the visual and the auditory presentation might have reduced chances

to observe an audiovisual congruency effect in ventral occipito-temporal areas. It is possible that, after early activation during the visual presentation, there was no additional VWFA recruitment with spoken syllables. More research on simultaneous and nonsimultaneous audiovisual presentations is needed to clarify this point. The lack of occipitotemporal effects might also relate to levels of reading automaticity, with the VWFA becoming more responsive to auditory/audiovisual stimuli as reading automaticity increases (Yoncheva et al., 2010; Monzalvo and Dehaene-Lambertz, 2013). The present findings suggest that at low levels of automaticity the left superior temporal cortex plays a crucial role in establishing cross-modal correspondences between letters and speech sounds. The VWFA does not seem to be as crucial at this stage but might become more relevant after several years of reading instruction (Froyen et al., 2009). These findings are in line with the idea that entrenched audiovisual brain networks represent an essential prerequisite for reading development that precedes the functional tuning of the VWFA (Blomert, 2011).

Previous research has reported a lack of occipitotemporal response during audiovisual processing (van Atteveldt et al., 2004; Karipidis et al., 2018), leading to the general claim that audiovisual congruency effects are more often observed in auditory than visual areas (van Atteveldt et al., 2004; Blomert and Froyen, 2010). However, such effects differ from those associated with the neural network for audiovisual speech, which requires a stronger involvement of visual areas (Calvert et al., 1998; Calvert, 2001). The source of this discrepancy might be related to the different nature of the audiovisual associations in question. While in audiovisual speech the visual component (i.e., lip movements) occurs simultaneously with speech input across the life span, the associations between letters and sounds are arbitrary and do not always occur simultaneously. Thus, although there is partial recycling of brain areas naturally evolved for audiovisual analysis, letter–sound associations maintain a certain degree of specificity (Blomert and Froyen, 2010).

We also found no effects in parietal areas, such as the supramarginal and angular gyri, which are generally thought to be involved in access to phonological representations of text (Pugh et al., 2000; Booth et al., 2004; Schlaggar and McCandliss, 2007). This might be because of differences in experimental design: audiovisual effects in parietal areas are more often observed in comparisons of unimodal and bimodal linguistic stimuli than in comparisons of matching and mismatching audiovisual conditions (Xu et al., 2018, 2019, 2020). These parietal areas may be more involved in audiovisual letter integration than in subsequent feedback to sensory brain areas.

Finally, although our participants were early bilinguals, the present results are compatible with those reported in monolinguals (Herdman et al., 2006; Hocking and Price, 2008; Karipidis et al., 2017, 2018). In addition, both writing systems learned by the children in this study (Spanish and Basque) were highly transparent and required similar learning strategies. Greater differences have been reported for late bilinguals (Bidelman and Heath, 2019a,b). Additional research is needed to understand to what extent neural correlates of audiovisual analysis can be generalized to diverse linguistic profiles.

In conclusion, the present study sheds light on the developmental changes of audiovisual syllable processing. Within the first months of reading instruction, children progressively set up letter–sound associations, which can be detected within the first 400 ms of bimodal presentation and recruit the left superior temporal cortex. This reading-dependent brain tuning supports the idea that general mechanisms of audiovisual processing are

applied (at least partially) to new arbitrary correspondences between letters and speech sounds.

## References

- Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ (2005) Functional imaging of human crossmodal identification and object recognition. *Exp Brain Res* 166:559–571.
- Bidelman GM, Heath ST (2019a) Enhanced temporal binding of audiovisual information in the bilingual brain. *Bilingualism* 22:752–762.
- Bidelman GM, Heath ST (2019b) Neural correlates of enhanced audiovisual processing in the bilingual brain. *Neuroscience* 401:11–20.
- Blau V, van Atteveldt N, Formisano E, Goebel R, Blomert L (2008) Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. *Eur J Neurosci* 28:500–509.
- Blau V, Reithler J, van Atteveldt N, Seitz J, Gerretsen P, Goebel R, Blomert L (2010) Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. *Brain* 133:868–879.
- Blomert L (2011) The neural signature of orthographic–phonological binding in successful and failing reading development. *NeuroImage* 57:695–703.
- Blomert L, Froyen D (2010) Multi-sensory learning and learning to read. *Int J Psychophysiol* 77:195–204.
- Booth JR, Burman DD, Meyer JR, Gitelman DR, Parrish TB, Mesulam MM (2004) Development of brain mechanisms for processing orthographic and phonological representations. *J Cogn Neurosci* 16:1234–1249.
- Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. *Cereb Cortex* 11:1110–1123.
- Calvert GA, Brammer MJ, Iversen SD (1998) Crossmodal identification. *Trends Cogn Sci* 2:247–253.
- Carreiras M, Seghier ML, Baquero S, Estévez A, Lozano A, Devlin JT, Price CJ (2009) An anatomical signature for literacy. *Nature* 461:983–986.
- Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9:179–194.
- Dehaene S, Pegado F, Braga LW, Ventura P, Filho GN, Jobert A, Dehaene-Lambertz G, Kolinsky R, Morais J, Cohen L (2010) How learning to read changes the cortical networks for vision and language. *Science* 330:1359–1364.
- Dehaene S, Cohen L, Morais J, Kolinsky R (2015) Illiterate to literate: behavioural and cerebral changes induced by reading acquisition. *Nat Rev Neurosci* 16:234–244.
- Ding SL, Van Hoesen GW, Cassell M, Poremba A (2009) Parcellation of human temporal polar cortex: a combined analysis of multiple cytoarchitectonic, chemoarchitectonic and pathological markers. *J Comp Neurol* 514:595–623.
- Fonov V, Evans AC, Botteron K, Almli CR, McKinstry RC, Collins DL (2011) Unbiased average age-appropriate atlases for pediatric studies. *Neuroimage* 54:313–327.
- Fox E (1994) Grapheme–phoneme correspondence in dyslexic and matched control readers. *Br J Psychol* 85:41–53.
- Froyen D, van Atteveldt N, Bonte M, Blomert L (2008) Cross-modal enhancement of the MMN to speech sounds indicates early and automatic integration of letters and speech sounds. *Neurosci Lett* 430:23–28.
- Froyen D, Willems G, Blomert L (2011) Evidence for a specific cross-modal association deficit in dyslexia: an electrophysiological study of letter–speech sound processing. *Dev Sci* 14:635–648.
- Froyen DJW, Bonte ML, van Atteveldt N, Blomert L (2009) The long road to automation: neurocognitive development of letter–speech sound processing. *J Cogn Neurosci* 21:567–580.
- Goswami U, Ziegler JC (2006) A developmental perspective on the neural code for written words. *Trends Cogn Sci* 10:142–143.
- Grill-Spector K, Malach R (2004) The human visual cortex. *Annu Rev Neurosci* 27:649–677.
- Henson R (2003) Neuroimaging studies of priming. *Prog Neurobiol* 70:53–81.
- Herdman AT, Fujioka T, Chau W, Ross B, Pantev C, Picton TW (2006) Cortical oscillations related to processing congruent and incongruent grapheme–phoneme pairs. *Neurosci Lett* 399:61–66.
- Hocking J, Price CJ (2008) The role of the posterior superior temporal sulcus in audiovisual processing. *Cereb Cortex* 18:2439–2449.



- Holloway ID, van Atteveldt N, Blomert L, Ansari D (2015) Orthographic dependency in the neural correlates of reading: evidence from audiovisual integration in English readers. *Cereb Cortex* 25:1544–1553.
- Horbach J, Scharke W, Cröll J, Heim S, Günther T (2015) Kindergarteners' performance in a sound-symbol paradigm predicts early reading. *J Exp Child Psychol* 139:256–264.
- Horbach J, Weber K, Opolony F, Scharke W, Radach R, Heim S, Günther T (2018) Performance in sound-symbol learning predicts reading performance 3 years later. *Front Psychol* 12:1716.
- Hurlbert A (2000) Visual perception: learning to see through noise. *Curr Biol* 10:R231–R233.
- Insausti R (2013) Comparative neuroanatomical parcellation of the human and nonhuman primate temporal pole. *J Comp Neurol* 521:4163–4176.
- Jost LB, Eberhard-Moscicka AK, Frisch C, Dellwo V, Maurer U (2014) Integration of spoken and written words in beginning readers: a topographic ERP study. *Brain Topogr* 27:786–800.
- Jung TP, Makeig S, Humphries C, Lee TW, McKeown MJ, Iragui VJ, Sejnowski TJ (2000) Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* 37:163–178.
- Karipidis II, Pleisch G, Röthlisberger M, Hofstetter C, Dornbierer D, Stampfli P, Brem S (2017) Neural initialization of audiovisual integration in prereaders at varying risk for developmental dyslexia. *Hum Brain Mapp* 38:1038–1055.
- Karipidis II, Pleisch G, Brandeis D, Roth A, Röthlisberger M, Schneebeli M, Walitza S, Brem S (2018) Simulating reading acquisition: the link between reading outcome and multimodal brain signatures of letter-speech sound learning in prereaders. *Sci Rep* 8:7121.
- Laasonen M, Tomma-Halme J, Lahti-Nuutila P, Service E, Virsu V (2000) Rate of information segregation in developmentally dyslexic children. *Brain Lang* 75:66–81.
- Laasonen M, Service E, Virsu V (2002) Crossmodal temporal order and processing acuity in developmentally dyslexic young adults. *Brain Lang* 80:340–354.
- Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods* 164:177–190.
- Monzalvo K, Dehaene-Lambertz G (2013) How reading acquisition changes children's spoken language network. *Brain Lang* 127:356–365.
- Oostenveld R, Fries P, Maris E, Schoffelen JM (2011) FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput Intell Neurosci* 2011:156869.
- Plewko J, Chyl K, Bola Ł, Łuniewska M, Deóbska A, Banaszkiwicz A, Wypych M, Marchewka A, van Atteveldt N, Jednoróg K (2018) Letter and speech sound association in emerging readers with familial risk of dyslexia. *Front Hum Neurosci* 12:393.
- Pugh KR, Mencl WE, Shaywitz BA, Shaywitz SE, Fulbright RK, Constable RT, Skudlarski P, Marchione KE, Jenner AR, Fletcher JM, Liberman AM, Shankweiler DP, Katz L, Lacadie C, Gore JC (2000) The angular gyrus in developmental dyslexia: task-specific differences in functional connectivity within posterior cortex. *Psychol Sci* 11:51–56.
- Raij T, Uutela K, Hari R (2000) Audiovisual integration of letters in the human brain. *Neuron* 28:617–625.
- Richlan F (2019) The functional neuroanatomy of letter-speech sound integration and its relation to brain abnormalities in developmental dyslexia. *Front Hum Neurosci* 13:21.
- Schlaggar BL, McCandliss BD (2007) Development of neural systems for reading. *Annu Rev Neurosci* 30:475–503.
- Taulu S, Hari R (2009) Removal of magnetoencephalographic artifacts with temporal signal-space separation: demonstration with single-trial auditory-evoked responses. *Hum Brain Mapp* 30:1524–1534.
- Taulu S, Kajola M (2005) Presentation of electromagnetic multichannel data: the signal space separation method. *J Appl Phys* 97:124905–124910.
- van Atteveldt NM, Ansari D (2014) How symbols transform brain function: a review in memory of Leo Blomert. *Trends Neurosci Educ* 3:44–49.
- van Atteveldt NM, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. *Neuron* 43:271–282.
- van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cereb Cortex* 17:962–974.
- Van Veen BD, van Drongelen W, Yuchtman M, Suzuki A (1997) Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Trans Biomed Eng* 44:867–880.
- Vellutino FR, Fletcher JM, Snowling MJ, Scanlon DM (2004) Specific reading disability (dyslexia): what have we learned in the past four decades? *J Child Psychol Psychiatry* 45:2–40.
- Wang F, Karipidis II, Pleisch G, Fraga-González G, Brem S (2020) Development of print-speech integration in the brain of beginning readers with varying reading skills. *Front Hum Neurosci* 14:289.
- Xu W, Kolozsvári OB, Monto SP, Hämäläinen JA (2018) Brain responses to letters and speech sounds and their correlations with cognitive skills related to reading in children. *Front Hum Neurosci* 12:304.
- Xu W, Kolozsvári OB, Oostenveld R, Leppänen PHT, Hämäläinen JA (2019) Audiovisual processing of Chinese characters elicits suppression and congruency effects in MEG. *Front Hum Neurosci* 13:18.
- Xu W, Kolozsvári OB, Oostenveld R, Hämäläinen JA (2020) Rapid changes in brain activity during learning of grapheme-phoneme associations in adults. *Neuroimage* 220:117058.
- Yoncheva YN, Zevin JD, Maurer U, McCandliss BD (2010) Auditory selective attention to speech modulates activity in the visual word form area. *Cereb Cortex* 20:622–632.
- Zachlud D, Rüttingers B, Bludau S, Mohlberg H, Langner R, Zilles K, Amunts K (2020) Four new cytoarchitectonic areas surrounding the primary and early auditory cortex in human brains. *Cortex* 128:1–21.
- Ziegler J, Muneaux M (2007) Orthographic facilitation and phonological inhibition in spoken word recognition: a developmental study. *Psychon Bull Rev* 14:75–80.